

THÈSE PRÉSENTÉE
POUR OBTENIR LE GRADE DE
DOCTEUR
DE L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE MATHÉMATIQUES ET
INFORMATIQUE

Par **Maxime LECOMTE**

Modélisation d'écosystèmes microbiens dans des environnements
contrôlés et non contrôlés

Microbial ecosystem modelling in natural and control
environment

Sous la direction de : **David SHERMAN**
Co-directrice : **Hélène FALENTIN**
Encadrante : **Clémence FRIOUX**

Membres du jury :

Muriel Cocaign-Bousquet	Directrice de recherche	INRAE	Rapportrice
Caroline Baroukh	Chargée de recherche HDR	INRAE	Rapportrice
Ludovic Cottret	Ingénieur de recherche	INRAE	Examinateur
Hélène Falentin	Ingénierie de Recherche HDR	INRAE	Co-Directrice
Clémence Frioux	Chargée de Recherche	Inria de l'université de bordeaux	Encadrante
Béatrice Laroche	Directrice de recherche	INRAE	Examinatrice
Gabriel Markov	Chargé de recherche	CNRS	Examinateur
Laurent Simon	Professeur de l'université de Bordeaux	LaBRI	Président
David Sherman	Directeur de Recherche	Inria de l'université de bordeaux	Directeur

Table des matières

1	Introduction	6
1.1	De la microbiologie à l'écosystème microbien	6
1.2	De la biologie des systèmes à l'écologie des systèmes	7
1.3	Objectifs de la thèse	8
2	Etat de l'art	10
2.1	L'abstraction du métabolisme comme support de l'analyse microbienne	11
2.1.1	Reconstruction de réseaux métaboliques	11
2.1.2	Représentation de la connaissance du métabolisme	12
2.1.3	Représentations mathématiques du métabolisme	12
2.1.3.1	Matrice stœchiométrique	13
2.1.3.2	Graphe	14
2.1.4	Comparaison de réseaux métabolique	14
2.1.5	Taille et qualité des réseaux métaboliques	15
2.2	Analyse métabolique à l'échelle d'une cellule	17
2.2.1	Méthodes numériques d'un réseau métabolique	17
2.2.1.1	Constantes cinétiques des réactions	17
2.2.1.2	Contraintes numériques	17
2.2.1.3	les modes de flux élémentaires	18
2.2.2	Méthodes qualitatives d'analyse d'un réseau métabolique	19
2.2.2.1	Graphes	19
2.2.2.2	Logiques	19
2.2.3	Défis rencontrés lors de la modélisation qualitative et numérique d'une cellule	21
2.2.4	Méthodes hybrides à l'échelle d'un modèle métabolique	21
2.2.4.1	Indépendants	21
2.2.4.2	Couplé	22
2.2.4.3	Inference temporelle dans la programmation logique	22
2.3	Analyse du métabolisme les écosystèmes microbiens	23
2.3.1	Méthodes numériques appliqués à plusieurs réseaux métaboliques	23
2.3.1.1	Equation différentielles ordinaires (ODE)	23
2.3.1.2	Analyse statique	23
2.3.1.3	Intégration de données multi-omiques	24
2.3.1.4	Analyse dynamique	24
2.3.2	Applications	25
2.3.3	Méthodes qualitatives de l'étude d'une communauté bactérienne	27
2.3.3.1	Basées sur les graphes	27
2.3.3.2	Basées sur le raisonnement	27

2.3.4	Défis rencontrés lors de l'analyse de la modélisation qualitative et quantitative du métabolisme	28
2.3.5	Modélisation hybride du métabolisme	28
3	TANGO : un jumeau numérique du métabolisme bactérien de la production d'un fromage	30
3.1	Introduction	31
3.2	Réultats	32
3.2.1	Les modèles métaboliques individuels identifient des voies métaboliques spécifiques dans le lait	32
3.2.1.1	Métabolisme des bactéries lactiques	33
3.2.2	Optimisation dynamique des modèles individuels	34
3.2.3	Le suivie de la production des métabolites dans les conditions de co-cultures par le modèle de communauté	38
3.2.4	Les interactions bactériennes dynamiques au sein de la communauté	42
3.3	Discussion	46
3.4	Materiels et méthodes	47
3.4.1	Données biologiques	47
3.4.1.1	composition communauté.	47
3.4.1.2	Fabrication de fromage et collecte des échantillons.	48
3.4.1.3	Expériences de culture souche spécifique.	48
3.4.1.4	Données métatranscriptomiques.	48
3.4.1.5	Métabolomiques ciblées.	49
3.4.2	Modélisation métabolique	50
3.4.2.1	Séquençage du génome.	50
3.4.2.2	Reconstruction de modèle métaboliques à l'échelle du génome.	50
3.4.2.3	Analyse par équilibre des flux.	50
3.4.2.4	Raffinement manuel des GEMs	51
3.4.2.5	Modélisation dynamique	51
3.4.2.6	Implementation numérique	52
3.4.3	Ajustement du modèle	53
3.4.4	Inférence des interactions métaboliques	53
3.4.5	Visualisation et analyses statistiques	53
3.4.6	Code et disponibilité des données	54
4	CoCoMiCo : Modélisation logique de communautés bactériennes au sein d'écosystèmes non-contrôlé	55
4.1	Introduction	57
4.2	Représentation de la connaissance	58
4.3	L'approche basé sur le raisonnement pour approximer la coopération et la compétition dans des grandes communautés	59
4.3.1	Les métabolites échangés utilisés comme un indicateur de la coopération.	59
4.3.2	La consommation de métabolites limitants utilisé pour approximer la compétition.	60
4.3.3	Description des interactions potentielles dans les communautés microbienne au moyen du raisonnement.	60
4.4	Calcul des potentiels de coopérations et de compétitions	60

4.4.1	Les propriétés voulus des critères d'optimisation	60
4.4.2	Calcule du potentiel de coopération CooP	60
4.4.3	Calcul du score du potentiel de compétition ComP	61
4.4.4	Illustration sur l'exemple jouet	61
4.4.5	Implémentation	62
4.5	Analyses comparatives	62
4.5.1	Données et génération de communautés artificielles	62
4.5.2	Potentiels de coopération et de compétition à travers les différents écosystèmes	63
4.5.3	Impact de la similarité des GEMs sur les potentiels de compétition et de coopération	64
4.5.4	Impact des changements de compositions bactériens de la communauté sur les potentiels	64
4.5.5	Performance	65
4.6	Comparaison à des méthodes numériques	65
4.6.1	Méthodes	66
4.6.2	Comparaison des approches par contraintes et celle par le raisonnement	66
4.7	Application à des données réelles	67
4.7.1	Méthodes	67
4.7.2	Application aux données expérimentales et impact de la méthode de reconstruction	68
4.8	Discussion	69
5	Enrichissement des modèles qualitatifs : Modèle de logique linéaire temporel et processus de selection de communauté à façon	71
5.1	SelCom : Sélection de communauté basé sur des scores	72
5.2	LTL ; intégration temporelle	72
6	Approche de modélisation hybride, [ML – pas vraiment dans le plan]	73
6.0.1	Approches hybrides découpées et couplées	74
7	Annexes	75
.1	Figures supplémentaires	75
.2	Matériel supplémentaire	78
.2.1	Pure culture data	78
.2.1.1	Strain pure culture data	78
.2.1.2	Acids dosage for <i>P. freudenreichii</i>	78
.2.2	Raffinement manuel des GEMs	79
.2.3	Données expérimentales communautaires	80
.2.4	Exchanged metabolites highlighted by computational approaches	80
.2.4.1	MICOM	80
.2.4.2	SMETANA	81
.2.4.3	Metage2Metabo	81
.2.5	Bornes dynamiques sur les réactions d'échanges	81
.2.6	Bornes dynamiques sur les réactions intracellulaires	84
.2.7	Estimation des bornes sur la production de métabolite pour <i>P. freudenreichii</i>	84
.2.8	Valeurs de paramètres	84

.1	Taxonomie des génomes utilisé pour reconstruire les communautés simulés	86
.2	Prédiction des potentiels de coopération et de compétition au sein des communautés simulées dans les divers écosystèmes.	86
.2.1	Absence d'association entre la topologie du réseau et les interactions potentielles	86
.2.2	Potentiels de coopération et de compétition à travers différentes taille de communautés.	86
.2.3	Les potentials de coopération et de compétition dans un écosystème non réel.	87

Chapitre 1

Introduction

Sommaire

1.1	De la microbiologie à l'écosystème microbien	6
1.2	De la biologie des systèmes à l'écologie des systèmes	7
1.3	Objectifs de la thèse	8

[ML – Commentaire général : pas assez info ?]

1.1 De la microbiologie à l'écosystème microbien

La science des micro-organismes, appelée microbiologie, est historiquement focalisée sur les maladies infectieuses affectant à la fois l'homme et les animaux. Un des pionniers au sein de cette communauté scientifique, Van Leeuwenhoek, réalisa une description précise des différents micro-organismes comprenant les protozoaires, champignons et bactéries. Dans les années 1800, Louis Pasteur découvrit que les bactéries étaient responsables de l'acidité du vin et des produits fermentés et l'âge d'or de la microbiologie suivit se basant sur les premières théories de croissance et propagations bactériennes (Cal and Pasteur, 1947). Au début du $XX^{\text{ème}}$ siècle, la microbiologie moderne voit le jour mettant en avant des étapes de contrôles de qualité dans les domaines des produits pharmaceutiques, l'alimentaire tel que la production de produits fermentés(Adams and Mitchell, 2002), l'eau potable (LeChevallier, 2013) et dans des applications industrielles. Ces étapes de contrôles consistent principalement à évaluer la croissance et l'abondance de micro-organismes présents dans leur écosystème garantissant ainsi, la qualité du produit final. Un écosystème microbien est ainsi décrit comme étant un ensemble dynamique de micro-organismes, interagissant entre eux *via* des **interactions bactériennes**, ou avec son environnement dans lequel ils vivent.

Ces écosystèmes microbiens sont présents partout et jouent un rôle précis pouvant être perturbé par des facteurs biotiques, i.e. les organismes vivants interagissant entre eux, ou abiotiques, i.e. tout ce qui est non vivants. Parmi les nombreux écosystèmes microbiens, tels que celui de l'intestin, de la peau, du sol, des produits alimentaires, des racines et feuilles de la plante, les micro-organismes portent des fonctions spécifiques de leur habitat. Grâce notamment à des interactions entre micro^{organismes}, les souches microbiennes de l'intestin sont, par exemple, chargées de notre protection contre divers pathogènes (Zhang et al., 2015), ceux du sol contribuent au recyclage des matières organiques (Hoorman, 2011), tout comme dans des écosystèmes agro (Yadav et al., 2018). Au sein des écosystèmes

de produits fermentés tels que le fromage ou le vin, les bactéries lactiques seront plus spécialisées dans la fermentation bactérienne, malolactique dans le cas du vin, libérant des produits aromatiques essentiels (Cao et al., 2021) à la qualité du produit final (Virdis et al., 2021). Sans perturbations, ces micro-organismes maintiennent un équilibre dans leur habitat. Une modification d'un des facteurs abiotiques tels que le pH, la température ou même la concentration d'oxygène, peuvent impacter l'équilibre de l'ecosystème(Beveridge et al., 2010). En effet, certaines bactéries ne peuvent se développer qu'à une température, un pH ou une concentration d'oxygène précise, là où d'autres ne peuvent survivre.

Être capable de capturer l'impact de ces facteurs biotiques et abiotiques à l'échelle d'un écosystème microbien, reconnus comme des moteurs pouvant modifier la structure et la fonction de l'écosystème (Lee et al., 2019), est un défi biologique. En effet, les microbiologistes sont confrontés à un problème d'échantillonnage et d'expertise (Wardle and Peltzer, 2007) et les études se focalisent plus sur le test d'hypothèses associées à un facteur biotique prédéfini. Pour être ainsi capable de mener au mieux ce but, il faut donc mélanger des approches interdisciplinaires, permettant notamment d'utiliser des approches d'échantillonnage et d'intégrer des variables représentant les facteurs biotiques et abiotiques, pour avoir une meilleure compréhension de l'écosystème : **Biologie des systèmes**.

1.2 De la biologie des systèmes à l'écologie des systèmes

Au moyen des données des expériences, de la modélisation et de la simulation, la biologie des systèmes permet une meilleure compréhension quantitative ou qualitative d'un système biologique, comme par exemple un écosystème microbien, en se basant sur des analyses mathématiques, statistiques et informatiques pour développer des modèles prédictifs. La biologie des systèmes a permis l'analyse à différents échelles, du génome à la communauté. Avec l'aide de ces modèles, la compréhension du fonctionnement d'une espèce a pu être ainsi expérimenté en procédant à de multiples perturbations sur la fonction des gènes (Butland et al., 2008 ; Tong et al., 2001). De plus, dans (Fischbach and Krogan, 2010) des interactions entre une espèce bactérienne et son hôte peut également être étudié en variant la composition de l'environnement nutritionnel ou ajoutant des mutants par exemple. Enfin, tous ces phénomènes peuvent être étudiés dynamiquement, ce qui renforce son utilisation auprès des biologistes. Étant donné que les interactions entre bactéries ou avec son hôte se fait via l'échange de molécules impactant ou non le phénotype du récepteur (la bactérie ou l'hôte) (Faust and Raes, 2012), modéliser le métabolisme en proposant des modèles prédisant des interactions et affinant la compréhension d'une espèce au sein d'un écosystème semble être une solution. Tout comme la microbiologie a évolué vers l'étude des micro-organisme à l'échelle d'un écosystème, les systèmes biologiques se sont également complexifié permettant l'analyse et la prédiction à l'échelle d'un écosystème entier. De ce fait, de nouvelles méthodes discrètes et numériques voient le jour, en mettant en avant les interactions bactériennes, via l'étude du métabolisme, à l'échelle d'un écosystème, et en essayant de trouver un compromis entre précision des résultats, passage à l'échelle de la méthode et disponibilité des données.

Les données du type **omiques** constituent les données de bases pour permettre l'analyse et la caractérisation *in silico* des écosystèmes microbiens. Parmi l'ensemble des données omiques, les données génomiques annotées, où au moins un gène est associé à une

séquence codante du génome, et les données méta-génomiques sont les plus fréquemment utilisées. Elles permettent notamment la reconstruction du métabolisme, sous forme d'un réseau métabolique à l'échelle d'un génome en décrivant l'ensemble des réactions bio-chimiques d'une espèce, sous forme d'une association de gene-protéine-reaction (GPR) (Thiele and Palsson, 2010), et ainsi, caractériser le comportement d'une espèce ou des plusieurs espèces au sein d'un écosystème.

1.3 Objectifs de la thèse

Du point de vue d'un modélisateur ou d'une modélisatrice, le défi principal est la création d'un modèle avec le meilleur compromis entre précision des résultats, versatilité du modèle et temps de calcul en fonction des données en entrées. Cette problématique est retrouvée à la fois lors de la modélisation à l'échelle d'une cellule qu'à l'échelle d'un écosystème. En effet, cibler des communautés naturelles rapidement, avec une parfaite compréhension de l'ensemble du métabolisme de chaque individu au sein de la communauté permettant la prédiction de nouveau processus de régulation est utopique. Dans l'état de l'art concernant la modélisation numérique du métabolisme, la construction de modèles précis requiert soit une grande connaissance *a priori* du système biologique (ODE, constantes cinétiques) ou est limité par le temps de calcul d'une optimisation numérique. A l'inverse, les modèles discrets basés sur les graphes ou sur une optimisation combinatoire permet ce passage échelle, sont moins coûteux en temps mais leur précision est de l'ordre du qualitatif. Un compromis doit-être fait pour résoudre le criblage de grandes communautés bactérienne. Dans cette thèse, nous nous sommes intéressé à la caractérisation de communautés bactériennes, à la fois dans un milieu contrôlé où des facteurs biotiques et abiotiques sont suivis, et à la fois dans un environnement non contrôlé, à l'aide de deux approches mathématiques distinctes : modélisation numérique et discrète du métabolisme.

Dans un premier temps, nous verrons l'apport d'un modèle numérique dans un environnement contrôlé. Le projet TANGO est composé de trois bactéries fromagères, *L. lactis*, *L. plantarum* et *P. freudenreichii*. Nous avons développé une stratégie numérique basée sur le FBA pour l'intégration de multiples données omiques. Elle consiste en un raffinement des modèles FBA individuels par rapport à la connaissance biologique que l'on trouve dans la littérature ou apporté par l'expert ou l'experte biologiste. A la suite, une calibration dynamique des modèles FBA par rapport aux données de croissance et à des données d'acides est faite. Enfin nous prédisons les croissances et les concentrations de chaque composé suivi dynamiquement en mettant en avant les contributions de chaque espèce. Une attention sera donnée aux interactions bactériennes mis en avant par cette stratégie.

Dans un deuxième temps, le développement d'une approche par raisonnement applicable sur des écosystèmes naturels pour l'optimisation d'une communauté sera présenté. En effet, nous criblons des communautés créés aléatoirement provenant du sol, de la racine et feuille de la plante, et de l'intestin. A l'aide d'une approche de raisonnement, nous décrivons les communautés bactérienne sous forme d'un vocabulaire contrôlé. Ce dernier est alors utilisé pour estimer la coopération et la compétition dans les différents écosystèmes pour la suite, comparer des communautés de tailles équivalentes.

Enfin, nous terminerons par proposer une méthode de sélection de communauté à façon basé sur le raisonnement. Les utilisateurs et les utilisatrices de la méthode renseignent, à

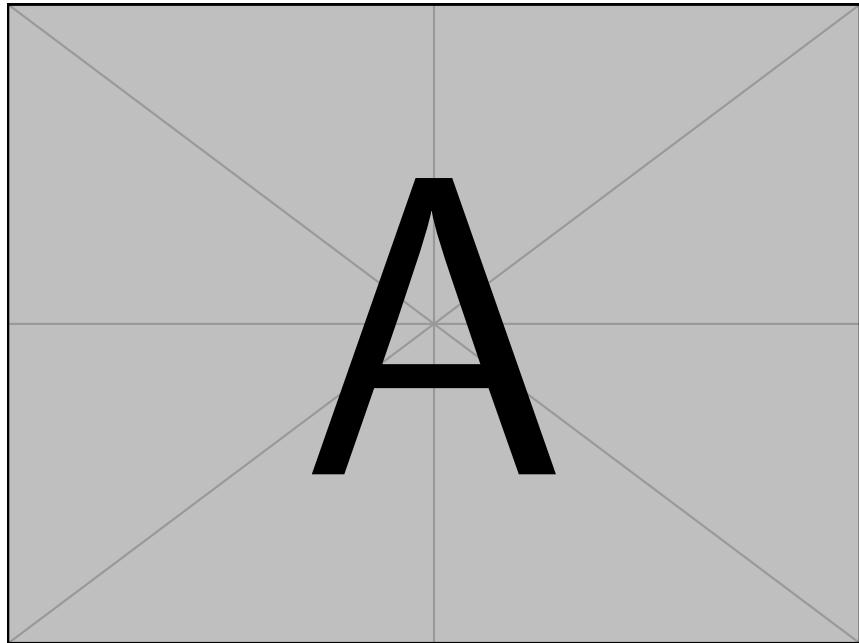


FIGURE 1.1 – a) Montrer qu'il y a des bactéries partout, à la fois dans les écosystèmes naturels (avec hôte ou non) ou industriels. b) montrer les différentes interactions bactériennes, leur impact sur les espèces (dans l'idée de celle de Faust et al., 2012). MESSAGE GLOBAL : peu importe la communauté, des interactions impactant l'écosystème.

partir d'une base de connaissance, les critères de sélection de la composition bactérienne souhaité. Puis, une maximisation ou une minimisation d'un ou de multiples objectifs pourra être ainsi fait. Cette méthode est une extension de l'outil MISCOTO (Frioux et al., 2018), qui permettait la sélection de communautés microbiennes minimales.

Dans le chapitre 2, nous vous proposons une revue de la littérature concernant les différents moyens de représenter le métabolismes ainsi que les méthodes d'analyses existantes. Le second chapitre présente les résultats de l'application du développement d'une stratégie numérique sur un environnement contrôlé. Le chapitre 3, est consacré au développement d'une méthode de criblage à haut débit de communauté microbienne. Enfin, dans le chapitre 4, nous explorons une stratégie permettant de connecter un modèle numérique avec un modèle discret.

Chapitre 2

Etat de l'art

Sommaire

2.1 L'abstraction du métabolisme comme support de l'analyse microbienne	11
2.1.1 Reconstruction de réseaux métaboliques	11
2.1.2 Représentation de la connaissance du métabolisme	12
2.1.3 Représentations mathématiques du métabolisme	12
2.1.4 Comparaison de réseaux métabolique	14
2.1.5 Taille et qualité des réseaux métaboliques	15
2.2 Analyse métabolique à l'échelle d'une cellule	17
2.2.1 Méthodes numériques d'un réseau métabolique	17
2.2.2 Méthodes qualitatives d'analyse d'un réseau métabolique	19
2.2.3 Défis rencontrés lors de la modélisation qualitative et numérique d'une cellule	21
2.2.4 Méthodes hybrides à l'échelle d'un modèle métabolique	21
2.3 Analyse du métabolisme les écosystèmes microbiens	23
2.3.1 Méthodes numériques appliqués à plusieurs réseaux métaboliques	23
2.3.2 Applications	25
2.3.3 Méthodes qualitatives de l'étude d'une communauté bactérienne	27
2.3.4 Défis rencontrés lors de l'analyse de la modélisation qualitative et quantitative du métabolisme	28
2.3.5 Modélisation hybride du métabolisme	28

2.1 L'abstraction du métabolisme comme support de l'analyse microbienne

2.1.1 Reconstruction de réseaux métaboliques

Le métabolisme décrit toutes les réactions biochimiques au sein d'un organisme permettant de produire de l'énergie, de dégrader et de produire du matériel métabolique. Un moyen de le représenter est d'utiliser le réseau métabolique comme une abstraction du métabolisme. Pour ce cela, un génome annoté est nécessaire. A partir d'une séquence génomique d'un organisme, deux annotations du génome sont possibles (M??digue et al., 2002) : structurelle, i.e. permettant d'identifier les gènes ~~associés~~ (par exemple, AUGUSTUS (Nachtweide and Stanke, 2019), Glimmer (Delcher et al., 2007)), et fonctionnelle, i.e. retrouver la biochimie et les fonctions biologiques portées par une protéine (par exemple, Blast (Altschul et al., 1990), par un processus d'alignement). Un grand panel de bases de données existe permettant une annotation à plusieurs échelles du génome. Il existe BRENDa (Chang et al., 2021) se concentrant sur les enzymes et les données liées au métabolisme, ou encore Prokka (Seemann, 2014), Rast (Aziz et al., 2008), MicrosCope (Vallenet et al., 2020) ou encore EggNOG-mapper (Cantalapiedra et al., 2021 ; Huerta-Cepas et al., 2019) se concentrant sur l'annotation fonctionnelle du génome.

Le génome étant annoté, le réseau métabolique peut-être reconstruit manuellement et/ou automatiquement avec l'approche ascendante (bottom-up) et/ou l'approche descendante (top-down), en utilisant une large collection d'outils (Mendoza et al., 2019). Tous les outils de reconstruction utilisent une base de données leur permettant d'assigner une fonction métabolique porté par un gène. Les approches bottom-up se caractérisent par un processus itératif de génération d'ébauche de réseaux métaboliques tandis que les approches descendante, sont définies par la production de réseaux métaboliques "prêt à l'emploi". En effet, au lieu d'assigner toutes les fonctions possibles directement à partir d'une base de donnée, les modèles métaboliques reconstruits à partir d'un meta-organisme, sont fonctionnels et manuellement curés, leur assurant ainsi une bonne qualité dû à leur faible nombre d'assignation faux-positif.

Un réseau métabolique se caractérise donc par l'association d'un gène à une protéine qui catalyse une réaction (GPR) Thiele and Palsson (2010). Cette association se base sur des règles d'association booléennes entre les gènes et les réactions. Considérons les réactions r_1 , r_2 associées aux gènes g_1 , g_2 , g_3 par les relations logiques suivantes :

$$r_1 : g_1 \wedge g_2 \vee g_1 \wedge g_3 \quad (2.1)$$

$$r_2 : g_1 \vee g_2 \wedge g_1 \vee g_3 \quad (2.2)$$

La règle booléenne 2.1 stipule que, pour que la réaction r_1 soit activée, il faut que le complexe g_1 et g_2 soit exprimé ou que le complexe g_1 et g_3 soit exprimé. Dans le cas de la règle d'association 2.2, la réaction r_2 est activée si g_1 est exprimé avec g_2 ou g_3 . Dans ce cas précis, si seul g_2 ou g_3 est exprimé mais que g_1 ne l'est pas, alors la réaction r_2 n'est pas activée.

Pour permettre la reconstruction d'un réseaux métabolique, et par extension, obtenir les associations gène-protéine-réactions, plusieurs outils s'offre à nous. On retrouve GEMSiRV (Liao et al., 2012), CarveME (Machado et al., 2018) et MetaDraft (Olivier, 2018) qui représentent les outils utilisant l'approche descendante, les deux derniers se servant de la base de données BIGG (King et al., 2016) pour construire leurs modèles universels. Une plus grande bibliographie existe concernant l'approche ascendante avec FAME (Boele and Olivier, 2012), CoReCo (Pitkänen et al., 2014), Merlin (Dias et al., 2015), RAVEN (Wang et al., 2018) et autoKEGGRec (Karlsen et al., 2018) se reposant sur la base de données KEGG (Kanehisa and Goto, 2000). Pathway-tools (Karp et al., 2022) ou encore AuReMe (Aite et al., 2018), tous deux basés sur la base de données MetaCyc (Caspi et al., 2018), ModelSEED (Seaver et al., 2021) et Kbase (Arkin et al., 2018) ou encore gapSeq (Zimmermann et al., 2021). Un peu moins récent, il y a rBioNet (Thorleifsson and Thiele, 2011), un outil hybride, alliant à la fois la reconstruction descendante et ascendante, permettant à l'utilisateur ou l'utilisatrice de créer elle même son réseau métabolique ou d'en importer un sur lequel peut se baser pour le reconstruire.

2.1.2 Représentation de la connaissance du métabolisme

On peut voir qu'il existe plusieurs outils permettant de reconstruire le métabolisme, cependant, standardiser sa représentation n'est pas encore acquis par la communauté scientifique (Stobbe et al., 2014). En effet, chaque base de données possèdent les fichiers de sortie différent permettant la représentation de la connaissance du métabolisme. On peut retrouver par exemple, BioCyc (Karp et al., 2018) qui utilise des fichiers plats pathway-tools, le format BioPAX (Demir et al., 2010) ou encore SBML (Hucka et al., 2019). La base de données KEGG (Kanehisa and Goto, 2000) utilise d'autres fichiers plats et sa propre façon de représenter le métabolisme : KEGG Markup Language. Même si le choix de représentation est différent, on retrouve majoritairement le format SBML pour System Biology Markup Language (Hucka et al., 2019), qui permet de représenter les voies signalétiques mais également, des paramètres mathématiques utiles pour la modélisation.

2.1.3 Représentations mathématiques du métabolisme

Le métabolisme peut se représenter de deux grandes façons : sous forme d'un graphe, orienté, non orienté ou encore bipartite, ou sous forme d'une matrice stœchiométrique. La représentation sous graphe nous informe sur les liens entre les réactions et métabolites d'un système biologique. La matrice nous apporte une notion quantitative : combien de métabolites sont consommés et/ou produits. Nous allons prendre l'exemple d'une voie de production connue, la glycolyse, chez une espèce lactique : *Lactiplantibacillus plantarum*, noté LP.

Représentation simplifiée de la glycolyse chez LP :

- (HEX1) : 1 glucose → 1 glucose-6-phosphate
- (PGI) : 1 glucose-6-phosphate ↔ 1 fructose-6-phosphate
- (PFK) : 1 fructose-6-phosphate → 1 D-Fructose 1,6-bisphosphate
- (FBA) : 1 D-Fructose 1,6-bisphosphate ↔ 1 Dihydroxyacetone phosphate
+1 Glyceraldehyde 3-phosphate
- (GAPD) : 1 Glyceraldehyde 3-phosphate ↔ 1 3-Phospho-D-glyceroyl phosphate
- (PGK) : 1 3-Phospho-D-glycerate ↔ 1 3-Phospho-D-glyceroyl phosphate

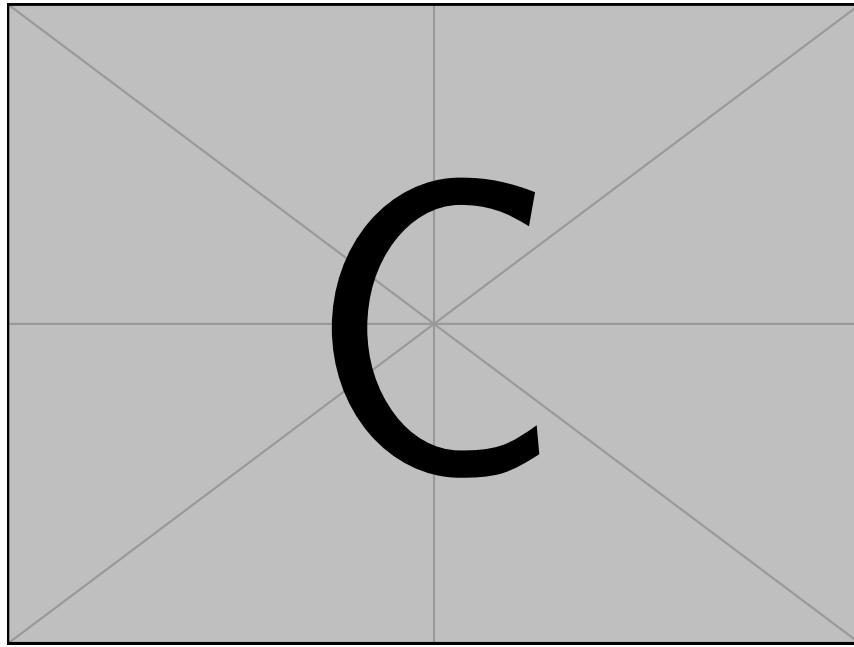


FIGURE 2.1 – Illustration approche top-down et bottom-up avec différents output et input (sbml,genome annotée ...)

(PGM) : 1 3-Phospho-D-glycerate \leftrightarrow 1 2-Phospho-D-glycerate

(ENO) : 1 2-Phospho-D-glycerate \leftrightarrow 1 Phosphoenolpyruvate

(PYK) : 1 Phosphoenolpyruvate \rightarrow 1 pyruvate

où HEX1, PGI, PFK, FBA, GAPD, PGK, PGM, ENO, PYK sont les ensembles de réactions et glucose, glucose-6-phosphate, fructose-6-phosphate, D-Fructose 1,6-bisphosphate, Dihydroxyacetone phosphate, Glyceraldehyde 3-phosphate, 3-Phospho-D-glyceroyl phosphate, 3-Phospho-D-glycerate, 2-Phospho-D-glycerate, Phosphoenolpyruvate et pyruvate sont les ensembles de métabolites principaux constituants la glycolyse de LP.

2.1.3.1 Matrice stœchiométrique

Le métabolisme peut-être représenté sous forme d'une matrice stœchiométrique dont l'ensemble des réactions et des métabolites du système constituent respectivement les colonnes et les lignes. Le nombre à l'intérieur de la matrice représente le coefficient stœchiométrique de la réaction biochimique. Ainsi, la glycolyse chez LP sous forme d'une matrice stœchiométrique est représenté de la manière suivante :

	HEX1	PGI	PFK	FBA	GAPD	PGK	PGM	ENO	PYK
glucose	-1	0	0	0	0	0	0	0	0
glucose-6-phosphate	1	-1	0	0	0	0	0	0	0
fructose-6-phosphate	0	1	-1	0	0	0	0	0	0
D-Fructose 1,6-bisphosphate	0	0	1	-1	0	0	0	0	0
Dihydroxyacetone phosphate	0	0	0	1	-1	0	0	0	0
Glyceraldehyde 3-phosphate	0	0	0	1	-1	0	0	0	0
3-Phospho-D-glyceroyl phosphate	0	0	0	0	1	-1	0	0	0
3-Phospho-D-glycerate	0	0	0	0	0	1	-1	0	0
2-Phospho-D-glycerate	0	0	0	0	0	0	1	-1	0
Phosphoenolpyruvate	0	0	0	0	0	0	0	1	-1
pyruvate	0	0	0	0	0	0	0	0	1

Nous pouvons remarquer que ce type de représentation ne permet pas de visualiser les réactions réversibles.

2.1.3.2 Graphe

Pour représenter des réseaux métaboliques, un graphe bipartite dirigé pondéré est souvent privilégié (Barabási and Oltvai, 2004). Le graphe est alors composé de deux types de sommets, un pour les métabolites et un pour les réactions, et d'arêtes symbolisant les relations "être produit par" ou "être le consommé par" entre les métabolites et les réactions. Mathématiquement, il peut être exprimé ainsi : soit un graphe bipartite G définit comme un tuple $G = \langle M, R, E \rangle$ où R , M et E représentent les ensembles respectifs de Réactions, de Métabolites et des Arc reliant les métabolites aux réactions. Un réactant m consommé par la réaction r est définie comme $e = \langle m, r \rangle$, et un produit m de la réaction r est définie comme suit $e = \langle m, r \rangle$ où $m \in M$ et $r \in R$.

Une représentation de la glycolyse de LP plus visuelle sous forme d'un graphe bipartite est décrite avec la figure 2.2, où les carrés représentent les réactions et les cercles représentent les métabolites.

2.1.4 Comparaison de réseaux métabolique

Le métabolisme est souvent représenté sous forme d'un graph bipartite (Belcour et al., 2020) où un ensemble de sommets est associé aux métabolites et l'autre ensemble aux réactions biochimiques. Les relations entre ces sommets peuvent être orientées ou non, permettant des analyses différentes. En effet, dans un graphe non orienté, seule la structure du réseau peut être analysée et comparée. La topologie peut se faire à l'échelle des substrats uniquement (Jeong et al., 2011) ou des réactions uniquement (Wagner and Fell, 2001). La topologique de réseaux nous permet d'identifier des différences fonctionnelles, de mettre en avant des motifs souches dépendantes ou encore, de formuler des hypothèses de capacité biosynthétiques entre les organismes, sans prendre en compte les interactions entre bactéries (Biggs et al., 2015). En effet, dans (Weiss et al., 2022), ils soulignent une grande diversité de super voie métabolique appartenant au métabolisme centrale et une prédominance de fonctions de dégradations des acides aminés au sein de la communauté. Ou encore, dans (Ay et al., 2012), ils utilisent une technique d'alignement de réseaux métabolique pour identifier des motifs connectés conservés parmi deux ou plusieurs organismes. Dans un réseau complexe, représenter la connectivité d'un substrat est très massivement utilisé par la communauté scientifique (Ma and Zeng, 2003), cependant, les co-facteurs tels que ATP, ADP etc sont sur représentés et sont peu informatifs pour les biologistes (Sweetlove and Fernie, 2005). En revanche, identifier les composés connectés ayant une importance pour la robustesse du réseau permettent de montrer des propriétés telles que la perturbation du réseau en raison d'une délétion enzymatique (Lemke et al., 2004), la comparaison structurelle entre organisme (Ma and Zeng, 2003) ou encore, le nombre moyen de plus petit chemins(Wagner and Fell, 2001).

Cette méthode de comparaison est purement qualitative mais permet de souligner des hypothèses sur ce que chaque réseau métabolique pourrait être capable de produire et de consommer entre autre. Pour vérifier ces hypothèses, on utilise la modélisation du métabolisme à travers des approches numériques et qualitatives permettant de générer des hypothèses testables en laboratoire sur des différents systèmes biologiques et de prédire

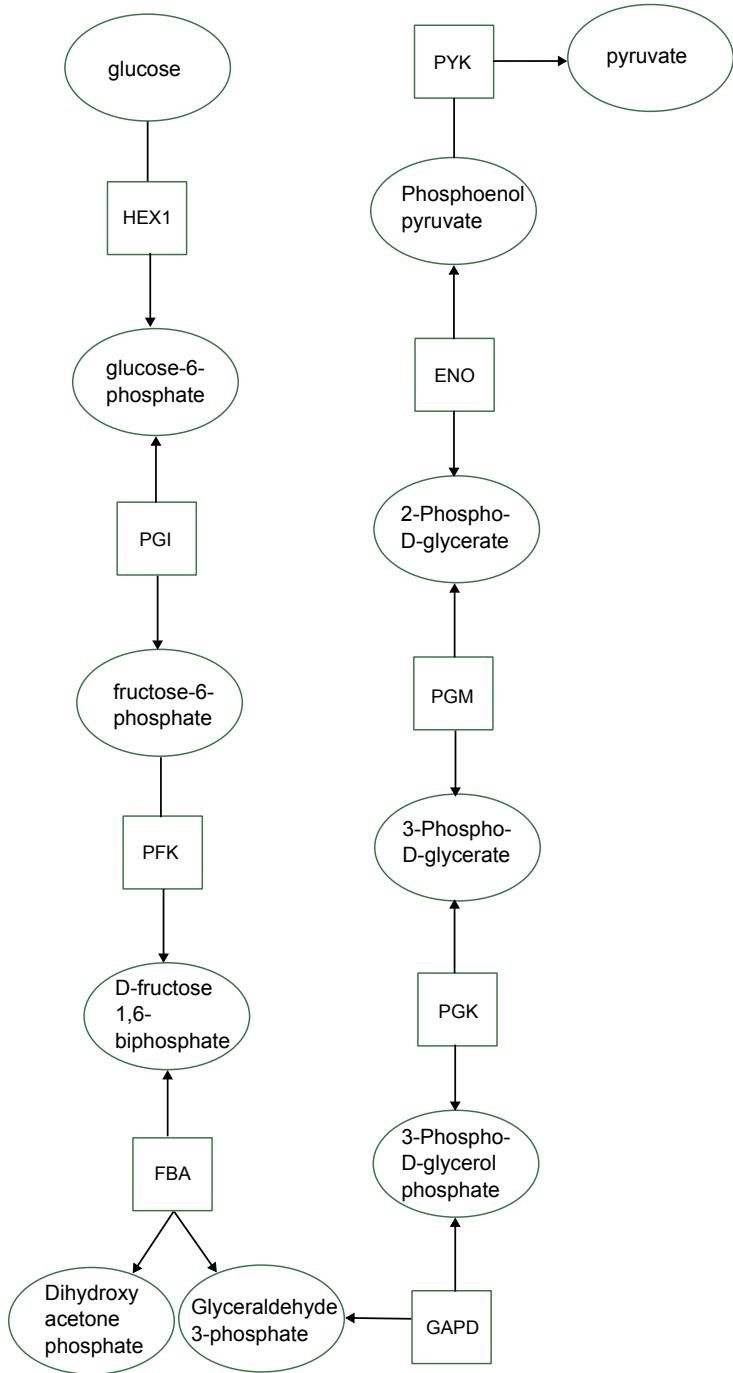


FIGURE 2.2 – Représentation du métabolisme sous forme d'un graphe bi-partite. Les carrés sont associés aux réactions, les cercles aux métabolites et les flèches représentent les arcs reliant les métabolites aux réactions.

de nouvelles connaissances. De plus, l'élaboration de tels modèles métaboliques permet l'intégration de diverses données omiques, affinant ainsi les prédictions (Passi et al., 2022).

2.1.5 Taille et qualité des réseaux métaboliques

Par convention, le plus petit réseau métabolique sur lequel une analyse peut être faite est constitué d'une réaction de production et d'une réaction de consommation et peut atteindre plus de dix mille réactions dans le cas du réseau métabolique de l'humain (Robinson et al., 2020).

Concernant l'évaluation de la qualité du réseaux métabolique, au delà de l'objectif

même de l'utilisateur ou de l'utilisatrice, il existe MEMOTE (Lieven et al., 2020) capable de vérifier la consistance du réseau métabolique en terme du nombre de gènes associés à des réactions, le nombre de réaction non associé à au moins un gène ou encore le nombre de métabolites qui ne sont jamais consommé mais produit, et donc, étant en accumulation dans le système. Le score évaluant la qualité du réseau est borné par la méthode de reconstruction choisi (Lieven et al., 2020).

2.2 Analyse métabolique à l'échelle d'une cellule

La modélisation du métabolisme est la capacité de construire, à partir d'un réseau métabolique, un modèle métabolique capable de prédictions et pouvant être contraint par des données multi-omiques. On peut trouver dans la littérature de nombreuses méthodes d'analyse du métabolisme, aussi bien à l'échelle d'un mécanisme métabolique particulier que à l'échelle d'une communauté de bactériennes de plusieurs centaines de bactéries, se basant sur divers formalismes mathématiques : numérique, logique et hybride. Une revue de l'ensemble de ces méthodes est ainsi proposée dans les paragraphes suivants.

2.2.1 Méthodes numériques d'un réseau métabolique

La modélisation numérique du métabolisme nous informe quantitativement sur des mécanismes biologiques intracellulaires d'une espèce ou inter-espèces en se basant sur la matrice stœchiométrique ou bien sur les données cinétiques.

2.2.1.1 Constantes cinétiques des réactions

Dans le but de prédire le taux de biomasse ou la concentration d'un métabolite dans le milieu extra-cellulaire, des modèles cinétiques se basant sur des constantes cinétiques et des équations du type Michaelis-Menten permettant de lier l'affinité d'un substrat à l'enzyme de la réaction (Costa et al., 2016) existent. L'avantage des modèles cinétiques est qu'ils assurent un haut niveau de confiance sur les prédictions de part l'inférence des paramètres cinétiques, en revanche, ces modèles nécessitent d'obtenir ces données cinétiques limitant ainsi leur utilisation qu'à des réseaux métaboliques plus petit (de l'ordre de la vingtaine de réactions) (van Rosmalen et al., 2021).

2.2.1.2 Contraintes numériques

Les méthodes basées sur des contraintes permettent de prouver mathématiquement une solution en se basant sur des limites d'un espace de recherche. Ainsi, toutes les variables qui partagent les mêmes contraintes seront présentes dans le même espace de recherche définit en amont. Ce type d'approches se base sur des techniques d'optimisations maximisant ou minimisant une fonction objective.

Parmi elles, on retrouve les méthodes d'analyse par équilibre des flux ou FBA. Elles ont pour but d'estimer la vitesse cinétique de chaque réaction présente dans le métabolisme. Cela peut se faire en respectant quelques contraintes d'optimisation, qui sont la maximisation ou la minimisation d'une fonction objectif, l'hypothèse que le système est à l'état stable et enfin, borner l'espace de solutions de chaque réaction (Orth et al., 2010). Ces contraintes peuvent se traduire sous le programme linéaire global suivant :

$$\begin{aligned} & \text{maximiser/minimiser } f_{obj} \\ & \text{tel que } (S.v)_{int} = 0 \\ & \text{et } v_{i_{min}} \leq v_i \leq v_{i_{max}} \end{aligned} \tag{2.3}$$

où f_{obj} est la fonction objectif, $(S.v)_{int} = 0$ suppose l'état stable du système et $v_{i_{min}} \leq v_i \leq v_{i_{max}}$. (Orth et al., 2010)

Un système est à l'état stable lorsque l'on observe aucune accumulation de métabolites à l'intérieur du système, c'est à dire, que pour un même métabolite m intracellulaire, la quantité de m produit est égale à la quantité de m consommé. Ainsi, le flux de chaque réaction, i.e. l'approximation de la vitesse cinétique, est ajustée de façon à ce qu'il y ait pas d'accumulation dans le milieu intracellulaire.

La fonction objective permet généralement de simuler la croissance d'un organisme en fonction des substrats. La maximiser revient ainsi à orienter les flux métaboliques de telles façon que l'ensemble des composés de la fonction objective ainsi que leurs précurseurs soient produits. En biologie des systèmes, la fonction ~~de~~ objective est une fonction de biomasse, simulant le taux de croissance d'un organisme [ML – ref dans papier FEIST2010344]. Comme chaque organisme ne pousse pas à la même vitesse, définir les composés de la biomasse pour chaque espèce est compliqué. En revanche, on retrouve des familles de composés en commun (Feist and Palsson, 2010).

Basé sur l'optimisation en contraintes, on retrouve des variations telles que l'analyse parcimonieuse de équilibre de flux (pFBA) (Lewis et al., 2010), trouvant une unique solution en minimisant la somme totale des flux tout en maintenant un flux optimal de croissance. L'analyse de la variabilité des flux (FVA) (Mahadevan and Schilling, 2003), qui cherche à évaluer pour chaque réaction, les distributions de flux possibles et ainsi identifier les réactions bloquées, i.e., pour lesquelles il existe pas de solution dans l'espace de solution où cette réaction sera activée. Une extension du problème d'optimisation énoncé plus haut permettant de prendre en compte la machinerie de production des protéines se fait grâce à l'analyse par équilibre des ressources (RBA) (Goelzer et al., 2015). Dans ce dernier, l'hypothèse mis en avant est que chaque cellule utilise les ressources internes de façon économique. Ces méthodes peuvent être utilisé dynamiquement grâce un processus itératif : L'analyse par équilibre dynamique de flux (dFBA) (Mahadevan et al., 2002a). Le processus dynamique se fait à l'aide de système d'équations différentielles ordinaire (ODE), permettant en plus, d'ajouter une couche de modélisation spatio-temporelle (Succurro and Ebenhöh, 2018). A chaque pas de temps δt , définit manuellement, le taux de croissance de chaque espèce et les concentrations des composés à suivre sont calculé grâce au FBA.

2.2.1.3 les modes de flux élémentaires

Une méthode dérivée du calcul du flux, appelée EFM pour *elementary flux mode*, est celui des ensembles minimaux et uniques de cascades réactionnelles (super voies métaboliques) permettant de satisfaire la contrainte d'équilibre du système biologique (SCHUSTER and HILGETAG, 1994; Trinh et al., 2009). Ce type de méthode a un fort coût calculatoire dû à sa grande combinatoire et est donc difficilement applicable à l'échelle du génome. Plus récemment, une méthode combinant l'answer set programming (ASP) (Lifschitz, 2008), connu pour résoudre les problèmes combinatoire, et les EFM sont utilisés sur le coeur du métabolisme de *E.Coli* permettant de réduire le coût calculatoire en intégrant différentes contraintes linéaires (Mahout et al., 2020).

2.2.2 Méthodes qualitatives d'analyse d'un réseau métabolique

La modélisation qualitative du métabolisme vise l'émergence de propriétés topologiques et fonctionnelles globales (Steuer, 2007). Ces méthodes peuvent se résumer à partir d'une analyse basée sur les graphes utilisant différentes représentation du métabolisme. D'autres méthodes se basent sur une autre représentation d'un graphe, et utilise des règles logiques permettant l'analyse de ce dernier : raisonnement. [ML – Steuer parle de stabilité]

2.2.2.1 Graphes

Comme les graphes sont des objets mathématiques, une analyse plus poussé que représenter la connectivité des nœuds peut-être ainsi faite, comme par exemple, trouver les ensemble de voies métaboliques à partir d'un ensemble de métabolites sources et de métabolites cibles, ou uniquement d'un ensemble de métabolite source ou uniquement à partir d'un ensemble de métabolites cibles(Heiden et al., 2002). Kuffner (R et al., 2000) et ses collègues démontrent que l'analyse des chemins métabolique à partir d'une paire source-cible est un non sens biologique au vue du nombre de chemins trouvés (plus de cinq cent mille). Une autre approche est de contraindre les voies les métaboliques par des données *a priori*, comme par exemple les données d'expression de gènes (Zien et al., 2000). Enfin, une analyse du plus court chemin peut être faite.

2.2.2.2 Logiques

2.2.2.2.1 Réseaux booléen permettent de révéler des mécanismes de régulation interne au système ou encore d'identifier les interventions de cible, dans le cas de cibles thérapeutiques. Brièvement, l'analyse se fait à l'aide de règles booléennes modélisant un comportement souhaité (voir (Schwab et al., 2020) pour une revue des différentes applications des réseaux booléens)

2.2.2.2.2 Problème SAT regarde la satisfaisabilité du problème, c'est à dire, vérifier qu'il existe une valeur pour chaque variable booléenne qui satisfait l'expression logique. Par exemple, dans la proposition booléenne 2.5, aucune valeur ne peut être assignée à x_1 qui puisse vérifier l'expression booléenne. En revanche, au sein de l'expression booléenne 2.4, il existe un nombre infini de solution comme $x_1 = 1$ et $x_2 = 0$ qui satisfont le problème logique. On retrouve les travaux de Corblin et al (Corblin et al., 2006) qui appliquent une méthode basée sur SAT pour déchiffrer des réseaux de régulation de gènes.

$$x_1 \wedge \neg x_2 \tag{2.4}$$

$$x_1 \wedge \neg x_1 \tag{2.5}$$

2.2.2.2.3 Extension de réseaux métabolique Une autre façon de modéliser le métabolisme est d'utiliser la notion de productibilité d'un composé (voir proposition logique ??). En effet, basé sur le principe de la représentation logique de la connaissance et du raisonnement (Levesque, 1986), les travaux d'Olivier Ebenhöh et al (Ebenhöh et al., 2004) font une abstraction booléenne du métabolisme à partir de règles logique. Une extension a été développé permettant d'introduire la notion d'ensemble putatif (Handorf et al., 2005). Ces concepts sont alors utilisés pour la modélisation du métabolisme à l'échelle de

génome (Aite et al., 2018 ; Belcour et al., 2020) dans le but de trouver un ensemble de voies métaboliques permettant la production de métabolites cibles. L'application de cette méthode de raisonnement nécessite, dans un premier temps, de créer une liste de faits adéquats représentant la connaissance, ici le réseau métabolique. Par exemple, le réseau métabolique est composé d'objets réactions, gènes, et métabolites, ainsi que des relations entre ces divers objets (gènes associés aux réactions, métabolites est un produit de la réaction etc). Un vocabulaire adéquat décrivant le réseau métabolique, c'est à dire la liste de faits, peut être constitué entre autre de $Produit(metabolite_id_1, reaction_id_1)$ qui signifie que le $metabolite_id_1$ est un $Produit$ de la réaction $reaction_id_1$.

Un fois que le vocabulaire a été créé, les règles logiques permettant d'analyser peuvent être inférée, notamment, celle permettant d'obtenir l'ensemble des métabolites productibles à partir d'un ensemble de nutriments, appelé graine. Elle est composée de deux règles : une d'initialisation, stipulant que la productibilité est initié par la présence de graine (voir règle 2.6), et une récursive explicitant que les produits d'une réactions sont productible si tous les réactants de cette réaction sont eux mêmes productibles (voir règle 2.7). Au vue de cette définition, la notion de productibilité d'un composé à l'échelle d'un réseau métabolique à l'échelle du génome peu être décrit ainsi :

$$productibilité(Composé) : -graine(Composé). \quad (2.6)$$

$$\begin{aligned} productibilité(Composé) : & -GEM(G), \\ & produit(Composé, Reaction, G), \\ & reaction(Reaction, G) \\ productibilité(new_Composé) \forall reactant(&new_composé, Reaction, G). \end{aligned} \quad (2.7)$$

2.2.2.2.4 Programmation par ensemble de réponse La méthode par raisonnement peut être combinée avec la programmation par ensemble de réponse (ASP) (Lifschitz, 2008), une programmation déclarative permettant la résolution de haut problème combinatoire. L'ASP est dites non monotone, c'est à dire, par défaut on considère que la proposition logique est fausse, même en cas de manque d'information. L'ASP décrit le problème en une liste de faits, de règles, de contraintes et d'optimisation. A l'aide d'un solveur logique, les variables vont être instanciées et le problème résolut satisfait ainsi l'ensemble des contraintes, règles et optimisations préalablement décrites. Par exemple, considérons un atome $x1$. Si $x1$ est vrai alors $x2$ est aussi. Cela se traduit par :

$$\begin{array}{l} x1. \\ x2 : -x1. \end{array} \quad (2.8)$$

L'ensemble de réponse associé a cette proposition logique est composé de $x1$ et $x2$, car $x1$ et $x2$ sont vrais.

A l'inverse, la proposition logique 2.9 stipulant que si $x1$ et $x3$ sont vrais, alors $x2$ l'est aussi donnera une réponse différente. En effet, $x3$ n'est pas connu, le solveur va donc considérer que le $x3$ est faux, par conséquent, $x2 : -x1, x3$. est fausse et donc l'ensemble de réponses associé a cette proposition est l'ensemble composé de $x1$.

$$\begin{aligned} & x_1. \\ x_2 : & -x_1, x_3. \end{aligned} \tag{2.9}$$

2.2.3 Défis rencontrés lors de la modélisation qualitative et numérique d'une cellule

Pour analyser une cellule du point de vue du métabolisme, des choix de modélisation doivent être pris en compte. Tout d'abord, les réactions transports permettant le passage de métabolites du milieu intracellulaire au milieu extracellulaire. Pour les deux types de modélisation, des réactions de transports ont du être ajoutées aux modèles métaboliques prenant en compte les différents éléments du milieu de culture. Un second défi rencontré est la prise en compte des systèmes de régulations des co-facteurs. Dans les deux cas, ils sont considérés comme disponibles car ils sont dépendant de cycles de régulations complexes et potentiellement thermodynamiquement non faisable, nécessitant une curation des modèles, ce qui aurait pour conséquence de réduire les réponses possibles et de manquer des vrais positifs.

Dans le but de comprendre le comportement d'un écosystème composé de plusieurs centaines de micro-organismes, étudier le métabolisme de chaque espèce ne semble pas être opportun. Il existe ainsi des méthodes à l'échelle de l'écosystème analysant des communautés de microorganisme. Dans cette section, vous trouverez une liste de méthodes, avec leurs avantages, leurs limites et leurs applications capable d'analyser un écosystème bactérien.

2.2.4 Méthodes hybrides à l'échelle d'un modèle métabolique

La modélisation est dite hybride lorsqu'une variable rassemble différentes informations issues deux formalismes mathématiques ou informatiques comme par exemple discret et continu (Cruz and Kemp, 2022). Il existe trois grands types de modèle hybride : (i) indépendant, (ii) couple ou (iii) imbriqué (Stéphanou and Volpert, 2016). (i) consiste en au moins deux modèles décrivant le même phénomène à des échelles de précision différentes et où aucune interaction entre les modèles existe. Puis vient le plus utilisé (ii) qui consiste en au moins deux modèles pouvant communiquer unidirectionnellement ou bidirectionnellement entre eux. Enfin, le modèle hybride imbriqué forme une seule entité composés de sous-parties sensible au changement du contexte, déclenché par un seuillage par exemple. Il ne sera pas détaillé dans cette section.

2.2.4.1 Indépendants

Au sein des modèles indépendants, c'est l'approche qui est qualifié d'hybride et non le modèle en lui-même. Les réseaux métaboliques possèdent la propriété d'être décomposable (Fisher and Piterman, 2010), c'est à dire qu'ils peuvent être décomposés en plusieurs petites et plus simples mécanismes qui sont mieux définis, tels qu'une réaction responsable de l'activation d'une chaîne réactionnelle impliquant la production d'un composé connu. On peut imaginer une approche de reconstruction de réseaux métaboliques hybride, utilisant comme premier modèle une reconstruction bottom-up et de l'autre, une

reconstruction top-down. Ce type d'approche est utile pour décrire un même phénomène à des échelles différentes permettant ainsi d'identifier lequel des modèles utilisés est le plus précis et le moins coûteux computationnellement.

2.2.4.2 Couplé

Dans le cadre des modèles couplés, on considère le modèle général comme hybride où les variables de sortie d'un modèle servent en entrée d'un autre. Ce type de modèle hybride sert principalement à traiter des problèmes à plusieurs échelles (Brook and Waters, 2008). Dans le cas des modèles hybrides appliquer à la biologie des systèmes, Frioux et al (Frioux et al., 2019) ont développé une méthode de complétion de réseaux de métabolique hybride basé sur l'ASP. L'unique modèle ASP combine à la fois le solver logique Clingo (Gebser et al., 2019) et le solver linéaire Clingo[LP] (Janhunen et al., 2017) résolvant respectivement les contraintes logiques et linéaires du programme. Les modèles trouvés satisfont ainsi les contraintes linéaires à partir des contraintes logiques données en entrée. Dans le cas où les contraintes linéaires ne sont pas satisfaites, une communication bidirectionnelle entre les deux solvers a lieu mettant en avant une possible incohérence.

2.2.4.3 Inference temporelle dans la programmation logique

Nous avons vu que les modèles numériques, contrairement aux modèles logiques, pouvaient intégrer une dimension temporelle et ainsi analyser d'autres propriétés du système biologique. Il existe un formalisme informatique, nommé *logique temporelle linéaire* ou LTL, qui permet d'introduire une notion de temporalité au sein des modèles logiques. Dans un monde logique, toutes les propositions sont évaluées, et peuvent être combinées à des connecteurs logiques tels que ET ou NOT. D'un point de vue de la modélisation, dire que la proposition logique ne varie pas au cours du temps est une forte approximation. Il existe dans un monde logique une extension permettant de moduler la notion temporelle d'une proposition à l'aide d'un vocabulaire contrôlé établissant les événements passés et futurs. On peut ainsi établir des contraintes temporelles permettant la description de phénomènes non stables dans le temps. On peut ainsi définir qu'une proposition logique fausse à un instant t , modélisant un processus biologique par exemple, deviendra vrai dans le futur à un moment donné. [ML – Faire une table qui récapitule ce vocabulaire]. Le problème est décrit en 3 parties qui sont l'ensemble des règles initiales, les règles dynamiques puis les règles finales (Cabalar). Dans l'exemple ci-dessous, on définit une proposition vraie tout le temps : Quelque chose est *chargé* si **au temps précédent** il a été *chargé* et qu'il n'est pas *déchargé*.

$$\square(\bullet loaded \wedge \neg unloaded \rightarrow loaded) \quad (2.10)$$

2.3 Analyse du métabolisme les écosystèmes microbiens

2.3.1 Méthodes numériques appliquées à plusieurs réseaux métaboliques

2.3.1.1 Equation différentielles ordinaires (ODE)

Un moyen pour décrire la dynamique spatio-temporelle d'une communauté microbienne est d'utiliser les équations différentielles (Succurro and Ebenhöh, 2018). In 1838, Verhulst a défini le premier modèle décrivant la croissance d'une bactérie dans une culture au cours du temps (F., 1838). Par la suite, les modèles proie-predateurs ont été utilisés pour décrire une interaction, positive ou négative, d'une espèce prédatrice sur la croissance d'une autre en communauté (Nair et al., 2019). On peut également retrouver les modèles de Lokta and Volterra, largement utilisés pour étudier l'évolution dans le temps d'un écosystème et principalement, les modèles généralisés de Lokta-Volterra, prenant en compte les données d'abondance relative de bactéries lors d'une analyse en métagénomique (Berry and Widder, 2014). En revanche, ces modèles ne se basent que sur les interactions trophiques entre les bactéries et ne prennent pas en compte l'environnement nutritionnel. Il existe cependant une extension du modèle de Lokta-Volterra généralisé permettant de tenir compte d'une potentielle perturbation environnemental sur le comportement d'un individu (Stein et al., 2013). Toutes ces méthodes sont malheureusement "data-driven", i.e. dépendant des données et de la connaissance que l'on peut avoir sur le système biologique d'étude. Ainsi pour permettre une analyse détaillée du système biologique permettant de répondre aux questions, qui produit quoi ? comment une perturbation dans la composition bactérienne ou dans les milieux nutritifs impacte la communauté bactérienne ? il faut l'inferer cette connaissance sous forme de paramètre. En d'autres termes, chaque mécanisme de régulation que l'on souhaite modéliser doit être décrite par, *a minima*, une équation.

2.3.1.2 Analyse statique

Les modèles de flux sont étendus à l'échelle de la communauté sous l'hypothèse d'état stable que l'on a vu précédemment. Le modèle de communauté peut se décrire de trois façons possibles : chaque espèce au sein de la communauté maximise sa propre biomasse et ainsi, la biomasse communautaire consiste en une somme des biomasse de chaque organisme, ou bien, une fonction de biomasse communautaire est créée en assemblant les compositions de chaque fonction de biomasse des organismes, ou enfin, une optimisation sur deux niveaux, une à l'échelle de l'individu et l'autre à l'échelle de la communauté.

En fonction du type de représentation du modèle de communauté il existe plusieurs outils (Colarusso et al., 2021). Dans Henry et al (Henry et al., 2016), les auteurs ont choisi de maximiser une fonction de biomasse communautaire se basant sur la composition des réactions de biomasse de chaque taxa présent dans la communauté. Microbiome modeling toolbox (Baldini et al., 2019), MiCOM (Diener et al., 2020) et SteadyCom (Chan et al., 2017) maximisent tous la somme des flux de la fonction de biomasse de chaque individu et permettent l'étude de plus grande communautés comparé à la méthode de Koch et al, (Koch et al., 2016) utilisant le FBA sur une communauté de taille 3 ou encore cFBA (Khanelwal et al., 2013). Enfin, dans la catégorie d'une optimisation à plusieurs niveaux, on peut retrouver les outils suivants : OptCom (Zomorrodi and Maranas, 2012), NECom

(Cai et al., 2020), maximisant chaque fonction de biomasse puis les flux de production de biomasse à l'échelle de la communauté. Pour des raisons de coût calculatoire, appliquer ces outils sur des grandes communautés microbiennes est non faisable.

Toutes ces méthodes basées sur des contraintes permettent d'étudier le métabolisme et par conséquent peuvent être utilisé pour mettre en évidence l'impact positif (coopération) ou négatif (compétition) d'une interactions bactérienne. En effet, dans (Freilich et al., 2011) les scores de potentiels de coopération et de compétition sont inférés en caractérisant les relations de donneurs-receveurs et de gagnants-perdants au sein de l'écosystème. Plus récemment, MMint (Mendes-Soares et al., 2016) prédit l'impact positif ou négatif d'une espèce sur une autre en comparant les croissances des espèces seules et lorsqu'elles sont en interactions deux à deux. L'analyse des interactions deux à deux ne permet pas de déduire l'ensemble des comportements de la communauté bactérienne puisque de nouvelles interactions se produisent, disparaissent ou encore se conservent (Morin et al., 2022). SMETANA (Zelezniak et al., 2015) et MiCOM (Diener et al., 2020) contournent cette limite et proposent d'étudier les interactions bactériennes de l'ensemble des espèces de la communauté. SMETANA définit la coopération via le potentiel d'interaction métabolique (MIP) représentant le nombre minimal de métabolites produit par une espèce pouvant être réutilisé pour la croissance d'une autre. La compétition est alors estimé par le maximum de chevauchement de métabolites (MRO) au sein du milieu minimum pour la croissance de toutes les espèces. MiCOM ne permet pas d'obtenir une mesure de la coopération et de la compétition mais, tout comme SMETANA, donne une liste de métabolites potentiellement échangés avec en plus le flux d'export et d'import.

2.3.1.3 Intégration de données multi-omiques

Ces méthodes d'optimisation sous contraintes permettent de mettre en évidence des interactions entre bactéries, ou révèlent des mécanismes de régulation de la communautés seulement à partir de données génomiques. Il existe des méthodes d'intégration de données multi-omiques, notamment de métatranscriptomiques, qui ajoutent des contraintes linéaires et donnant des informations supplémentaires sur différents processus de régulation ou encore sur l'état d'activation d'une réaction ou non. Une liste non exhaustive est présentée ci-dessous permettant l'intégration de données omiques : GIMME (Becker and Palsson, 2008), créant un modèle métabolique consistant avec les données biologiques et la fonction objective à maximiser, en minimisant les réactions trouvé inactive par l'outil. INIT (Agren et al., 2012), à partir de données protéomiques ou transcriptomiques, il maximise l'utilisation des réactions basé sur un score de confiance et minimise les réactions où aucune protéine est associée. Plus récemment, RIPTiDe (Jenior et al., 2020) qui utilise une approche parcimonieuse pour refléter au mieux la transcription de la cellule.

2.3.1.4 Analyse dynamique

Au sein d'un écosystème, la composition bactérienne, la quantité de ressource disponible et l'espace utilisé varient au cours du temps. On retrouve la méthode DMMM (Zhuang et al., 2011) qui utilise les équations différentielles ordinaires pour décrire la loi de conservation de la masse pour tous les métabolites échangés et la production de biomasse de chaque espèce au sein de la communauté. Le temps est alors discrétisé, et à chaque pas de temps, une optimisation est faite. Au sein de cette méthode, la concen-

tration des métabolites disponibles dans le milieu extra-cellulaire est alors ajustée après la consommation et la production des différents métabolites échangés ainsi que la densité cellulaire (méthode explicit Euler). Avec ce choix temporel d'ajuster la concentration des métabolites extra-cellulaires après l'importation des nutriments, le modèle peut prédirer que les espèces consomment un nutriment dont sa concentration n'est pas suffisante par rapport la densité cellulaire prédit par le modèle. Une première possibilité consiste à rendre indisponible tous les métabolites dont la concentration est négative. Une seconde possibilité revient à diminuer l'intervalle de temps entre deux calculs, en revanche, cela augmenterait le temps calculatoire. *μbialSim* (Popp and Centler, 2020) solutionne ce problème en réduisant le pas de temps seulement lorsque la concentration d'un métabolite échangé est négative.

Les modèles dynamiques de communautés doivent prendre en compte le cas où il n'y a pas de solutions possibles pour le système dynamique étudié, on parle de résolution non faisable, ou bien, lorsqu'il y a pas un flux unique de production ou de consommation pour un métabolite. La question de faisabilité peut s'expliquer notamment lorsqu'il y a pas une quantité de nutriments limitants suffisants. DFBALab (Gomez and Barton, 2018) tente de résoudre ces deux problèmes en priorisant le choix des métabolites d'intérêts à maximiser. Ainsi, une unique solution est trouvée par le solveur linéaire permettant d'outrerpasser la seconde contraintes. En revanche, l'utilisateur ou l'utilisatrice doit donner une liste de métabolites échangés en amont pour que le résultat de la simulation reste cohérent. En plus, cela suppose une bonne connaissance du système biologique d'étude pour effectuer ce choix et est donc peu recommandé pour l'étude de communauté à grande échelle. D'autres méthodes existent utilisant une méthode de résolution mathématiques différentes (Schroeder and Saha, 2020) ou encore, un optimisation à deux niveaux (Zomorodi).

Une des limites fondamentales de ces utilisations est que tous les échanges métaboliques sont conditionnés par l'optimisation de la fonction de biomasse. En effet, lors d'une maximisation d'une fonction de biomasse, tous les flux sont calculés et orientés pour la production de biomasse, ainsi, un ensemble de métabolites pouvant être échangés au sein d'une communauté mais n'étant pas essentiel pour la biomasse d'un individu ne sera donc pas identifié. Dans le papier de Zomorodi et al (Zomorodi and Segrè, 2017), ils attaquent ce problème en ajoutant des contraintes supplémentaires qui pré-traitent différentes stratégies comme la prise en compte de métabolites échangés coûteux. En revanche, identifier l'ensemble des stratégies à l'échelle d'une communauté ne semble pas être faisable. Au sein de l'outil NECom (Cai et al., 2020), se basant sur une optimisation à deux niveaux, les auteurs relâchent les contraintes de flux acceptant plus de métabolites potentiellement échangés.

2.3.2 Applications

En plus de l'étude des interactions au sein de communautés bactériennes, la création de communautés synthétiques semble être un secteur prometteur. En effet, elle permettrait (1) d'identifier des communautés microbiennes d'intérêt à partir d'une liste de bactéries candidates, (2) d'évaluer l'impact d'une modification génétique sur la synergie de la communauté, et enfin (3), d'évaluer les perturbations dans le milieu de culture ou dans la composition de la biomasse sur la communauté. Dans (Chiu et al., 2014), l'utilisa-

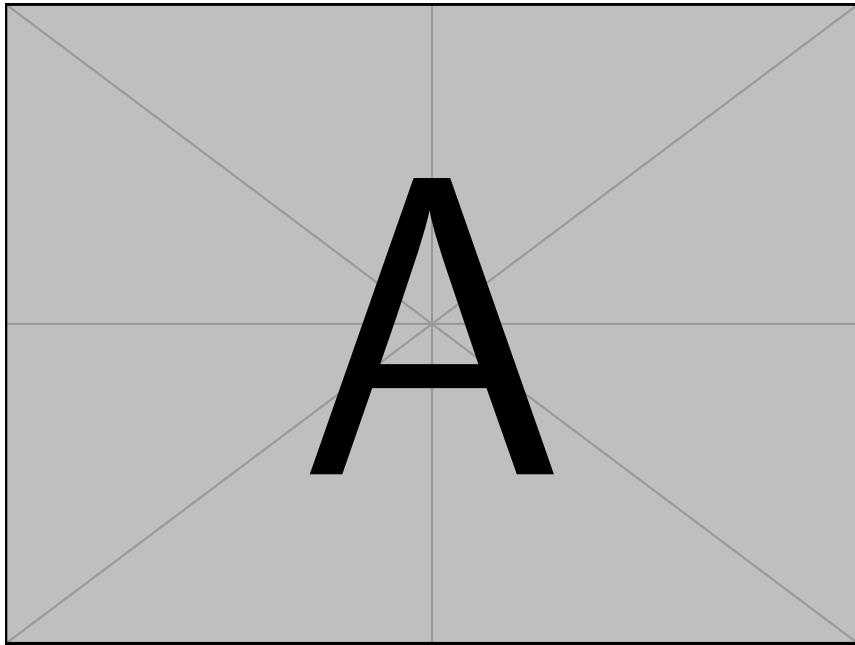


FIGURE 2.3 – Review des méthodes numériques pour une bactéries/communauté avec avantages et inconvénients. Mettre si cela peut répondre à notre défis : utiliser à grande échelle, mettre en avant les interactions etc

tion d'un modèle dynamique a permis de mettre en avant une communauté composée de deux espèces qui ne produisent et échangent des composés uniquement lorsqu'ils sont en co-culture, ou encore, dans (Zuñiga et al., 2020), le modèle dynamique sélectionne des souches facilitant la croissance. D'autre méthode, comme par exemple OptAux, s'intéresse à la création de souches auxotrophes, c'est à dire ne pouvant synthétiser un composé par eux même, qui sont vérifiables expérimentalement (Lloyd et al., 2019). Un autre point de vue d'analyse est celui du milieu de culture, comme par exemple dans (Klitgord and Segrè, 2010), où l'objectif est de sélectionner un milieu de culture permettant d'observer soit des interactions du type mutualisme ou du commensalisme entre des paires d'organismes, soit une production de métabolites cibles (Pacheco and Segrè, 2021). Dans le même esprit, FLYCOP (García-Jiménez et al., 2018) créé un consortium à partir d'un milieu de culture dans le but de trouver tous les échanges métaboliques permettant la production de bio-plastic. Enfin, à ma connaissance, la communauté scientifique a peu étudié *in silico* l'impact d'une perturbation sur la communauté. On retrouve ainsi (Machado et al., 2021), ils étudient *in silico* l'impact de facteurs abiotiques (introduction d'une nouvelle espèce) et biotiques (disponibilité des nutriments) sur la communauté.

Toutes ces méthodes permettant d'analyser des communautés microbiennes sous différents points de vue : identification statique ou dynamique des interactions métaboliques au sein d'une communauté donnée, la recherche d'un milieu de culture pour la création d'une communauté avec le comportement souhaité, la compréhension détaillée de mécanismes métaboliques intra-cellulaire ou encore, l'impact d'une perturbation sur le potentiel métabolique de la communauté. Malgré la précision numérique que l'on obtient, l'utilisation de ces modèles ne permet pas une analyse sur des communautés naturelles du notamment à la combinatoire exponentielle des échanges métaboliques. Il existe des méthodes qualitatives, s'appliquant à l'échelle d'une communauté, sacrifiant la précision numérique pour permettre résolution de ce problème combinatoire sur des communautés naturelles.

2.3.3 Méthodes qualitatives de l'étude d'une communauté bactérienne

Les interactions bactériennes sont au cœur de la régulation des communautés microbiennes. Dans cette sous-section, une liste des méthodes qualitatives sera présentée permettant de révéler un potentiel de compétition et des échanges métaboliques. Pour rappel, un échange métabolique se traduit par la production d'un composé par un microorganisme dans l'environnement nutritionnel qui est lui-même consommé par un autre impactant ou non la dernière (Faust and Raes, 2012). Cette coopération est un mécanisme biologique important et devient au cœur des préoccupations pour mieux comprendre comment une communauté bactérienne fonctionne.

2.3.3.1 Basées sur les graphes

Certaines approches étudient la structure de réseaux métaboliques, sous forme d'un graphe, en déterminant la relation de dépendance du milieu nutritif sur les éléments productibles du graphe. Les éléments identifiés sont ainsi utilisés comme une approximation des métabolites échangés (Levy and Borenstein). Ainsi, en se basant sur Le principe écologie associé, qui l'écologie inversée consistant à déduire des processus écologiques en utilisant les données génomiques massive, il existe des algorithmes permettant de définir des potentiels de coopération et de compétition. Un moyen de déterminer les potentiels de coopération et de compétition au sein d'une communauté bactérienne est de déterminer la composition nutritionnelle (graine) requise pour la communauté. NetSeed propose de trouver un ensemble de graines possibles en se basant à la fois sur le calcul de composantes fortement connexes (SCC) et sur le concept de l'écologie inversée (Carr and Borenstein, 2012). Dans le but de quantifier la complémentarité entre toutes les paires de combinaison, NetCooperate (Levy et al., 2015) mesure à la fois la capacité d'un organisme hôte à répondre aux besoins nutritionnels d'un parasite endosymbiotique (BSS score) et quantifie l'aide mutuelle de deux organismes en fournissant un indice de potentiel syntrophie (MCI score). En se basant sur une analyse du milieu nutritionnel, on peut également déterminer un indice de compétition métabolique en analysant les composés connectés à plusieurs organismes (Levy and Borenstein, 2013). Ainsi, NetCompt estime le potentiel de compétition en calculant le réel ~~de~~ chevauchement métabolique (Kreimer et al., 2012). Toutes ces méthodes sont limitées par une analyse par paire d'organisme ne reflétant pas le réel potentiel coopératif ou compétitif d'une communauté microbienne mais informant l'état coopératif ou compétitif d'un individu sur un autre. Il existe à ce jour, aucun outil qualitatif permettant d'évaluer une communauté dans son ensemble.

2.3.3.2 Basées sur le raisonnement

On a vu dans une sous section précédente que la méthode par raisonnement pouvait identifier des composés potentiellement productible et consommable. Une extension de cet algorithme a été proposé dans MISCOTO (Frioux et al., 2018) qui s'intéresse à l'échelle d'une communauté entière sélectionnant des communautés minimales et en les criblant. Une première approche de mettre en avant des interactions au sein de ce raisonnement est proposé dans (Belcour et al., 2020), avec l'outil Metage2Metabo. Il permet d'identifier la valeur ajoutée de la coopération métabolique au sein d'une communauté au moyen du métabolisme de chaque individu, et identifier et cribler des espèces d'intérêts parmi tous les membres de la communauté.

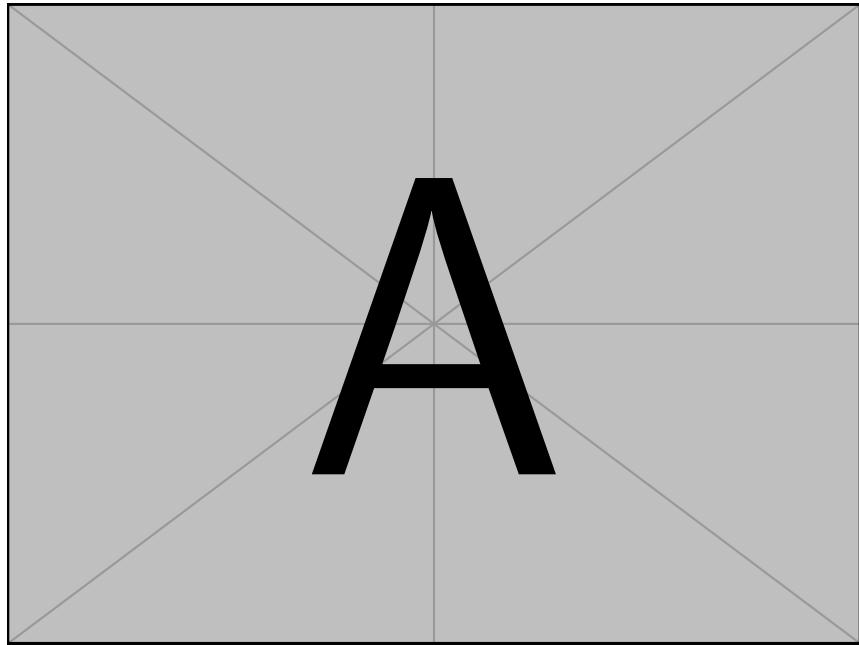


FIGURE 2.4 – Review des méthodes qualitatives pour une bactéries/communauté avec avantages et inconvénients. Mettre si cela peut répondre à notre défis : utiliser à grande échelle, mettre en avant les interactions etc

2.3.4 Défis rencontrés lors de l’analyse de la modélisation qualitative et quantitative du métabolisme

Des questions similaires que celles vu dans la sous section 2.2.3 se posent également au niveau de la communauté. Concernant les réactions de transports, on exige que chaque individu présent dans l’écosystème puisse être capable de consommer tous les nutriments présents dans l’environnement nutritionnel mais tout le monde ne peut pas les produire. Ensuite, concernant la modélisation logique nous avons décidé le postulat suivant : les ensembles de métabolites produits en intracellulaire puisse être disponible pour l’ensemble des membres de la communauté. En effet, nous savons que les cellules sont capables de lyse cellulaire, libérant ainsi tous les produits métaboliques dans le milieu extra-cellulaire (Fazzino et al., 2020) et parce que les sécrétions métaboliques passives, c’est à dire sans coût énergétique pour la cellule, ont été montré comme étant dominante dans les interactions microbiennes (Pacheco et al., 2019). Une seconde notion à prendre en compte est celle de la frontière des compartiments métaboliques propre à chaque espèce de la communauté. Dans nos modèles, numériques et logiques, nous avons considéré qu’un métabolites produit dans le milieu extra-cellulaire peut être disponible pour les autres membres de la communautés.

2.3.5 Modélisation hybride du métabolisme

[ML – Je peux développer les modèles hybrides niveau biblio encore]

Le formalisme des modèles hybrides imbriqués sont fondamentalement et exclusivement mathématiques, ce qui représente une forte contrainte et rend la résolution difficile. Un premier modèle imbriqué, cherchant à modéliser la compétition entre les systèmes immunitaires et tumoral (Cattani and Ciancio, 2012). Ici, Cattani et ses collègues s’appuient ainsi sur un modèle proie-prédateur de Lotka-Volterra avec un paramètre stochastique simulant le temps. Dans la biologie des systèmes, Assar et al (Assar et al., 2014) ont développé un modèle hybride combinant le formalisme numérique et discret. Le modèle est

capable, en fonction de la nature du processus à modéliser et des changements conditions
~~logiques~~, de choisir lequel, du formalisme numérique ou discret, le modèle doit choisir
a fin de modéliser au mieux.

Chapitre 3

TANGO : un jumeau numérique du métabolisme bactérien de la production d'un fromage

Sommaire

3.1	Introduction	31
3.2	Réultats	32
3.2.1	Les modèles métaboliques individuels identifient des voies métaboliques spécifiques dans le lait	32
3.2.2	Optimisation dynamique des modèles individuels	34
3.2.3	Le suivie de la production des métabolites dans les conditions de co-cultures par le modèle de communauté	38
3.2.4	Les interactions bactériennes dynamiques au sein de la communauté	42
3.3	Discussion	46
3.4	Materiels et méthodes	47
3.4.1	Données biologiques	47
3.4.2	Modelisation métabolique	50
3.4.3	Ajustement du modèle	53
3.4.4	Inférence des interactions metaboliques	53
3.4.5	Visualisation et analyses statistiques	53
3.4.6	Code et disponibilité des données	54

Soumis dans metabolic engineering

3.1 Introduction

La fermentation des aliments et des boissons par un consortium bactérien compte parmi les plus vieux processus technologiques alimentaires, associés au bénéfices alimentaires et utilisés aujourd’hui à la fois dans les contextes traditionnels et industriels (Tamang et al., 2016a,b). La transformation de matériel alimentaires brutes par les micro-organismes est le résultat de processus métaboliques and interactions entre les membres de la communauté bactérienne. The contenu de ce dernier peut être contrôlé, comme dans les débuts industriels (Somerville et al., 2021), pour la sécurité alimentaire ou pour la prédictabilité de propriétés organoleptiques (Galimberti et al., 2021). Alors que les propriétés fonctionnelles des souches bactériennes associées aux produits fermentés sont mieux compris (Tamang et al., 2016a), les mécanismes moléculaires précis mis en jeu commencent seulement à être élucidés (Blasche et al., 2021). En particulier, la compréhension du métabolisme individuel des membres d’un consortium n’est pas suffisante pour capturer ou prédire les comportements globaux de la communauté, qui est impacté par les interactions microbe-microbe.

Les approches en biologie des systèmes combinées à l’acquisition de données omiques sont une grande promesse pour caractériser toutes les fonctions et les interactions se déroulant dans le consortium microbien de produits fermentés (Manna et al., 2021). Dans le cas d’une communauté contrôlée, une comparaison en mono et en co-cultures peut nous informer sur l’impact des interactions sur la dynamique and les profils métaboliques de la communauté (Özcan et al., 2020). L’utilisation de modèles métaboliques à l’échelle du génome (GEMs) (Fang et al., 2020) semble particulièrement opportun comme un moyen d’associer les souches du consortium à la fonction harbourer par les systèmes and ultimement aux propriétés organoleptiques alimentaires (Özcan et al., 2020). De tels modèles d’un consortium bactérien peuvent être utiles pour un grand nombre d’applications dans les biotechnologies alimentaires, de l’assemblage de souches dans la conception de communautés, à l’optimisation de procédés pour les propriétés organoleptiques désirées (Rau and Zeidan, 2018).

La nature de la matrice, la complexité de la communauté bactérienne, ainsi que la grande spécificité souches-dépendantes des pathway métaboliques secondaires impliquée dans les processus organoleptiques rendent difficile la mise en place d’un modèle mechanistique ou d’un jumeau numérique de la biologie des systèmes. Ainsi, travailler à petite échelle avec des systèmes contrôlés où la compléxité est réduite est particulièrement pertinent (Rau and Zeidan, 2018). Toutefois, même des systèmes simples peuvent être considérés et modélisés sur le temps, comme la fermentation est intrinsèquement un processus dynamique. Dans ce travail, nous nous sommes intéressés dans la modélisation du métabolisme de la fabrication du fromage, en partant de lait jusqu’à l’étape d’affinage en utilisant une communauté bactérienne utilisée dans le processus de fabrication de fromage industriel. Elle est composée de deux bactéries lactiques (LAB), *Lactococcus lactis*, *Lactiplantibacillus plantarum* et d’une bactérie propionique, *Propionibacterium freudenreichii*. Un travail précédent (Cao et al., 2021) a décrit comment le réglage des paramètres d’un processus standardisé de fabrication d’un fromage peut moduler la production de composés aromatiques. Ici, en utilisant des données omiques, nous avons construit and exploité les réseaux métaboliques et les modèles communautaires dans le but de (i) évaluer la contribution métaboliques de chaque espèce microbienne durant la fabrication de fromage, (ii) de simuler les dynamiques des métabolismes microbiens ainsi que l’effet de la communauté sur le comportement des individus et enfin (iii) proposer des interactions possibles entre les espèces bactériennes.

Notre travail identifie les chemins métaboliques responsables de la transformation des nutriments présent initialement dans le lait. Nous avons optimisé chacun de nos modèles métaboliques individuels et avons été capable de reproduire la dynamique des cultures pures. À partir de ces modèles, nous avons reconstruit un modèle dynamique de la communauté bactérienne qui prédit avec précision la production de composés aromatiques dans le fromage.

3.2 Resultats

3.2.1 Les modèles métaboliques individuels identifient des voies métaboliques spécifiques dans le lait

La communauté bactérienne de départ est composé de *Propionibacterium freudenreichii* CIRM-BIA122, *Lactiplantibacillus plantarum* CIRMB-BIA465 and de *Lactococcus lactis* biovar *diacetyl lactis*. Pour chacun d'entre eux, le longueur total de l'ADN a été récupéré. Les génomes bactériens ont été assemblé à partir d'une combinaison de lectures courtes et longues, produisant un assemblage précis et complète, révélant un chromosome circulaire and la présence de sept plasmides pour la souche de *L. lactis* et quatre plasmides pour la souche de *L. plantarum*. La souche de *P. freudenreichii* porte pas de plasmide. Les genomes annotés peuvent être retrouvés sous le numero d'accesion ENA suivant : ENA : PRJEB42478. Le nombre de gènes codant pour des protéines, de séquences ribosomiques, tRNA et le pourcentage de GC ainsi que le numéro d'accesion sont rassemblés dans la Table 3.1. L'annotation a identifiée une list de 8234 séquences codantes incluant ceux du chromosome (2460 pour *P. freudenreichii*, 2417 for *L. lactis* et 3016 pour *L. plantarum*) et des plasmides (139 pour *L. lactis* et 82 pour *L. plantarum*).

Les réseaux métaboliques à l'échelle du génomes ont été reconstruit pour chaque souche individuelle depuis leur sequence génotiques annotées respective. Après une première étape automatique de reconstruction, nous avons réalisé un raffinement manuel du métabolisme carboné central basé sur la littérature et la connaissance biologique dans le but de prendre en compte la spécificité de chaque souche bactérienn. Les réseaux métaboliques qui en résultent satisfont la qualité attendu dans le domaine : une faible fraction de réactions universellement bloqués (< 3%), et une faible fraction of réaction sans association de gène-proteine-réactions (GPR) (< 30%) (Lieven et al., 2020). La Table 3.2 décrit les caractéristiques des trois modèles. Nous observons que *L. lactis* a le plus petit nombre de réactions que les deux autres souches, une différence qualitative qui est aussi observée pour les autres souches de la même espèces dans d'autres bases de données (Noronha et al., 2018).

Nous avons construit les réseaux métaboliques à l'échelle de chaque génomes (GEMs) depuis les trois reconstructions métaboliques en utilisant l'analyse par équilibre des flux (FBA) (Orth et al., 2010 ; Palsson and Varma, 1994) assurant la production des molécules de la biomasse sur un environnement nutritionnel reflétant la composition du lait (voir Matériel et méthodes). Nous avons obtenu des taux de croissances allant de 0.187 mmol.gDW⁻¹.hr⁻¹ à 2.17 mmol.gDW⁻¹.hr⁻¹, où le premier est obtenu par *P. freudenreichii* et la plus haute valeur obtenu par *L. lactis*. Une attention particulière a été donnée sur le rôle de l'oxygène pour les 3 métabolismes. Depuis que les croissances bactérienne sur le fromage sur du fromage correspondent à des conditions microaérobique (Miyoshi et al.,

2003), nous avons modifié la borne supérieur de la réaction de transport de l'oxygène pour limiter l'import de l'oxygène dans le compartiment cytosolique.

Le métabolisme carboné central produit du NADH and NADPH ainsi que les formes réduites, l'ATP and les précurseurs nécessaires pour la biosynthèse des molécules essentielles pour la croissance et l'activité métabolique des cellules bactériennes, e.g. acides aminés, purine, pyrimidine, glycerol-3-phosphate, acides gras, N-acetyl-glucosamine, vitamines. Les trois espèces étudiées partagent presque la même glycolyse and la voie des pentose phosphate. Les sources de carbones, dans ce cas le lactose, l'acide citrique et lactique du lait, sont converties en acide pyruvique, qui est par la suite métabolisé *via* des voies métaboliques spécifiques. Le métabolisme carboné central spécifique à chaque souche microbienne est illustré par la Figure 3.1. Nous décrivons par la suite les caractéristiques de chaque souche métaboliques avec les principaux raffinement métaboliques performés pour chaque GEMs, où une attention particulière est donnée sur les voies métaboliques secondaires permettant la production de composés organoleptiques.

3.2.1.1 Métabolisme des bactéries lactiques

A la fois *L. lactis* et *L. plantarum* utilisent le lactose et le citrate, les deux principales sources de carbone du lait (Widyastuti et al., 2014). Le lactose est dégradé par la beta-galactosidase en glucose and galactose. Le glucose nourrit la glycolyse, tandis que le métabolisme du galactose et de l'acide citrique (Palles et al., 1998) est souche dépendant. Le galactose est convertie en glucose-6-phosphate *via* la voie métabolique de Leloir avant d'entrer dans la glycolyse. L'acide citrique est convertie en pyruvate par l'enzyme multimérique citrate lysase à travers la production des acides lactiques et citriques (Roissart, 1994). La production du butanediol et du diacetyl, responsable de la saveur beurré, apparaît dans chacun des deux modèles à travers la voie métabolique de l'acétolactate depuis la transformation du pyruvate. Toutes ces voies métaboliques sont présentes et fonctionnelles dans le modèle (Fig. 3.1 a, b).

3.2.1.1.1 *L. lactis*

La production du butanediol a requis une curation du GEMs(Fig. 3.1 a). Le flux de l'acetoin-dehydrogenase a été bloqué (ACTD2) permettant la production du butanediol, et les bornes de l'acétolactate décarboxylase (ACLDC), acétolactate synthase (ACLS) ont été modifiée permettant l'activation de la voie de l'acétolactate, comme retrouvé dans la littérature (Carroll et al., 1999 ; Makhlof, 2006 ; Swindell et al., 1996). Enfin, la consommation de lactose a été régulée en modifiant les bornes de la réaction LACZ qui a permis un flux de production de lactate pour *L. lactis*(Fig. 3.1 a).

Pour valider ce modèle, nous avons considéré un exemple d'activation de voie métabolique durant la production du fromage. *L. lactis* consomme le lactose à la fois par la voie du tagatose, encodée par l'opéron plasmidique *lacABCD*, ou par la voie de Leloir, encodée par l'opéron du chromosome *galKTE*. Les données metatranscriptomiques montrent que *L. lactis* exprime les gènes associés à la voie du tagatose au début de la production de fromage quand la concentration de lactose est grande, and ceux de la voie de Leloir durant l'affinage quand la concentration de lactose est faible. Cela suggère fortement que la concentration de lactose détermine quelle voie métabolique est employée, et prédit que une grande quantité de lactose devrait produire un grand de lactose à travers la voie du tagatose. Le modèle FBA de *L. lactis* mène à un plus grand flux de lactose à travers la voie du tagatose lors de l'inoculation et ainsi, coincide avec les attentes (Fig. 3.2 a).

3.2.1.1.2 *L. plantarum* La dégradation du glucose et du galactose mène à la production du pyruvate qui contribue au relargage de l'acide lactique dans le cas d'une fermentation homolactique or de l'éthanol, acetate ou lactate dans le cas d'une fermentation heterolactique. Comme *L. plantarum* est une bactérie heterolactique, l'enzyme spécifique de la transketolase est impliquée (4.1.2.9), décrit précédemment dans (Abedi and Hashemi, 2020), et ainsi, l'acide acétique peut être produit par la voie de l'acetate kinase. La production du butanediol a requis la même curation que pour *L. lactis*. La conversion du pyruvate en acide acétique a été permis en bloquant la pyruvate formate lyase (PFL) (Quatravaux et al., 2006). Pour reproduire la dégradation du galactose et du glucose à travers les voies de Leloir et de la glycolyse par la β -galactosidase *LACZ*, la réaction *LCTSt3ipp* fut bloqué. Pour compléter la dégradation du galactose, la réaction UTP-glucose-1-phosphate uridylyltransferase (*GALUi*) est rendu reversible.

3.2.1.1.3 Le métabolisme de *P. freudenreichii* La souche CIRM-BIA122 de *P. freudenreichii* consomme à la fois l'acide lactique, sa principale source de carbone (Dank et al., 2021 ; Thierry et al., 2011). L'acide lactique et le lactose sont catabolisés à travers les voies de la glycolyse et des pentose-phosphate, menant à la production de l'acide pyruvique. Une part du pyruvate alimente le cycle des tricarboxylique (TCA) et la voie de Wood-Werkman (Fig. 3.1 c). La production de l'acide propionique, spécifique au genre *Propionibacterium*, est assurée en routant le succinate à partir du TCA. La production de l'acid acétique et du CO₂ (Turgay et al., 2020) à partir du pyruvate est permis grâce à l'activité de deux enzymes : la pyruvate déhydrogénase et la décarboxylative oxidation. Assurer la proportion correct de propionate-sur-acétate a requis de plus amples curation du GEM. Les bornes des réactions *PPAKr*; *PTAr*; *2131pyrpp* ont été modifié pour diriger la production du propionate dans sa voie métabolique de prédilection, i.e. Wood-Werkman, et la réaction *POX2* fut bloqué pour réguler le flux d'acétate.

Pour valider le modèle, nous avons calculé le ratio de propionate sur l'acétate et obtenu la valeur de 2.19, ce qui est confirmé par les observations biologiques de 1:2 (Crow, 1986 ; Dank et al., 2021 ; Turgay et al., 2020 ; ?).

Après raffinement, les modèles FBA des trois espèces ont prédit la production de plusieurs composés aromatiques tels que les acides lactiques, acétiques et propioniques, le diacetyl and butanediol (Smid and Kleerebezem, 2014 ; ?) (Fig. 3.1). Ils ont en plus reproduit avec précision les spécificités métaboliques de chaque souche en fonction de leur métabolisme primaire, i.e. la voie de Wood-Werkman pour *P. freudenreichii*, la voie du tagatose pour *L. lactis* et la voie de la transketolase pour *L. plantarum*.

3.2.2 Optimisation dynamique des modèles individuels

Chaque souche microbienne fut cultivé en culture pue : monitorant le pH pour les LAB, et les croissances, pour toutes les souches, performé en lait sous conditions microaerophilique (Supp. Table S1). Pour imiter la disponibilité du lactate en co-culture, le milieu de culture a été supplié en lactate et un hydrolyzat de protéine du lait simulant la protéolyse pour *P. freudenreichii*. Sous ces conditions, toutes les souches ont atteint un seuil de culturabilité au dessus de 9 log₁₀ CFU/g, et le pH, inoculé à 6.7, atteint 5.7 et 5.1 pour les cultures respectives de *L. plantarum* et *L. lactis*. Le plateau de culturabilité correspond à l'entrée dans la phase stationnaire et a été atteint en 48 heures pour *P. freudenreichii*, 14 heures pour *L. plantarum* et 6 heures pour *L. lactis*. Nous notons que les modèles

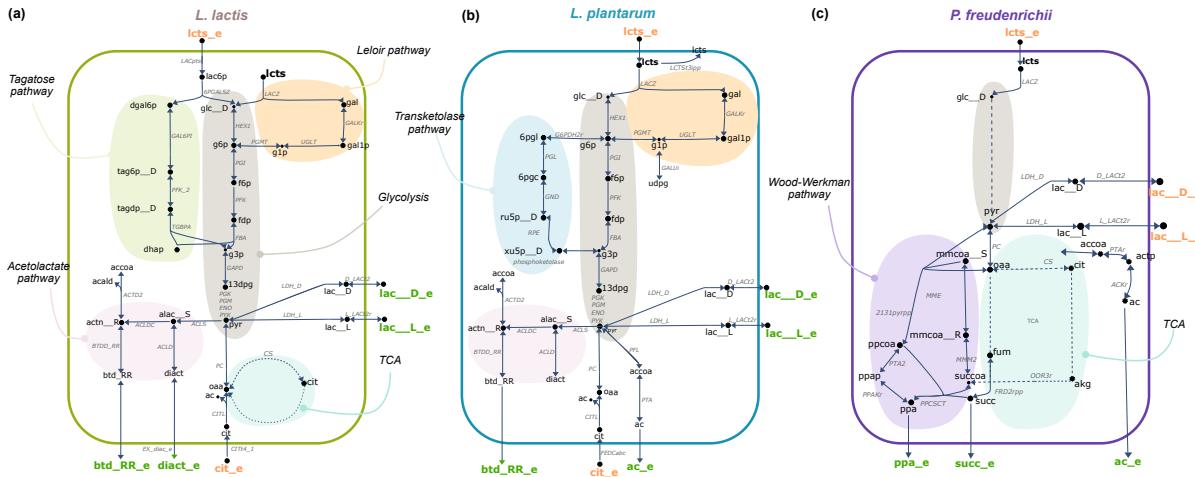


FIGURE 3.1 – Metabolic maps of specific and shared pathways for *L. lactis*(a), *L. plantarum*(b) and *P. freudenreichii*(c). The glycolysis pathway (grey) is shared by the three species, while Leloir (salmon), acetolactate (pink) and citrate degradation pathways are shared by the two lactic acid bacteria. The TCA cycle (light blue) is shared by *P. freudenreichii* and *L. lactis*. Tagatose (green), transketolase (blue) and Wood-Werkman (purple) pathways are specific to *L. lactis*, *L. plantarum* and *P. freudenreichii* respectively. Orange and green metabolites are respectively inputs and outputs of the metabolic models. *Compound Abbreviations* : 13dpg : 3-Phospho-D-glyceroyl phosphate; 6pgc : 6-Phospho-D-gluconate; 6pgl : 6-phospho-D-glucono-1,5-lactone; ac : acetate; acald : acetaldehyde; accoa : acetyl-coa; actn__R : acetoine; akg : 2-Oxoglutarate; alac__S : (S)-2-Acetylacetate; btd__RR : butanediol; cit : citrate; dgal6p : d-galactose-6-phosphate; dhap : Dihydroxyacetone phosphate; diact : diacetyl; f6p : fructose-6-phosphate; fdp : D-Fructose 1,6-bisphosphate; fum : fumarate; g1p : glucose-1-phosphate; g3p : Glyceraldehyde 3-phosphate; g6p : glucose-6-phosphate; gal : galactose; gal1p : galactose-1-phosphate; glc__D : glucose; lac__D : lactate; lac6p : lactose-6-phosphate; lcts : lactose; mmcoa__R : (R)-Methylmalonyl-CoA; mmcoa__S : (S)-Methylmalonyl-CoA; oaa : Oxaloacetate; ppa : propionate.; ppap : Propanoyl phosphate; ppcoa : propionyl-coa; pyr : pyruvate; ru5p__D : D-Ribulose 5-phosphate; succ : succinate; succo : succinyl-coa; tag6p__D : D-tagatose-6-phosphate; tagdp__D : D-Tagatose 1,6-biphosphate; udpg : UDPglucose; xu5p__D : D-Xylose 5-phosphate. *Reaction Abbreviations* : 2131pyrpp : Methylmalonyl-CoA carboxyltransferase 5S subunit; 6PGALSZ : 6-phospho-beta-galactosidase; ACCL : Acetolactate decarboxylase; ACTD2 : Acetoin dehydrogenase; ACLDC : Acetolactate decarboxylase; ACSL : Acetolactate synthase; BTDD__RR : R R butanediol dehydrogenase; CITL : Citrate lyase; CIT4_1 : Citrate transport via sodium symport; CS : Citrate synthase; D_LACt2 : D lactate transport via proton symport; ENO : Enolase; FBA : Fructose-bisphosphate aldolase; FEDCabc : FEDCabc; FRD2rpp : Fumarate reductase / succinate dehydrogenase (irreversible) (periplasmic, membrane potential dissipating); G6PDH2r : Glucose 6-phosphate dehydrogenase; GALkr : Galactokinase; GALU1 : UTP-glucose-1-phosphate uridylyltransferase (irreversible); GAPD : Glyceraldehyde-3-phosphate dehydrogenase; GND : Phosphogluconate dehydrogenase; HEX1 : Hexokinase (D-glucose:ATP); L_LACt2r : L lactate reversible transport via proton symport; LACpts : Lactose transport via PEP:Pyr PTS; LACZ : B-galactosidase; LCTSt3ipp : Lactose transport via proton antiport (periplasm); LDH_D : D-lactate dehydrogenase; LDH_L : L-lactate dehydrogenase; MME : Methylmalonyl-CoA epimerase; MM2 : Methylmalonyl-CoA mutase; OOR3r : 2-oxoglutarate synthase (rev); PC : Pyruvate carboxylase; PFK : Phosphofructokinase; PFK_2 : Phosphofructokinase; PFL : Pyruvate formate lyase; PGI : Glucose-6-phosphate isomerase; PGK : Phosphoglycerate kinase; PGL : 6-phosphogluconolactonase; PGM : Phosphoglycerate mutase; PGMT : Phosphoglucomutase; phosphoketolase : phosphoketolase; PPAKr : Propionate kinase; PPCSCT : Propanoyl-CoA : succinate CoA-transferase; PTA : Phosphotransacetylase; PTA2 : Phosphate acetyltransferase; PYK : Pyruvate kinase; RPE : Ribulose 5-phosphate 3-epimerase; TGBPA : Tagatose-bisphosphate aldolase; UGLT : UDPglucose-hexose-1-phosphate uridylyltransferase

Bioproject Number	PRJEB42478		
Genome name	<i>Propionibacterium freudenreichii</i>	<i>Lactiplantibacillus plantarum</i>	<i>Lactococcus lactis</i>
Strain name	CIRM-BIA122	CIRM-BIA465	<i>lactis</i> biovar <i>diacetylactis</i> CIRM-BIA1206
Taxon id	1744	1590	1358
Accession number	ERS5564379	ERS5564378	ERS5564377
number of plasmids	Ø	4	7
size of plasmid (bases)	Ø	p1 : 40748, p2 : 30463, p3 : 9152, p4 : 2012	p1 : 40748, p2 : 30463, p3 : 9152, p4 : 2012, p5 : 30463, p6 : 9152, p7 : 2012
Size of the circular genome (bases)	2622405	3121980	2365039
number of rRNA	6	16	19
number of tRNA	45	67	62
percentage of GC (genome, min & max in plasmids)	67	44 (36-41)	35 (30-36)
number of CDS	2472	3112	2563
number of genes in each replicon	chromosome : 2557	p1 : 41, p2 : 27, p3 : 13, p4 : 2 chromosome : 3157	p1 : 69, p2 : 43, p3 : 14, p4 : 9, p5 : 5, p6 : 4, p7 : 3, chromosome : 2542

TABLE 3.1 – Genomic overview of *P. freudenreichii*, *L. lactis* and *L. plantarum* after assembly.

	<i>P. freudenreichii</i>	<i>L. plantarum</i>	<i>L. lactis</i>
Number of genes	1473	1433	1272
Number of metabolites	1284	1045	939
Number of reactions	1790	1523	1337
Percentage of reactions associated to genes	76.6	79.0	82.1
Growth rate in $\text{mmol.gDW}^{-1}.\text{hr}^{-1}$	0.187	0.645	2.172

TABLE 3.2 – Characteristics of the three genome-scale metabolic network reconstructions and their associated FBA models in milk medium at the inoculation point.

FBA à l'inoculation ont prédit un plus grand taux de croissance pour *L. lactis*, et le plus faible pour *P. freudenreichii*, ce qui est consistant avec les observations expérimentales présentées dans la Table 3.2. Dans une seconde expérience, un dosage de métabolites a été performé pour *P. freudenreichii* au temps final, montrant une large production de propionate (8g/L de lait), une forte consommation de lactate (9g/L de lait) ainsi qu'une faible production d'acétate et de succinate (respectivement 3.07 g/L et 0.371g/L de lait) (voir supplémentaire S2).

Nous avons utilisé les conditions expérimentales des cultures pures afin calibrer dynamiquement les modèles correspondant au métabolisme de chaque souches. Expérimentalement, nous avons observé que le métabolisme des bactéries lactiques est toujours actif pendant la phase stationnaire. Comme l'acide lactique est produit par le lactose, et que le lactose est pas un susbtrat limitant, nous avons ainsi régulé le flux de production de l'acide lactique et de l'acétate par le flux de consommation du lactose. En plus, en suivant l'approche de (Özcan et al., 2020), la consommation de lactose fut par la suite contraint par la forme non dissocié de l'acid lactique (voir matériel supplémentaire §??). Dans le cas de *P. freudenreichii*, des bornes de régulations de productions de métabolites supplémentaires ont été estimé en se basant sur les données de dosage des métabolites décrit plus haut dans le but de converger la valeurs de flux vers les valeurs physiologiques (voir matériel supplémentaire §??). En outre, comme les sources de carbonnes ne sont pas limitantes dans le milieu de cultures, nous supposons une croissance logistiques pour modéliser la phase plateau.

Pour calibrer les modèles métaboliques, nous avons estimé les paramètres des modèles dynamiques basés sur les données de mono-cultures en résolvant le problème de minimisation décrit par les équations (3.8)-(3.9). Pour éviter une surévaluation, un nombre réduit de paramètres d'ajustement a été conservé pour chaque modèle. En l'occurrence, seulement deux paramètres ont été utilisé pour calibrer *L. lactis* et la régulation du pH, trois pour *L. plantarum*, qui sont impliqué dans la régulation du pH ainsi que pour la croissance logistique, et enfin, un paramètre pour *P. freudenreichii*, ajustant sa croissance (voir la section matériels et méthode, §??, pour de plus amples détails).

Après optimisation, les simulations des modèles FBA dynamiques (dFBA) calibré avec les données expérimentales de croissance sont représentés avec les figures 3.2, 3.3 and 3.4, (b). Les courbes de pH (Fig. 3.2, 3.3, (c)) sont également correctement restituées, montrant un bon accord du modèle de pH, ainsi que les concentrations finales des métabolites produits par *P. freudenreichii*(Fig. 3.4, (c)). Le coefficient de determination globale retrouvé est de 0.99 (voir données supplémentaires, Fig. S1, fenêtre de gauche).

Dans le but de valider les modèles FBA optimisés représentant correctement le métabolisme, les flux des voies métaboliques d'intérêts décrit dans la figure 3.1 sont représentés dans les figures 3.2, 3.3 and 3.4 (a) lors de l'inoculation où aucun métabolites est limitant.

Le modèle de *L. lactis* montré une préférence de consommation du lactose à travers la voie métabolique du tagatose plutôt que celui de Leloir et une production des isomères de l'acide lactique par la glycolyse, ce qui explique l'acidification du lait et de la caille durant la fabrication du fromage (Fig. 3.2a). Les flux de production du butanediol et du diacetyl surviennent avec l'activation de la voie de l'acétolactate. Un flux de consommation du citrate, *via* la réaction *CITL* codé par le gène *citL* est observé.

Le modèle de *L. plantarum* présenté un flux de consommation du lactose and un flux de production des isomères L et D de l'acide lactique à travers la voie métabolique de la glycolyse et l'acide acétique *via* la voie métabolique de l'acetatekinase, indiquant que *L. plantarum* contribue à l'acidification du fromage (Fig. 3.1b). Le modèle montre différentes valeurs de flux à différents pas de temps à travers le metabolisme hétérolactique (voie métabolique de la transketolase) et homolactique (voie métabolique de la glycolyse). L'activation de la voie métabolique de la transketolase dans un contexte laitier était, à notre connaissance, observé pour la première fois. La consommation du citrate est souche dépendante pour *L. plantarum* (Palles et al., 1998) et a été observé dans notre cas à travers la réaction de transport *FEDCabc*, consuisant la dégradation du citrate pour augmenter la quantité (pool) de pyruvate.

Le modèle de *P. freudenreichii* montré un plus grand flux de consommation de lactose comparé aux isomères L et D de l'acide lactique (Fig. 3.4a). Les cycles du TCA et de Wood-Werkman sont activés simultanément (Deborde and Boyaval, 2000) : le premier permet la régénération des sources de carbones nécessaires pour l'activation du cycle de Wood-Werkman et ainsi, la production de propionate.

3.2.3 Le suivi de la production des métabolites dans les conditions de co-cultures par le modèle de communauté

Un suivi de la croissance des bactéries ainsi que de la production des métabolites ont été réalisé durant le fabrication du fromage. *L. lactis*, *L. plantarum* et *P. freudenreichii* ont atteint respectivement un seuil de culturabilité de $8.45 \log_{10}$ CFU/g, $8.47 \log_{10}$ CFU/g et $8.59 \log_{10}$ CFU/g en approximativement 1200 heures. Nous avons construit un modèle de communauté en regroupant les modèles individuels calibrés sans procéder des optimisations supplémentaires, mais en intégrant la nouvelle valeur de la phase plateau dans la régulation de la croissance logistique au sein de l'équation 3.5 dans le but de reproduire le comportement métabolique de la communauté *in silico* (voir eq. (3.3)-(3.4) et matériels et méthodes §??). Nous notons que, comme aucun ajustement du modèle de communauté n'a été fait, l'expérience de co-culture peuvent être considéré comme des données tests impliquant de nouveaux points de validations ([ML – unseed]).

Le modèle de communauté a prédit de chaque espèce avec une précision acceptable (Fig. 3.5(a)-(c), Table supplémentaire S4) en dépit d'une légère surestimation de la croissance de *P. freudenreichii* durant la phase exponentielle. Les dynamiques des métabolites sont correctement rendues. La consommation du lactose est plus lente dans le modèle que lors des expériences biologiques, tandis que la dynamique du citrate s'ajuste bien avec les données expérimentales. La production de lactate est légèrement surprédit durant la phase exponentielle résultant d'une légère suracidification du fromage durant les étapes de moulage et de démoulage. La production d'acétate est sous-prédit durant les premières phases de la fabrication du fromage, mais le taux de production pendant la maturation est correctement modélisé. Les courbes du propionate, et dans une moindre mesure, celle du succinate sont prédit avec précision par le modèle suggerant que la production de compo-

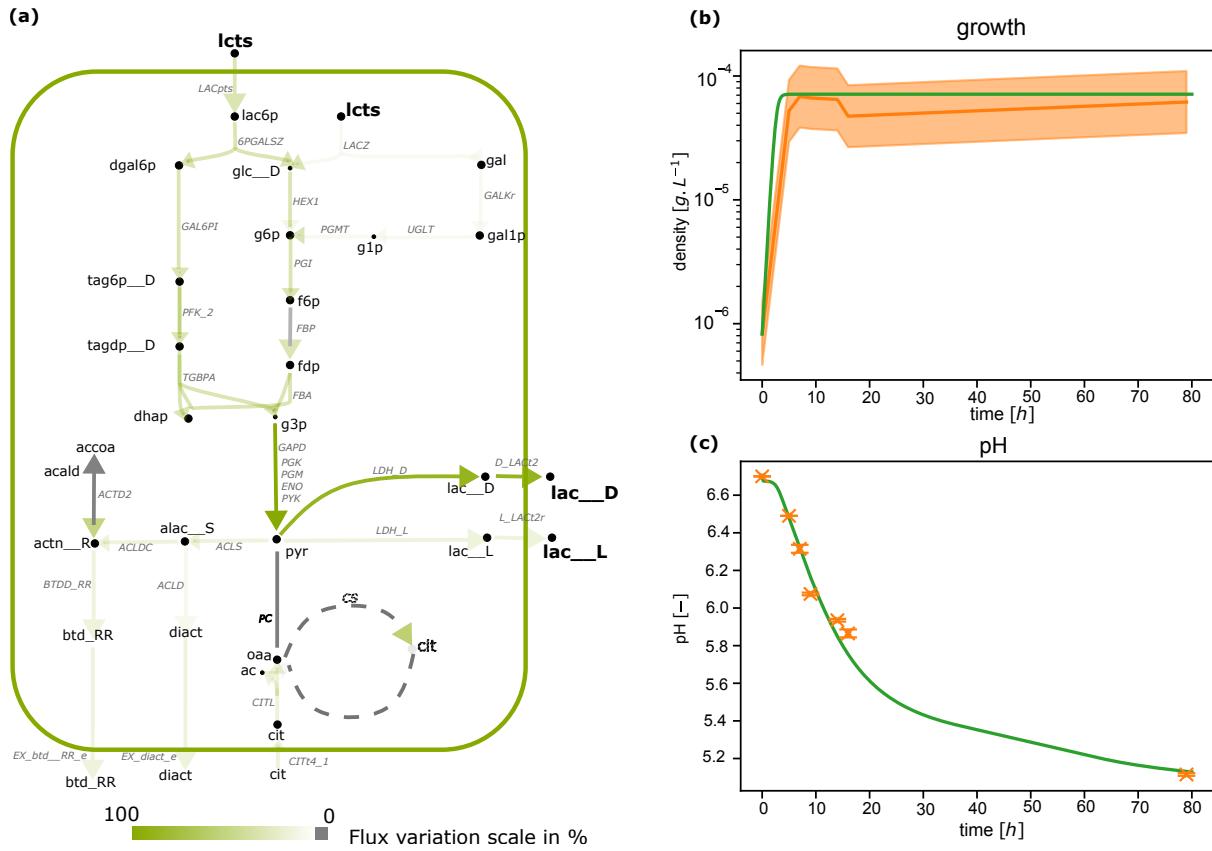


FIGURE 3.2 – *L. lactis*GEM fitted on pure culture data. (a) FBA optimization with optimal parameters applied to the central carbon metabolism at the inoculation step. The color scale represents the reaction flux values predicted by the FBA and normalized by the highest flux value of the illustrated pathways. (b) Dynamics of *L. lactis* population in pure culture after parameter inference in the model (green line) and in the experimental data (orange line, with $\pm 1/4 \log$, which is the admitted range of precision for plating numbering). (c) pH in pure-culture of *L. lactis* in the model (green line) and in the experimental data (orange crosses, with standard deviation). *Compound abbreviations*: ac : acetate; actn_R : acetoine; alac_S : (S)-2-Acetylacetate; btd_RR : butanediol; cit : citrate; dgat6p : d-galactose-6-phosphate; dhap : Dihydroxyacetone phosphate; diact : diacetyl; f6p : fructose-6-phosphate; fdp : D-Fructose 1,6-bisphosphate; g1p : glucose-1-phosphate; g3p : Glyceraldehyde 3-phosphate; g6p : glucose-6-phosphate; gal : galactose; gal1p : galactose-1-phosphate; glc_D : glucose; lac_D, lac_L : lactate; lac6p : lactose-6-phosphate; lcts : lactose; oaa : Oxaloacetate; pyr : pyruvate; tag6p_D : D-tagatose-6-phosphate; tagdp_D : D-Tagatose 1,6-biphosphate. *Reaction abbreviations*: 6PGALSZ : 6-phospho-beta-galactosidase; ACLD : Acetolactate decarboxylase; ACTD2 : Acetoin dehydrogenase; ACLDC : Acetolactate decarboxylase; ACSL : Acetolactate synthase; BTDD_RR : R R butanediol dehydrogenase; CITL : Citrate lyase; CIT4_1 : Citrate transport via sodium symport; CS : Citrate synthase; D_LACt2 : D lactate transport via proton symport; ENO : Enolase; FBA : Fructose-bisphosphate aldolase; GAL6PI : Galactose-6-phosphate isomerase; GALkr : Galactokinase; GAPD : Glyceraldehyde-3-phosphate dehydrogenase; HEX1 : Hexokinase (D-glucose:ATP); L_LACt2r : L lactate reversible transport via proton symport; LACpts : Lactose transport via PEP:Pyr PTS; LACZ : B-galactosidase; LDH_D : D-lactate dehydrogenase; LDH_L : L-lactate dehydrogenase; PC : Pyruvate carboxylase; PFK : Phosphofructokinase; PFK_2 : Phosphofructokinase; PGI : Glucose-6-phosphate isomerase; PGK : Phosphoglycerate kinase; PGM : Phosphoglycerate mutase; PYK : Pyruvate kinase; TGBPA : Tagatose-bisphosphate aldolase; UGLT : UDPglucose-hexose-1-phosphate uridylyltransferase

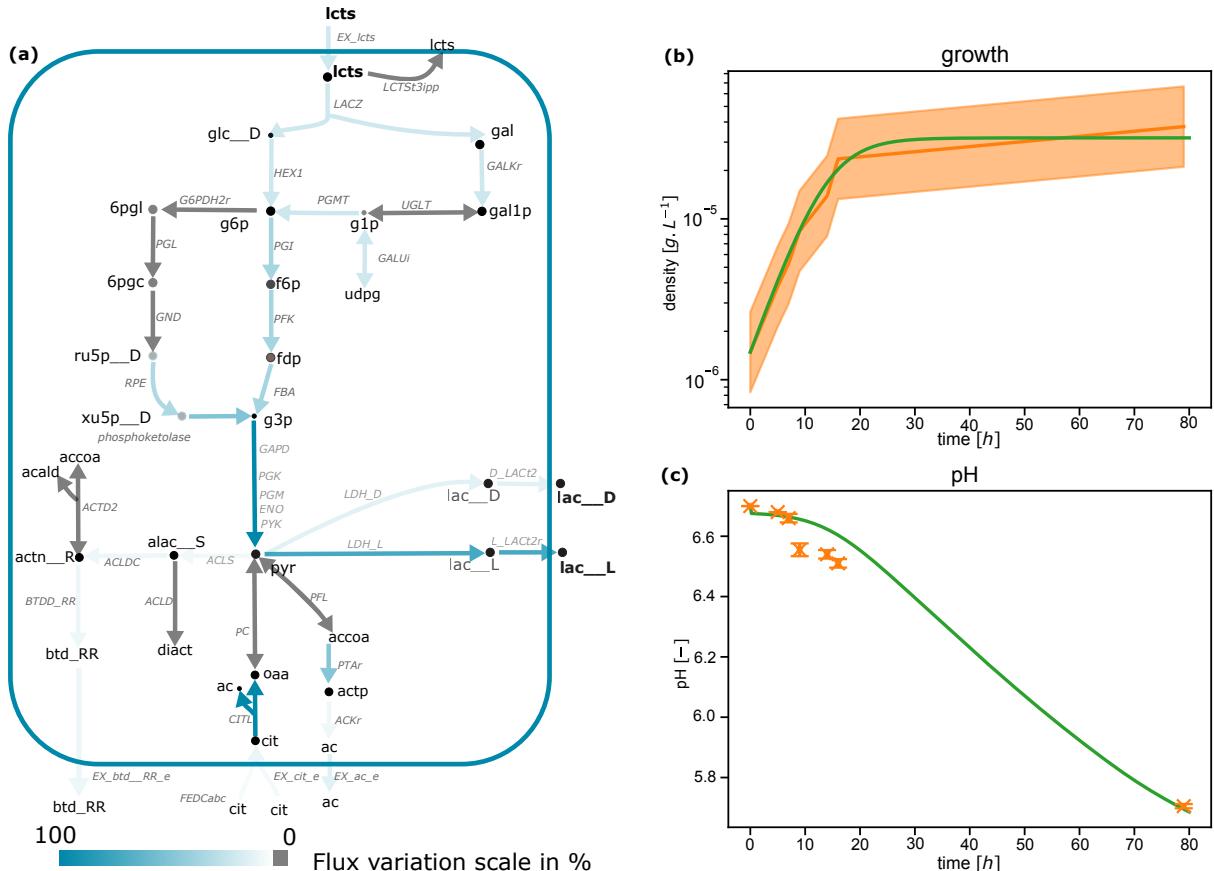


FIGURE 3.3 – *L. plantarum*GEM fitted on pure culture data. (a) FBA optimization with optimal parameters applied to the central carbon metabolism at the inoculation step. The color scale represents the reaction flux values predicted by the FBA and normalized by the highest flux value of the illustrated pathways. (b) Dynamics of *L. plantarum* population in pure culture after parameter inference in the model (green line) and in the experimental data (orange line, with $\pm 1/4$ log, which is the admitted range of precision for plating numbering). (c) pH in pure-culture of *L. plantarum* in the model (green line) and in the experimental data (orange crosses, with standard deviation). 6pgc : 6-Phospho-D-gluconate; 6pgl : 6-phospho-D-glucono-1,5-lactone; ac : acetate; acald : acetaldehyde; accoa : acetyl-coa; actn_R : acetoine; alac_S : (S)-2-Acetolactate; btd_RR : butanediol; cit : citrate; dhap : Dihydroxyacetone phosphate; f6p : fructose-6-phosphate; fdp : D-Fructose 1,6-bisphosphate; g1p : glucose-1-phosphate; g3p : Glyceraldehyde 3-phosphate; g6p : glucose-6-phosphate; gal : galactose; gal1p : galactose-1-phosphate; glc_D : glucose; lac_D, lac_L : lactate; lcts : lactose; oaa : Oxaloacetate; pyr : pyruvate; ru5p_D : D-Ribulose 5-phosphate; tag6p_D : D-Tagatose 6-phosphate; tagdp_D : D-Tagatose 1,6-biphosphate; udpg : UDPglucose; xu5p_D : D-Xylulose 5-phosphate. Reaction abbreviations : ACLD : Acetolactate decarboxylase; ACTD2 : Acetoin dehydrogenase; ACLDC : Acetolactate decarboxylase; ACLS : Acetolactate synthase; BTDD_RR : R R butanediol dehydrogenase; D_LACt2r : D lactate transport via proton symport; ENO : Enolase; FBA : Fructose-bisphosphate aldolase; FEDCabc : FEDCabc; G6PDH2r : Glucose 6-phosphate dehydrogenase; GALkr : Galactokinase; GALUi : UTP-glucose-1-phosphate uridylyltransferase (irreversible); GAPD : Glyceraldehyde-3-phosphate dehydrogenase; GND : Phosphogluconate dehydrogenase; HEX1 : Hexokinase (D-glucose:ATP); L_LACt2r : L lactate reversible transport via proton symport; LACZ : B-galactosidase; LCTSt3ipp : Lactose transport via proton anion (periplasm); LDH_D : D-lactate dehydrogenase; LDH_L : L-lactate dehydrogenase; PC : Pyruvate carboxylase; PFL : Pyruvate formate lyase; PTA : Phosphotransacetylase; PGL : 6-phosphogluconolactonase; PFK : Phosphofructokinase; PGK : Phosphoglycerate kinase; PGI : Glucose-6-phosphate isomerase; PGM : Phosphoglycerate mutase; PGMT : Phosphoglucomutase; phosphoketolase : phosphoketolase; PYK : Pyruvate kinase; RPE : Ribulose 5-phosphate 3-epimerase; UGLT : UDPglucose-hexose-1-phosphate uridylyltransferase.

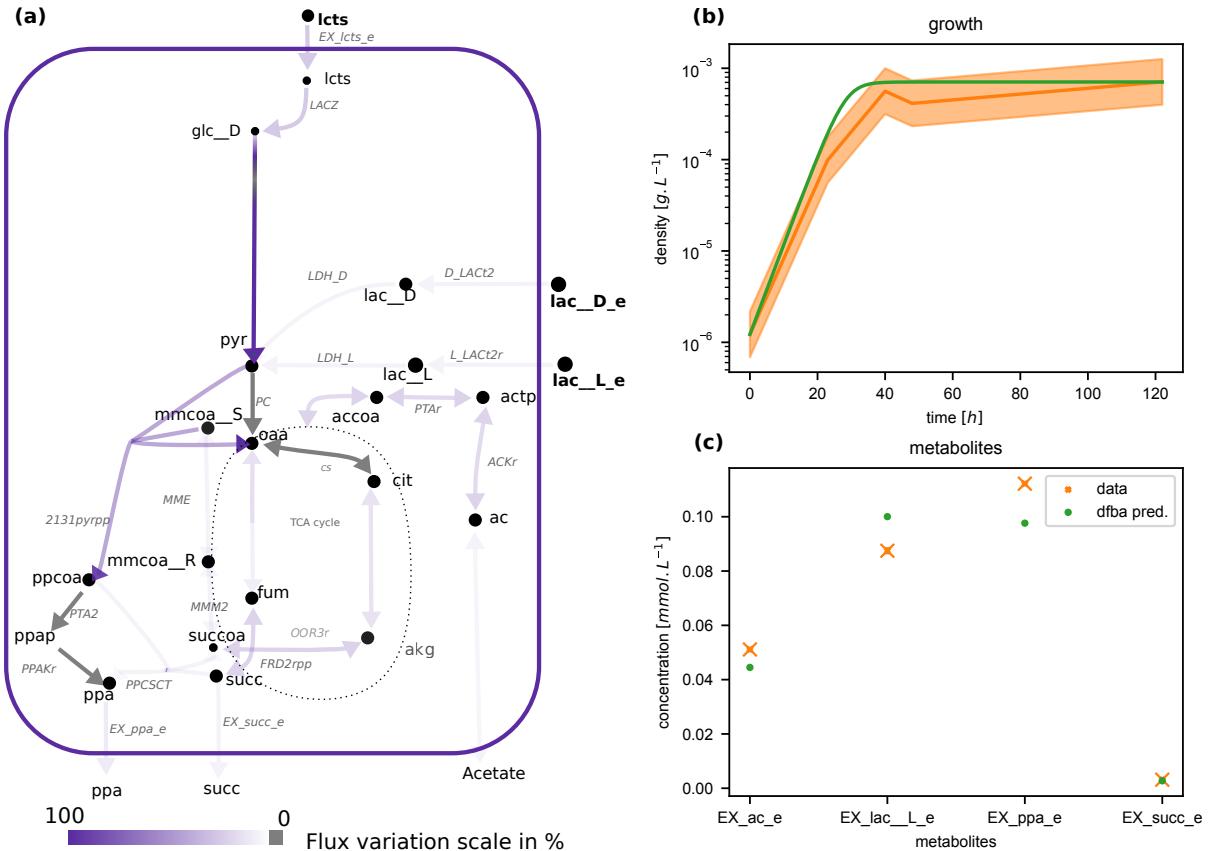


FIGURE 3.4 – *P. freudenreichii*GEM fitted on pure culture data. (a) FBA optimization with optimal parameters applied to the central carbon metabolism at the inoculation step. The color scale represents the reaction flux values predicted by the FBA and normalized by the highest flux value of the illustrated pathways. (b) Dynamics of *P. freudenreichii* population size in pure culture after parameter inference in the model (green line) and in the experimental data (orange line, with $\pm 1/4 \log$, which is the admitted range of precision for plating numbering). (c) Lactate, acetate, propionate and succinate concentrations simulated (green points) and observed (orange crosses, with standard deviation) at the final time point of a second experiment with different lactate inoculation (see Material and method) *Compound abbreviations :* ac : acetate; akg : 2-Oxoglutarate; fum : fumarate; glc_D : glucose; lac_D, lac_L : lactate; lcts : lactose; mmcoa_R : (R)-Methylmalonyl-CoA; mmcoa_S : (S)-Methylmalonyl-CoA; oaa : Oxaloacetate; ppa : propionate.; ppap : Propionyl phosphate; ppcoa : propionyl-coa; pyr : pyruvate; succ : succinate; succoa : succinyl-coa. *Reaction abbreviations :* 2131pyrpp : Methylmalonyl-CoA carboxyltransferase 5S subunit ; CS : Citrate synthase ; D_LACt2 : D lactate transport via proton symport ; FRD2rpp : Fumarate reductase / succinate dehydrogenase (irreversible) (periplasmic, membrane potential dissipating); L_LACT2r : L lactate reversible transport via proton symport ; LACZ : B-galactosidase ; LDH_D : D-lactate dehydrogenase; LDH_L : L-lactate dehydrogenase; MME : Methylmalonyl-CoA epimerase; MM2 : Methylmalonyl-CoA mutase; OOR3r : 2-oxoglutarate synthase (rev); PC : Pyruvate carboxylase; PPAKr : Propionate kinase; PPCSCT : Propionyl-CoA : succinate CoA-transferase ; PTA2 : Phosphate acetyltransferase.

sés organoleptiques par *P. freudenreichii* est correctement capturée. Un R^2 globale de 0.98 est obtenu pour cette étape de validation (données supplémentaires, Figure S1, fenêtre de droite). En outre les voies métaboliques activées en mono-cultures le sont également en co-culture (Figure supplémentaire S3)

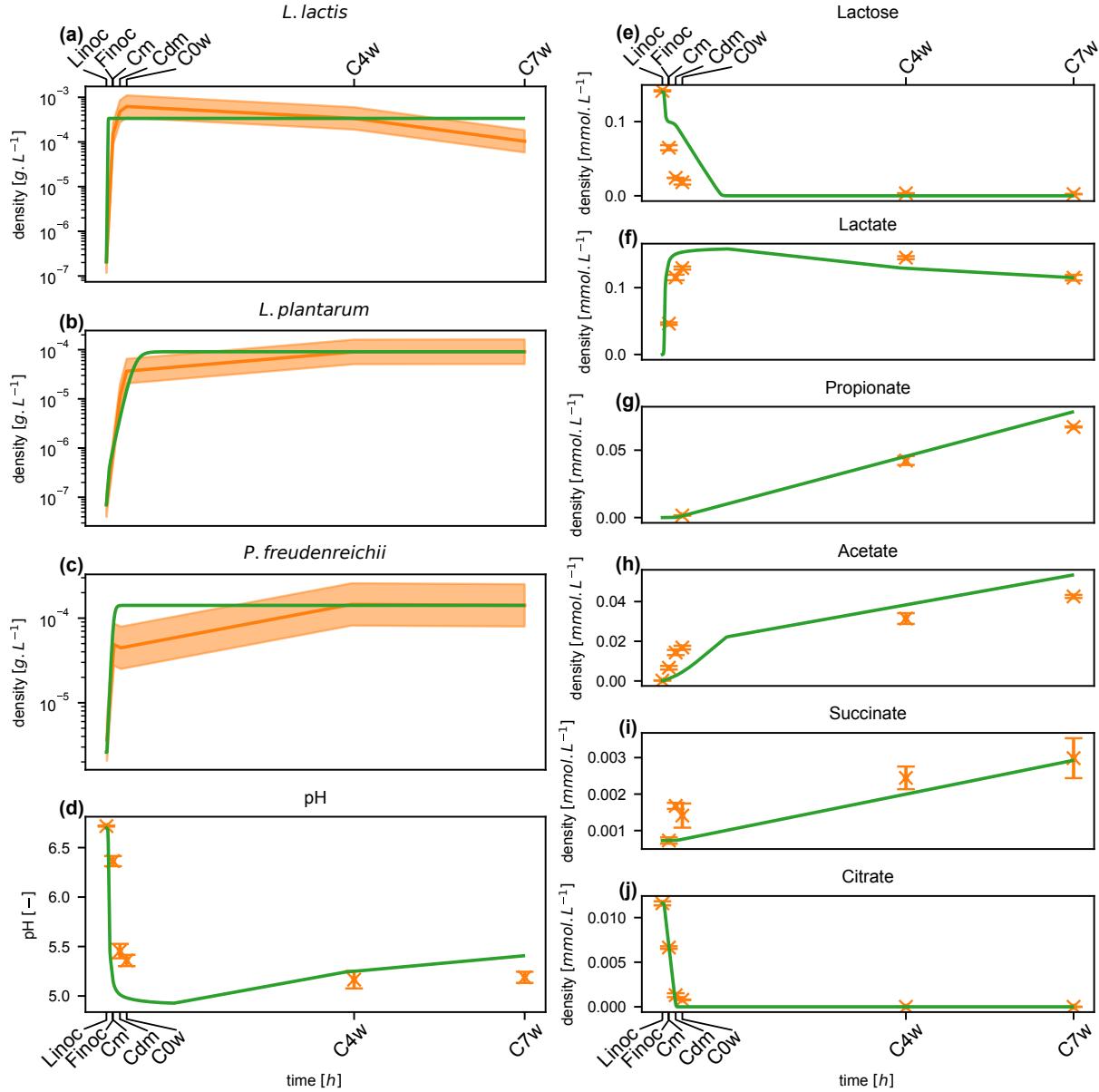


FIGURE 3.5 – (a-c) Simulated growth of respectively *L. lactis*, *P. freudenreichii* and *L. plantarum* species computed by the community model (green lines) and experimentally-observed (orange lines, confidence interval of $\pm 1/4 \log$). (d-j) Computational and experimental co-culture pH and metabolic profiles. Each data point is represented with its corresponding standard deviation (error bars). Abbreviations : Linoc, Lactic acid bacteria inoculation ($t = 0$) ; Finoc, *P. freudenreichii* inoculation ($t = 18$ hours) ; Cm, molding ($t = 19.5$ hours) ; Cdm, demolding ($t = 40$ hours) ; C0w, start of ripening ($t = 60$ hours) ; C4w, Fourth week ripening ($t = 732$ hours) ; C7w, seventh week ripening ($t = 1236$ hours).

3.2.4 Les interactions bactériennes dynamiques au sein de la communauté

Le criblage de la consommation et de la production des metabolites durant en condition de co-culture peut révéler des interactactions bactéries-bactéries, telles que l'alimentation croisée ou la compétition pour un nutriment. Tout d'abord, nous avons enquêté sur les

dynamiques des flux des métabolites échangés pour chaque espèce (Fig. 3.6, (a)) en calculant, à chaque pas de temps de la dynamique, et à l'échelle de la population, les flux d'échange $\mu_{i,j}b_i$ pour chaque métabolite j et bactérie i et les concentrations des métabolites prédictes associées. Pour le citrate et le lactose, *L. lactis* montre une forte importation durant sa phase exponentielle, qui disparaît rapidement après la déplétion complète du citrate et active de la régulation du pH par le lactose. Curieusement, les flux de *L. plantarum* sont négligeables pour ce substrat, suggérant que d'autres métabolites contribuent à sa croissance, comme par exemple, les acides aminés. Cela peut-être du à la faible croissance de *L. plantarum* durant les premières heures de la fabrication de fromage : le citrate est épuisé et l'acidification active la consommation du lactose sous-régulé quand *L. plantarum* est toujours à faible densité. L'épuisement du lactose est atteint par *P. freudenreichii* durant la première phase d'affinage. La production de lactate reflète la consommation de lactose : le lactate est fortement produit par la population de *L. lactis* jusqu'à $t = 250\text{h}$ (affinage), alors qu'il n'y a qu'une faible production par *L. plantarum* observée. Lactate est cependant consommé par *L. plantarum* durant l'affinage, ce qui contribue à l'augmentation du pH dans le système observé dans la Figure 3.5 d. L'acétate est produit par les trois bactéries : *L. lactis* montre une forte production jusqu'à $t = 250\text{h}$, tandis que la contribution de *L. plantarum* augmente doucement jusqu'à atteindre la phase plateau, puis, chute à $t = 250\text{h}$. Après avoir atteint son plateau, *P. freudenreichii* garde un taux constant de production d'acétate durant la maturation du fromage. Le même patron est observé pour la production du butanediol par les deux bactéries lactiques. Le diacétyle est seulement produit par *L. lactis* et sa production s'arrête après la déplétion du lactose. Le propionate et le succinate sont produits par *P. freudenreichii*, qui les relâche à un taux constant durant l'affinage du fromage.

For each strain, we have also calculated the net exchange flux total of metabolites by integrating the FBA prediction of $\int_0^T \mu_{i,j}(t)b_i(t)dt$ over time. After normalizing this net exchange by the total exchange rates within the community (i.e. the sum of net exchanges for each individual), we obtained the contributions of each strain for each metabolite (Fig. 3.6, (b)). These contributions confirm that the succinate and propionate are uniquely produced by *P. freudenreichii*, while the acetoin (production) and citrate (consumption) are metabolized by *L. lactis*. *L. plantarum* contributes to the production of butanediol, even though it is predicted to be the main producer of lactate, along with *L. lactis* and *P. freudenreichii*. A co-dependent relationship between *L. lactis* (producer) and *P. freudenreichii* (consumer) is observed for lactose consumption, which is shared between *L. lactis* and *P. freudenreichii*: curiously, *P. freudenreichii* begins consuming lactose after *L. lactis* stops metabolizing this substrate, indicating a temporal segregation of lactose utilization and no direct competition for this substrate.

The model described above demonstrates that capturing the complexity of the metabolic process in a microbial community during cheese production requires considering the dynamics underlying the system. We asked if the application of computationnal approaches *a priori* based on models of metabolism at the genome scale can predict new interactions between strains. To this end, we used two methods based on fluxes, SMETANA (Zelezniak et al., 2015) and MiCOM (Diener et al., 2020), and a reasoning-based approach, Metage2Metabo (Belcour et al., 2020) to suggest new metabolic complementarity within the consortium.

MiCOM and SMETANA have identified 14 and 25 metabolites exchanged respectively (Supplementary Tables S5 and S6), while Metage2Metabo identified 11 metabolites which

ne peuvent pas être produits par une espèce sans interactions dans la communauté (Table supplémentaire S7). Une première observation est qu'un relativement grand nombre de métabolites sont en communs dans les prédictions des composés échangeables de SME-TANA et MiCOM : lactate, qui est aussi prédit dans le modèle dynamique, phénylalanine, serine, malate, succinate, xanthhine, H₂S, 2-ketoglutarate, glycollate et acetaldehyde. Les autres composés échangés prédit incluent principalement des acides aminés (isoleucine, proline, glycine, alanine). Nous avons utilisé les données metatranscriptomiques dans le but d'évaluer la validité des interactions le plus plausible selon le score SMETANA (voir méthodes et méthodes). H₂S (de *L. lactis* et *L. plantarum* à *P. freudenreichii*), ribose (de *L. lactis* et *L. plantarum* à *P. freudenreichii*), glycerol (de *L. lactis* et *L. plantarum* à *P. freudenreichii*) et phénylalanine (de *L. plantarum* à *P. freudenreichii*). Nous avons vérifié l'expression des gènes associés à la production de ces métabolites chez les espèces donneurs, et à la consommation de ces métabolites chez les espèces receveuses (voir méthodes et Figure supplémentaire S2). Les résultats suggèrent que les interactions pour H₂S, ribose et le glycerol sont plausibles à plusieurs étapes de la fabrication du fromage. A l'inverse, alors que les données d'expression de gènes montrent que les voies métaboliques de la consommation de la phénylalanine est fortement exprimées chez *P. freudenreichii*, leur production par *L. plantarum* n'est pas, suggérant que cette interaction est moins convaincante. Enfin, l'ensemble des métabolites dont la production est prédit par Metage2Metabo comme nécessitant des interactions dans la communauté inclus divers acides gras, comme galactose-1-phosphate, benzyl-Coa, Glyceraldehyde and xanthosine, indiquant une complémentarité métabolique entre les souches.

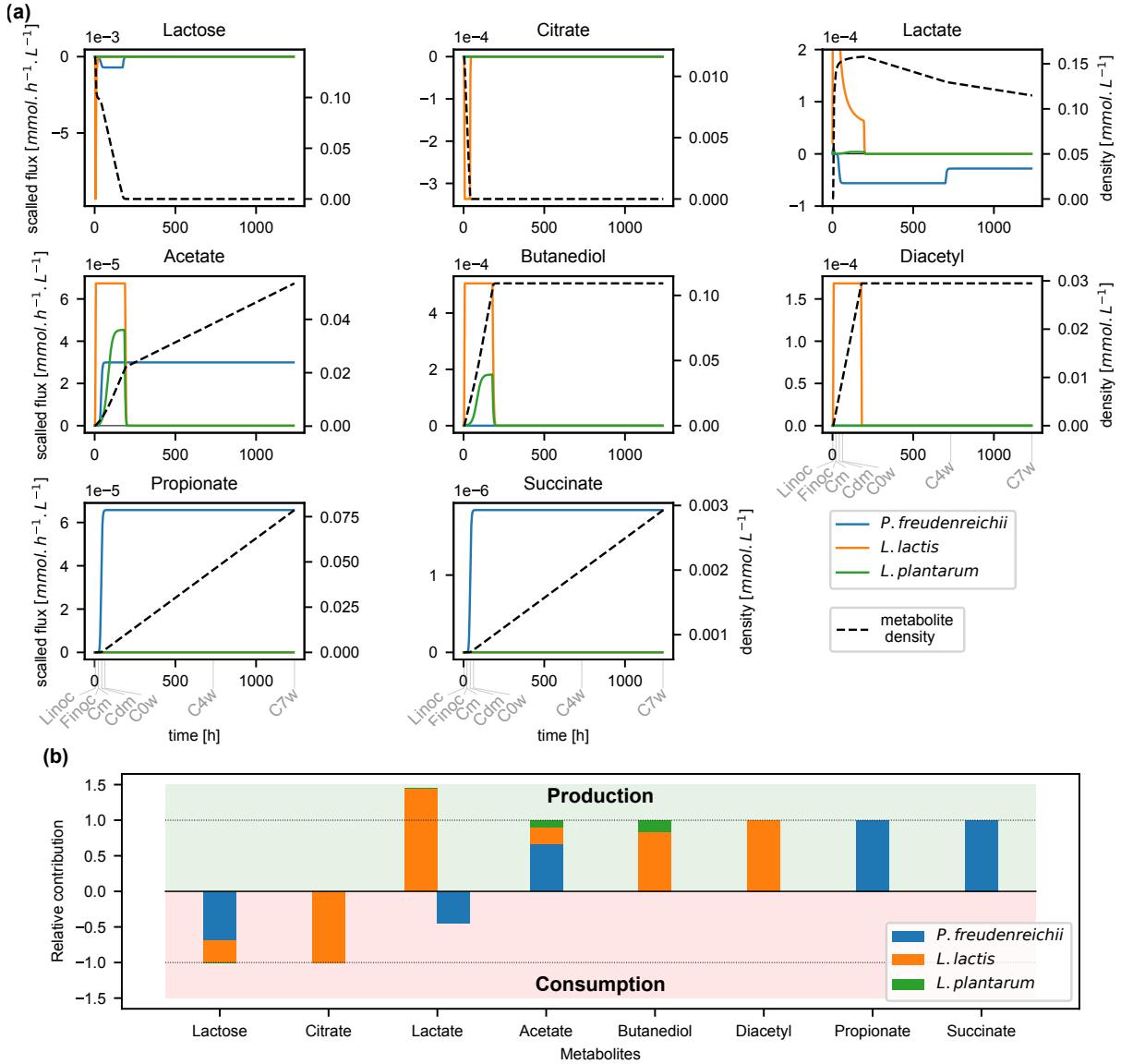


FIGURE 3.6 – (a) **Consumption and production flux dynamics.** We represent for each bacterium the dynamics of the consumption and production fluxes of the followed-up metabolites (colored plain lines). We also represent the dynamics of the concentration of the corresponding metabolite in the cheese (dashed black line, right axis for scale). (b) **Bacterial relative contribution to the metabolite fate.** We represent the relative contribution of each bacteria to the final metabolite concentration. Namely, we computed each microbial overall contribution by integrating in time its consumption or production flux ($\int_0^t \mu_{i,j} b_i$ for metabolite j and bacteria i , see eq.3.4), and normalized the result by the sum over the three bacteria. The value 1 (or -1) then represents the net production (or consumption), i.e. the difference between the final and initial concentration of the metabolite (horizontal dashed lines). For each metabolite, the bacterial contribution is represented by a colored bar, with positive value for production, and negative value for consumption. Abbreviations : Linoc, Lactic acid bacteria inoculation ($t = 0$) ; Finoc, *P. freudenreichii* inoculation ($t = 18$ hours) ; Cm, molding ($t = 19.5$ hours) ; Cdm, demolding ($t = 40$ hours) ; C0w, start of ripening ($t = 60$ hours) ; C4w, Fourth week ripening ($t = 732$ hours) ; C7w, seventh week ripening ($t = 1236$ hours).

3.3 Discussion

Dans ce travail, nous fournissons un modèle de jumeau numérique du métabolisme bactérien pendant le processus de fabrication de fromage. En utilisant les modèles métaboliques de trois souches bactériennes inoculées et des données omiques, we etions capable de reproduire, avec précision, les motifs de production de metabolites ainsi que la dynamique des populations bactériennes. Nous nous sommes reposés sur les conditions expérimentales de cultures pures pour calibrer les modèles métaboliques individuels avec peu de paramètres pour chaque, ainsi limitant la surévaluation du modèle durant les simulations communautaires. Une originalité du modèle est sa capacité de prédire les dynamiques sur tout le procedé de fabrication du fromage (sept semaines). Notre travail soulève des hypothèses sur les voies métaboliques mis en jeu dans le lait, ainsi que la contribution de chaque souche bactérienne dans la consommation de nutriments et la productibilité de composé organoleptiques.

La connaissance des microbiologistes, de la littérature scientifique et des données multi-omiques étaient nécessaire pour accomplire des modèles métaboliques à l'échelle du génome de grandes qualités et les simulations dynamiques qui en découlent. Le métabolisme carbonné central des bactéries lactiques produisent principalement l'acide lactique à partir de la dégradation du lactose *via* les voies métaboliques de la glycolyse, du tagatose et de Leloir (Kleerebezem et al., 2003 ; Van Rooijen et al., 1991 ; Widjastuti et al., 2014) ; pour la première fois, l'utilisation du citrate et de la fermentation hétérolactique ont été observé dans le modèle curé de *L. plantarum* lait. La vérification que chaque modèle métabolique individuel peut activer les différentes voies métaboliques décrites plus haut était ainsi une étape important pour valider les modèles, permettant la reproduction d'activités enzymatiques observé dans (Carroll et al., 1999 ; Quatravaux et al., 2006). La bactérie propionique *P. freudenreichii* convertie le lactose, et préferenciellement le lactate en propionate selon (Loux et al., 2015 ; Thierry et al., 2011). En suivant (Borghei et al., 2021), nous avons ajouté l'enzyme propionyl-CoA:succinate Coa transferase (2.8.3.-) pour completer le cycle de Wood-Werkman, et ainsi assurer la production de propionate.

Dans (Özcan et al., 2020), les paramètres individuels de calibration pour tous les nutriments du milieu de culture sont imposés dans le but d'expliquer la métabolomique à l'échelle de l'espèce, et de les prédire à l'échelle de la communauté. En comparaison de ce travail, nous avons implémenté un modèle FBA dynamique en se basant sur (Mahadevan et al., 2002b), qui a utilisé des paramètres de calibration souche spécifique pour prédire la croissance, le pH et les concentrations métaboliques. Pour éviter la surévaluation du modèle, nous avons réduit le nombre de paramètres inférés pour chaque bactérie, où seulement deux paramètres sont gardés pour la régulations du pH pour les bactéries lactiques et un paramètre se focalisant sur les croissances de *L. plantarum* et *P. freudenreichii*. Des calibrations additionnelles ont été mené pour *P. freudenreichii*sans aucune optimisations mathématiques supplémentaires, en calculant la valeur de la borne supérieure des métabolites dosés en monoculture.

Le modèle de co-culture a donné des indices sur le comportement de la communauté pendant la fabrication du fromage et a suggéré temporellement la contribution de chaque espèce impliqué dans la production des composés aromatiques. Il est montré que la production de propionate et d'acetate peut être attributé uniquement à *P. freudenreichii*, comme indique dans (Cao et al., 2021), et que le diacetyl semble être seulement produit par *L. lactis*. D'après le modèle de communauté, le butanediol a été produit pendant l'étape du moulage et au début de l'affinage par à la fois, *L. lactis* et *L. plantarum*, comme observé dans les modèles de mono-cultures respectifs. En regardant la production d'acé-

tate, le modèle déduit une production précoce de l'étape du moulage jusqu'à l'affinage par *L. lactis* et *L. plantarum* suivi par *P. freudenreichii* pendant l'affinage, responsable de la majeur partie de production de l'acétate à cette étape. Concernant l'utilisation des sources de carbonnes, alors que *L. plantarum* possède les voies métaboliques de dégradation du lactose et du citrate, qui sont complètement activées dans le modèle en monoculture, il apparaît que ces voies métaboliques sont fortement sous régulés en co-culture. Lors de la mise en culture avec *L. lactis*, et comme la vitesse de croissance de *L. plantarum* est la plus faible des deux, il accomplit son efficacité métabolique maximale après la phase exponentielle de *L. lactis*. Comme *L. lactis* pousse principalement sur le lactose produisant le lactate, sa phase exponentielle est associée à la forte acidification de milieu, ce qui active la régulation du pH par la répression du lactose sur le métabolisme de *L. plantarum*. *L. plantarum* peut utiliser des acides aminés ou d'autres métabolites non suivis par le modèle dynamique permettant sa croissance. Pour explorer cette piste, nous avons utilisé des outils dans le but d'identifier des interactions bactériennes en dehors de l'ensemble des composés suivis par le modèle dynamique en co-culture, révélant d'autres interactions métaboliques putatives, où les acides aminés représentent une grande proportion.

Dans le modèle de co-culture, *P. freudenreichii* garde activé son métabolisme dans la dernière phase de production d'un fromage : l'affinage. Ce métabolisme est reflété par la consommation de lactate durant tout l'affinage, et le relargage de composés organoleptiques à flux constant (Figure 3.6). Ce comportement est cohérent avec les capacités métaboliques connues de *P. freudenreichii* et avec les données métabolomiques obtenues durant le procédé de production du fromage. Dans le modèle, *P. freudenreichii* consomme aussi le lactose restant après la croissance de *L. lactis*. En effet, durant sa phase exponentielle, *L. lactis* metabolise principalement le lactose, produisant fortement du lactate et réduisant le pH. La réduction du pH inhibe l'absorption du lactose régulée par le pH dans les deux bactéries lactiques, et donc stoppant la consommation de lactose par les *Lactobacilli*. Comme le lactose reste dans le milieu de culture et que *P. freudenreichii* est l'unique bactérie capable de l'utiliser, il épuise donc le lactose encore présent. Dans les données métatranscriptomiques, le gène LACZ est hautement activé par *P. freudenreichii* durant l'affinage confirmant la dégradation du lactose par cette souche après l'acidification (Figure S2).

Ce travail revèle comment la combinaison d'expert biologique, des données omiques et de la modélisation métabolique permet d'obtenir des prédictions précises sur les mécanismes responsables du comportement dynamique de la communauté microbienne. D'autre part, cela souligne également que la quantité de données et les efforts nécessaires pour créer de tels modèles de haute qualité restent un prix élevé à payer malgré l'amélioration des approches de simulation. Des développements méthodologiques doivent encore être proposés afin d'automatiser l'étalonnage des modèles avec des données et de garantir à la fois l'exactitude des prédictions et l'extensibilité à des communautés plus importantes ou à des communautés de composition empirique.

3.4 Matériels et méthodes

3.4.1 Données biologiques

3.4.1.1 composition communauté.

Les souches qui ont été utilisées dans le projet sont décrites dans (Cao et al., 2021). Briefement, la communauté bactérienne contrôlée est composée de deux bactéries lactiques

(LAB), *L. lactis* subsp. *lactis* biovar *diacetylactis* CIRMBIA1206 (nommé *L. lactis*), *L. plantarum* CIRMBIA465 (nommé *L. plantarum*), et une bactérie propionique *P. freudenreichii* CIRMBIA122 (nommé *P. freudenreichii*) provenant du Centre international pour les ressources microbiennes - Bactéries associées aux aliments (CIRM-BIA¹)

3.4.1.2 Fabrication de fromage et collecte des échantillons.

Le protocole standard de production d'un fromage décrit dans (Cao et al., 2021) a été utilisé. Brievement, après la collecte de lait frais, pasteurisé (76°C, 20 secondes), écrêmer, standardisé (jusqu'à 30 grams de gras et 36 mg de calcium par kg de lait, *L. lactis* et *L. plantarum* sont inoculés à 5.7 et 5.2 log₁₀ CFU/g (Unité de Formant Colonie). Après pré-maturation du lait inoculé (à 14 °C pour 18 h), *P. freudenreichii* a été inoculé à 6.1 log₁₀ CFU/g et remué pendant 20 min. Le lait est versé dans des cuves et chauffé à 33°C pendant 30 min. De la presure commerciale a été ajouté et le coagulum fut coupé, mélangé, lavé, drainé, pré-pressé et transféré dans des moules pour le présurage. Le demoulage apparaît le troisième jour de la fabrication du fromage, et les fromages ont été salé (immersion à 12°C pendant 10h avec de la saumure en excès), puis séché toute la nuit et emballés sous vide dans des sacs en plastique le quatrième jour de la fabrication du fromage pour un affinage de 7 weeks à 13 °C.

Après l'inoculation des bactéries lactiques (Linoc, t=h) et de *P. freudenreichii* (Finoc, t=18h), les échantillons sont collectés à cinq étapes de production pour de plus amples analyses comme décrit dans (Cao et al., 2021) : l'étape de mouillage (Cm, t=19.5h), l'étape de démouillage (Cdm, t=40h), le commencement de l'affinage (C0w, t=60h) après quatre et sept semaines d'affinage (C4w, t=732h et C7w, t=1236h). Les fromages sont répliqués quatre fois (des répétitions biologiques ont été réalisés à différents jours avec des incolores et laits différents), et échantillonés chaque fois pour la numérisation et la biochimie et deux fois (répétitions biologiques) pour le séquençage ARN.

3.4.1.3 Expériences de culture souche spécifique.

Dans le but de calibrer les modèles métaboliques, un jeu de données d'entraînement a été réalisés en utilisant les données expérimentales de chaque souche individuel. Ces croissances individuelles des trois bactéries fut réalisé dans un lait gras UHT (Delisse) en tube dans des conditions microaérobie, c'est à dire, sans secouer dans l'incubateur à 30°C, avec deux réplicats. Pour *P. freudenreichii*, le lait fut supplémenté en lactate (20 g/L) et avec de la peptone caseine (10 g/L). Une seconde expérience de croissance fut réalisé pour *P. freudenreichii* dans le but de mesurer la concentration finale de métabolites (acétate, lactate, propionate, et succinate). *L. lactis* fut inoculé à 1×10^6 et 4×10^6 CFU/g, *L. plantarum* à 3×10^6 et 6×10^6 CFU/g et *P. freudenreichii* à 1.4×10^7 et 9×10^6 CFU/g. Les données de croissance en culture pure sont disponibles en Table S1.

3.4.1.4 Données métatranscriptomiques.

Dix grammes de fromage ont été dispersé dans 90 grammes d'une solution de citrate de sodium à 2% durant 1 minute à vitesse maximal suivie de 1 min à faible vitesse dans un stomacher (Humeau, Treillières, France) à température ambiante. Dix millilitres de la dispersion ont été centrifugés à 10000 grammes à 4°C pendant cinq minutes. Les culots cellulaires ont été conservés à -80°C pour l'extraction de l'ARN. Après décongélation, les

1. <https://collection-cirmbia.fr/>

échantillons ont été centrifugés à nouveau pendant 5 minutes à 4°C, 10 000 grammes pour retirer le surnageant. Les culots cellulaires ont été suspendus dans 200 L de tampon de lyse (50 mM Tris-HCl, 1 mM EDTA ; pH 8,0) 20 mg/g de lysozyme (MP Biomedicals, Illkirch, France) et 50 U/g de mutanolysine (Sigma, Saint Quentin Fallavier, France) et incubés pendant 30 min à 24°C. Les suspensions ont ensuite été transférées dans des tubes de 2 mL contenant 50 mg de billes de zirconium (diamètre : 0,1 mm ; BioSpec Products, Bartlesville) et 50 L de SDS (20 %). Les échantillons ont ensuite été agités dans un Precellys Evolution (Bertin, Montigny-le-Bretonneux, France) pendant deux cycles de 40 s à 6500 rpm. L'ARN a ensuite été extrait des lysats cellulaires à l'aide de Qiazol et du kit RNeasy Mini (Qiagen) selon (Falentin et al., 2010) avec deux ou trois étapes de traitement ultérieur à la DNase (Dnase Rnase free, Ambion) selon les instructions du fabricant. L'intégrité de l'ARN (RIN) a été évaluée à l'aide d'un bioanalyseur Agilent 2100 (Agilent Technologies, Santa Clara, CA). Les concentrations d'ARN ont été quantifiées à l'aide de la quantification fluorométrique Qubit (Thermofisher, France). L'absence de contamination par l'ADN a également été vérifiée à l'aide du Qbit et du protocole de quantification de l'ADN conformément aux instructions du fabricant. Les échantillons d'ARN ayant une valeur RIN supérieure à 8, indiquant une bonne intégrité de l'ARN, ont été conservés pour la suite de l'analyse.

Les données métatranscriptomiques ont été obtenues à cinq étapes de la fabrication du fromage. Les librairies Illumina et le séquençage ont été réalisés par GeneWiz (Leipzig, Allemagne). Les lectures sont disponibles à l'ENA (Cambridge) sous le numéro d'accession ENA : PRJEB42478. Les lectures séquencées ont été découpées à l'aide de Trim Galore et mises en correspondance avec les génomes de référence à l'aide de Bowtie 2 (Langmead et al., 2009). Aucune erreur d'appariement n'a été autorisée et seules les lectures correspondant à des séquences uniques ont fait l'objet d'une analyse plus poussée. Les lectures correspondant à chaque séquence codante ont été comptées à l'aide de HTSeq-count (-stranded=reverse, -a 0, -t CDS, -i db_xref, -m union) (Putri et al., 2022).

Les analyses statistiques ont été réalisées à l'aide du logiciel statistique R version 4.0.3 (R Core Team, 2021). Les données brutes de comptage de l'ARN-Seq ont été normalisées en deux étapes. Tout d'abord, un facteur d'échelle spécifique à l'espèce a été appliqué pour éliminer les biais de composition entre les bibliothèques. Pour chacune des trois espèces, elles ont été calculées en utilisant tous les gènes de l'espèce considérée et deux autres rangées pour l'autre espèce, calculées comme la somme des comptages attribués à chacune de ces deux autres espèces, en utilisant la méthode TMM (Trimmed Mean of M-Values) telle qu'implémentée dans le package edgeR version 3.32.1 (Robinson et al., 2009). Une étape supplémentaire de normalisation au sein de l'échantillon a été réalisée pour corriger la longueur des gènes et permettre la comparaison. Enfin, la moyenne des réplicats a été calculée.

3.4.1.5 Métabolomiques ciblées.

Les méthodes sont décrites dans (Cao et al., 2021). Brièvement, les sucres et les acides organiques dans les échantillons sont quantifiés en utilisant la chromatographie liquide à haute performance (HPLC). Les composés volatils sont analysés en utilisant l'extraction par sorption de l'espace de la tête (HS) et couplés à la chromatographie en phase gazeuse-spectrométrie de masse (GC-MS).

3.4.2 Modelisation métabolique

3.4.2.1 Séquençage du génome.

Des culots cellulaires (équivalent à 1e^{10} CFU) ont été obtenus par centrifugation pendant 10 minutes à 5000 g à partir de cultures de souches dans un milieu de bouillon. L'ADN fut extrait en utilisant le kit DNeasy midi (Qiagen) d'après le protocole décrit dans (Falentin et al., 2010). Les génomes sont séquencés par Illumina NovaSeq 2 x 1500 pb et PacBio Sequel (Eurofins Genomics, Constance, Allemagne). Les génomes sont assemblés en utilisant Unicycler (Wick et al., 2017a,b) sur la plateforme bioinformatique Galaxy Genotoul (Toulouse, France). Les séquence génomiques sont intégrés sur la plateforme MicroScope hébergé au Genoscope (CEA, Evry, France) pour une annotation automatique selon (Vallenet et al., 2019). Les génomes annotés sont disponibles à l'ENA (Cambridge) sous le numéro d'accès ENA : PRJEB54980.

3.4.2.2 Reconstruction de modèle métaboliques à l'échelle du génome.

Les génomes bactériens individuels sont annotés fonctionnellement en utilisant la plateforme MicroScope (Vallenet et al., 2019). Les réseaux métaboliques à l'échelle du génomes sont par la suite reconstruits à partir de ces génomes annotés avec CarveMe (Machado et al., 2018) version 1.4.1. Chaque modèle métaboliques à l'échelle du génome est poli en utilisant ModelPolisher (Römer et al., 2016) qui améliore l'annotation en ajoutant des informations liés aux identifiants BIGG depuis la base de connaissances BIGG (King et al., 2016). La consistance de chaque modèle est vérifié avec MEMOTE (Lieven et al., 2020) à la fin de l'étape de polissage (<https://memote.io>). Le milieu nutritionnel inclus le milieu minimal trouvé par CarveMe et les composé du lait (<https://ciqual.anses.fr/#/aliments/19024/lait-entier-pasteurise>, où les acides aminés sont considérés en excès (fichier supplémentaire 1).

3.4.2.3 Analyse par équilibre des flux.

Un modèle métabolique à l'échelle du génomes (GEM) est reconstruit pour chaque souche en utilisant le réseaux métabolique et la composition nutritionnelle, et analysé avec la méthode de l'analyse par équilibre des flux (FBA) (Orth et al., 2010). Le FBA calcule une distribution de flux optimal satisfaisant la stoechiométrie et la faisabilité biophysique donnant un maximum de biomass, en solvant le programme linéaire suivant :

$$\begin{aligned} \text{trouver } v^* \in \mathbb{R}^{N_r} \text{ tel que } v^* = & \quad \arg \max_{S.v = 0} v_{growth} \\ & v_{min}^{in} \leq v \leq v_{max}^{in} \\ & v_{min}^{ex} \leq v \leq v_{max}^{ex} \end{aligned} \tag{3.1}$$

ou v_{growth} est le flux de la réaction de biomasse de modèle, S est la matrice stoechiométrique dérivé du modèle métabolique, v_{min}^{in} et v_{max}^{in} (resp. v_{min}^{ex} et v_{max}^{ex}) sont les valeurs minimales et maximales de flux intracellulaires (resp. échangés) définissant les bornes de flux pour le vecteur de flux v contenant \mathbb{R}^{N_r} dimension. Les ensembles de valeurs $(v_{min}^{in}, v_{max}^{in})$ et $(v_{min}^{ex}, v_{max}^{ex})$ dans les GEMs sont disponibles dans le matériel supplémentaire.

L'équation (3.1) définit alors, pour chaque bactérie i , la correspondance μ_i entre les vecteurs de contraintes sur les réactions d'échange définissant l'environnement nutritionnel

et le flux optimal v^* obtenu dans ce contexte nutritionnel.

$$\mu_i(v_{min}^{ex}, v_{max}^{ex}) = v^* \quad (3.2)$$

Le problème d'optimisation FBA (3.1) fut résolu en utilisant CobraPy (Ebrahim et al., 2013) version 0.17.1 utilisant l'ordre lexicographique pour garantir une solution unique (Gomez et al., 2014).

3.4.2.4 Raffinement manuel des GEMs

Une étape de raffinement manuel fut réalisé pour tous les modèle pour assurer la bonne production de composés organoleptiques et l'activité des fonctions métaboliques d'intérêt au regard de la littérature. Les modifications sont récapitulées dans le matériels supplémentaire Table S3. Comme les trois espèces sont micro-aérobique, la borne d'import de l'oxygène est défini comme la valeur minimal permettant la production de composés aromatiques.

3.4.2.5 Modélisation dynamique

Le comportement dynamique de chaque système de culture est calculé en utilisant la l'analyse par équilibre des flux dynamique (dFBA) (Mahadevan et al., 2002b) en ajoutant des mécanismes dynamiques de la population. Notons $b = (b_i)$ pour $i \in \mathcal{B} = \{L. lactis, L. plantarum, P. freudenreichii\}$ les densités de la population bactérienne, et $m = (m_j)_{1 \leq j \leq N_m}$ la densité de composés métaboliques N_m qui sont expérimentalement suivi, nous avons ainsi le système, pour $i \in \mathcal{B}$ et $1 \leq j \leq N_m$

$$\partial_t b_i = \mathcal{R}_i(b_i) \mu_{i,i} ((c_{min,i}^{ex}, c_{max,i}^{ex})(b, m)) b_i \quad (3.3)$$

$$\partial_t m_j = \sum_{i \in \mathcal{B}} \mu_{i,j} ((c_{min,i}^{ex}, c_{max,i}^{ex})(b, m)) b_i \quad (3.4)$$

où le terme $\mu_{i,i}$ est la composante correspondant au métabolite j (ou à la biomasse i) de μ_i , calculé à partir de 3.2 compte tenu de l'ensemble des contraintes $(c_{min,i}^{ex}, c_{max,i}^{ex})(b, m)$. Ces contraintes dependent dynamiquement de l'état de la variable b et m .

Le vecteur \mathcal{R}_i modélise un processus de régulation de la population sur la croissance de la population de manière phenomenologique avec le terme

$$\mathcal{R}_i = \lambda_i (1 - \frac{b_i}{\beta_i}), \text{ for } i \in \mathcal{B} \quad (3.5)$$

où λ_i est un poids et β_i est la capacité de charge de la population, correspondant à la valeur de la phase plateau de chaque bactérie dans l'expérience.

Pour un substrat j et une modèle métabolique i , la bornes usuelles sont

$$c_{min,i,j}^{ex} = \max(-\frac{m_j}{\Delta t \sum_{i \in \mathcal{M}_j} b_i}, v_{i,j}^{int}) \quad c_{max,i,j}^{ex} = 0, \quad (3.6)$$

où \mathcal{M}_j est le sous ensemble de bactérie métabolisant j and Δt est une caractéristique temporelle du temps d'importation. Ces équations reflète le partage uniforme des ressources parmis les bactéries durant la fenêtre temporelle Δt . Quand les substrats sont en excès, le flux est limité par la limite d'import intrinsèque $v_{i,j}^{int}$.

Certains substrats et produits sont également régulés. En suivant l'approche de (Özcan et al., 2020), comme le métabolisme du lactose par les bactéries lactiques est régulé négativement par la forme non dissociée de l'acide lactique avec une décroissance exponentielle, pour $j = lcts_e$ et $i \in \{L. lactis, plantarum\}$, nous avons

$$c_{min,i,j}^{ex} = \max\left(-\frac{m_{lcts_e}}{\Delta t * \sum_{i \in \mathcal{M}(lcts_e)} b_i}, -\mu_{max,lcts} * 10^{(-k_{lac} * \phi_{undiss})} - \mu_{min,lcts}\right)$$

où k_{lac} correspond à la décroissance exponentielle, $\mu_{max,lcts}$ et $\mu_{min,lcts}$ sont les valeurs maximales et minimales et ϕ_{undiss} est une fonction calculant la concentration de la forme non dissociée de l'acide lactate à partir du lactate. Cette fonction ϕ_{undiss} est dérivée de l'équation de Henserson-Hasselbalch comme dans (Özcan et al., 2020) et lit

$$\phi_{undiss}(m_{lac_L_e}, m_{lac_D_e}) = \frac{m_{lac_L_e} + m_{lac_D_e}}{1 + 10^{c_1 * (m_{lac_L_e} + m_{lac_D_e}) + c_2}}. \quad (3.7)$$

Dans cette équation le terme $c_1 * (m_{lac_L_e} + m_{lac_D_e}) + c_2$ est une approximation linéaire de $(pH - pKa)$, alors que les paramètres c_1 et c_2 sont inférés par une régression linéaire des données de co-culture. La production de lactate est régulée de la même manière, alors que la production d'acétate est régulé en fonction de la disponibilité en lactose. Voir matériel supplémentaire section ?? pour une description détaillée et justifications des bornes dynamiques configurées sur les réactions d'échange.

Pour *P. freudenreichii*, les bornes supérieures de productions sont évalué en calculant par le flux de production à partir des dosages de métabolites et des données de croissance dans les expériences de culture pure, en supposant un flux constant (voir le matériel supplémentaire ??).

La condition initiale, l'inoculum et l'environnement nutritionnel constitué des composés du lait, acides aminés et co-facteurs, a été défini comme spécifié dans (Cao et al., 2021), après conversion des unités (voir matériel supplémentaire pour la liste des métabolites). Une masse cellulaire de 0.33×10^{-12} g/CFU a été appliquée pour la conversion d'unité de biomasse.

3.4.2.6 Implementation numérique

Le système dynamique est résolu avec le schéma semi-implicit d'Euler. A chaque étape n , après le calcul de F_j le flux total du composé j comme défini par l'ensemble des GEMs,

$$F_j = \sum_{i \in \mathcal{B}} \mu_{i,j} ((c_{min,i}^{ex}, c_{max,i}^{ex})(b^n, m^n)) b_i,$$

nous calculons,

$$\begin{aligned} b_i^{n+1} &= b_i^n + \Delta t * F_{b_i} \\ m_j^{n+1} &= \begin{cases} m_j^n + \Delta t * F_j & \text{si } F_j > 0 \text{(cas explicit)} \\ m_j^n / (1 - \Delta t * F_j / m_j^n) & \text{sinon (cas implicit)} \end{cases} \end{aligned}$$

Ce schéma garantie la positivité de la solution pour tous les temps.

Les problèmes dFBA (3.3)–(3.4) ainsi que les paramètres inférés sont résolu en créant un script Python et en utilisant les bibliothèques numpy v.1.19.5(Harris et al., 2020), pandas v.0.25.3 (McKinney et al., 2010), and scipy v.1.5.3 (Wes McKinney, 2010). Durant le dFBA, le FBA est résolu en utilisant l'ordre lexicographique défini (Gomez and Barton, 2018).

3.4.3 Ajustement du modèle

Pour ajuster les paramètres du modèle, nous avons adapté les équations (3.3)–(3.4) aux expériences de monocultures. Nous avons ensuite inféré le vecteur de paramètre du modèle θ du résultat du modèle de dFBA en maximisant le logarithme de la ressemblance des données observé en monoculture. Formellement, pour $i \in \{L. lactis, L. plantarum\}$, nous avons minimisé la fonction coût suivante

$$J(b_i, pH|\theta_i, b_{i,exp}, pH_{exp}) = \left\| \frac{\log_{10}(b_i) - \log_{10}(b_{i,exp})}{\sigma_{log,i,exp}} \right\|^2 + \alpha \left\| \frac{pH - pH_{exp}}{\sigma_{pH,exp}} \right\|^2 \quad (3.8)$$

où σ (resp. $\sigma_{log,i}$) est l'écart-type (resp. l'écart-type log transformé) des données correspondantes, et $\theta_i = (k_{lac}, \mu_{max,lcts})$ si $i = L. lactis$ et $\theta_i = (k_{lac}, \mu_{max,lcts}, \lambda_i)$ si $i = L. plantarum$. Pour $P. freudenreichii$, la fonction coût est

$$J(b, m|\theta_i, b_{exp}, m_{exp}) = \left\| \frac{\log_{10}(b) - \log_{10}(b_{exp})}{\sigma_{log,b,exp}} \right\|^2 + \alpha \left\| \frac{m - m_{exp}}{\sigma_{m,exp}} \right\|^2 \quad (3.9)$$

où m_{exp} est le dosage final d'acétate, lactate, propionate et succinate, et $\theta_i = (\lambda_i)$. Nous avons gardé le nombre de paramètres d'ajustement petit (deux paramètres pour $L. lactis$, trois pour $L. plantarum$ et un pour $P. freudenreichii$) pour éviter la sur-prédiction. Les résultats d'optimisation sont présentés dans la Table S8.

3.4.4 Inférence des interactions métaboliques

Dans le but de prédire des échanges métaboliques comme ceux qui apparaissent dans le consortium microbien, MICOM v 0.32.0 (Diener et al., 2020) et SMETANA v1.2.0 (Zelezniak et al., 2015) ont été utilisé sur l'environnement nutritionnel du lait. En plus, la prédiction d'une complémentarité métabolique entre les trois modèles métaboliques poussant dans le lait fut révélé avec Metage2Metabo v1.5.0 (Belcour et al., 2020). Tous les métabolites prédict par les outils sont classés dans des catégories supérieures : les identifiants de la base de données Metacyc v25.5 (Caspi et al., 2018) des composés BIGG ont été retrouvé et l'ontologie de ces composés ont été utilisé. La liste des composés prédict comme intervenant dans une interaction d'alimentation croisé est disponible à la Table supplémentaire S7.

Un sous ensemble d'interaction bactérienne putative fut sélectionné parmi ceux prédict par SMETANA, selon son propre score avec un seuil choisi à 0.5. La pertinence des métabolites échangés sélectionné fut vérifié en utilisant les données de métatranscriptomiques (RPKM et moyenne des réplicats). Pour chaque métabolite, les réactions de production (dans l'espèce productrice) et de consommation (dans l'espèce consommatrice) des métabolites fut extraites en utilisant l'association gene-proteine-reactions. Dans le but d'évaluer l'expression putative de ces gènes, les valeurs d'expressions ont été comparé à l'expression globale des gènes dans chaque espèce, séparé en quartiles.

3.4.5 Visualisation et analyses statistiques

Les figures ont été générées en utilisant `matplotlib` version 3.3.4 (?). Les cartes métaboliques des voies métaboliques ont été générées avec Escher (King et al., 2015) pour Python v3.6.9 (Van Rossum and Drake, 2009)², après avoir lancé un FBA pour chaque

2. <https://escher.readthedocs.io/en/latest/escher-python.html>

modèle. Les flux des réactions d'intérêt pour chaque espèce ont été normalisé par la valeur de flux maximal des voies métaboliques ciblées de l'espèce correspondante.

3.4.6 Code et disponibilité des données

Le code pour les simulations métaboliques et les modèles métaboliques sont disponibles dans https://forgemia.inra.fr/tango/tango_models. Les génomes des trois souches bactériennes et leurs annotations sont disponibles dans l'ENA sous le numéro d'accès PRJEB54980. Les données métatranscriptomiques sont disponibles dans l'ENA sous le numéro d'accès PRJEB42478. Les génomes annotés, les lectures ARN et RPKM sont déposés dans <https://entrepot.recherche.data.gouv.fr/privateurl.xhtml?token=4bf01466-29dd-4d71-b25e-b775d9cc39dc>. Le dosage biochimiques et métaboliques sont disponibles donc (Cao et al., 2021).

Chapitre 4

CoCoMiCo : Modélisation logique de communautés bactériennes au sein d'écosystèmes non-contrôlé

Sommaire

4.1	Introduction	57
4.2	Représentation de la connaissance	58
4.3	L'approche basé sur le raisonnement pour approximer la coopération et la compétition dans des grandes communautés	59
4.3.1	Les métabolites échangés utilisés comme un indicateur de la coopération.	59
4.3.2	La consommation de métabolites limitants utilisé pour approximer la compétition.	60
4.3.3	Description des interactions potentielles dans les communautés microbienne au moyen du raisonnement.	60
4.4	Calcul des potentiels de coopérations et de compétitions	60
4.4.1	Les propriétés voulus des critères d'optimisation	60
4.4.2	Calcule du potentiel de coopération CooP	60
4.4.3	Calcul du score du potentiel de compétition ComP	61
4.4.4	Illustration sur l'exemple jouet	61
4.4.5	Implémentation	62
4.5	Analyses comparatives	62
4.5.1	Données et génération de communautés artificielles	62
4.5.2	Potentiels de coopération et de compétition à travers les différents écosystèmes	63
4.5.3	Impact de la similarité des GEMs sur les potentiels de compétition et de coopération	64
4.5.4	Impact des changements de compositions bactériens de la communauté sur les potentiels	64
4.5.5	Performance	65
4.6	Comparaison à des méthodes numériques	65
4.6.1	Méthodes	66
4.6.2	Comparaison des approches par contraintes et celle par le raisonnement	66

4.7 Application à des données réelles	67
4.7.1 Méthodes	67
4.7.2 Application aux données expérimentales et impact de la méthode de reconstruction	68
4.8 Discussion	69

soumis a CMSB (Computational methods in systems biology)

4.1 Introduction

Les communautés microbiennes sont gouvernées par des réseaux complexe d’interactions, principalement animés par la production et la consommation de biens métaboliques, et l’inférence au débit de ces interactions est crucial pour un large champs d’application (Cao et al., 2021 ; Fournier et al., 2022 ; Rousk et al., 2016 ; Wilmes et al., 2022). Les biens métaboliques échangés peuvent conduire à des effets de *coopération* positif, ou de l’alimentation croisée, où les *substrats nutritionnels* échangés impactent les organismes receveurs sans effets néfastes sur le donneur ; les effets de *compétition* négatif, où la dépendance sur les *substrats limitants* impacte négativement la production biosynthétique ou la croissance d’organismes ; les effets *neutres* (Faust and Raes, 2012).

Utiliser la modélisation métaboliques pour identifier le potentiel de coopération or de compétition entre les organismes est une voie prometteuse. Les réseaux métaboliques à l’échelle du génome inférés en utilisant les données omiques associent les réactions métaboliques et de transports aux génomes de chacun des membres d’un microbiote (Ankrah et al., 2021). En utilisant les relations réactions-protéines-gènes d’un GEM, les approches mathématiques, basées sur les graphes ou encore sur le raisonnement peuvent être appliqués pour déchiffrer le potentiel de coopération ou de compétition au sein d’une communauté microbienne. Les méthodes basées sur les graphes permettent une analyses par paires d’espèces, caractérisant les potentiels de coopération (Levy et al., 2015) et de compétition (Kreimer et al., 2012), mais ne sont pas appropriées pour des grandes communautés avec de nombreuses interactions. Les méthodes basées sur contraintes (Diener et al., 2020 ; Zelezniak et al., 2015) solutionne la limitation de l’analyse par paire pour calculer les potentiels de compétition et de coopération en utilisant une optimisation numérique et l’analyse par équilibre des flux (FBA), mais les coûts de calcul augmentent considérablement en fonction de la taille de la communauté. Malheureusement, pour les grandes communautés contenant des centaines ou des milliers de taxa différents, des approches numériques standard sont entravées par leur coûts de calcul et par leur dépendance des réseaux métaboliques curés. Dans un second temps, l’abstraction discrète de la productibilité métabolique, a été vu comme étant un bon compromis entre précision et passage à l’échelle des analyses d’espèces individuelles ou dans la sélection de communautés microbiennes (Belcour et al., 2020 ; Frioux et al., 2018).

Dans ce papier, nous étendons le travail de Frioux et ses collègues (Frioux et al., 2018), où le raisonnement logique au moyen d’une base de connaissance encodant les modèles métaboliques a été utilisé pour cribler and classifier la productibilité métabolique de grandes communautés, en particulier sur les termes de faisabilité, redondance, et de processus de coopération. La modélisation de la productibilité métabolique se prête au raisonnement machine parce que les données sont naturellement encodés comme une très grande base de connaissance de faits biologiques, et le problème lisiblement exprimé comme un ensemble de règles d’inférence directement inspirés de la connaissance des mécanismes biologiques. La Réalisation de l’inférence en utilisant le paradigme de "generer-filter" de la programmation par ensemble de réponse (ASP) (Lifschitz, 2008) a été prouvé plutôt avantageuse : il offre de solides garanties d’exactitude des résultats, une résolution efficace à l’aide de solveurs incrémentaux et, pour le raisonnement avec des données négatives, fait explicitement la distinction entre ce qui est inconnu et ce que l’on sait être faux.

En se basant sur ce raisonnement, notre extension offre un nouvel ensemble de règles

d'inférence logiques et de contraintes dérivés des définitions biologiques de coopération et de compétition. En emboitant l'algorithme d'extension de réseaux métaboliques (Ebenhöh et al., 2004), nous inférons les échanges métaboliques que chaque organisme peut être impliqué pour modéliser le potentiel de coopération ; et nous inférons les substrats qui limitent chaque organisme pour modéliser le potentiel de compétition. Ces échanges sont inférés efficacement à partir d'une grande base de connaissance encodant chaque communauté microbienne. De plus, dans le but d'utiliser les potentiels de coopération et de compétition dans le criblage de communauté, nous avons conçu des scores pour chaque qui peuvent être utilisés comme critère d'optimisation. Nous avons étalonné ces scores contre une grande collection de communautés et les avons comparés avec deux méthodes numériquement précises mais coûteuses.

4.2 Représentation de la connaissance

L'encodage judicieux de la connaissance est une préoccupation de toute approche par raisonnement effectif. Depuis les GEMs comprenant les communautés d'intérêts, nous construisons une base de connaissance en traduisant les modèles d'entrée en faits représentés par des relations logiques, définies par des ensembles de métabolites \mathcal{M} , réactions \mathcal{R} et d'un taxon \mathcal{T} . Dans le but de suivre des métabolites à travers les échanges entre les organismes, nous enrichissons l'encodage de (Frioux et al., 2018) en donnant pour chaque instance d'un métabolite ou d'une réaction d'un organisme, un unique identifiant et en les étiquetant explicitement leurs organismes donneurs, définissant la projection `source` : $\mathcal{M} \rightarrow \mathcal{T}$ et `source` : $\mathcal{R} \rightarrow \mathcal{T}$ pour extraire la source, et `name` : $\mathcal{M} \rightarrow \text{term}$ pour extraire le terme ontologique générique. N'importe quel ensemble de métabolites étiquetés (resp. réactions) peut être filtré par un taxon $\mathcal{P}_{t \in \mathcal{T}} = \{m \in \mathcal{M} \mid \text{source}(m) = t\}$.

GEMs sont des collections de réactions que nous encodons comme des relations entre les métabolites et les réactions. Les réactions réversibles sont encodées par les relations dans les deux directions. Nous adoptons une vue conservatrice stipulant qu'un métabolite produit dans un organisme est potentiellement disponible pour les autres organismes, due à la mort cellulaire et à la lyse (Fazzino et al., 2020), et parce que la sécrétion métabolique sans coût a été démontré pour être un mécanisme dominant dans les interactions microbiennes (Pacheco et al., 2019). De ce fait, nous ne contraignons pas les échanges métaboliques à ceux possédant une réaction de transport. Pour les métabolites \mathcal{M} et les réactions \mathcal{R} , un modèle

$$G \in \mathcal{M} \times \mathcal{R} \times (\mathcal{M} \times \mathcal{R} \cup \mathcal{R} \times \mathcal{M})$$

est un tuple $G = \langle M, R, E \rangle$ encodant une relation bipartite définie par les réactions $r \in \mathcal{R}$; où chaque $e = \langle m, r \rangle$ encode un *reactant* consommé par r et chaque $e = \langle r, m \rangle$ encode un *produit* produit par r . Les ensembles de réactants et de produits d'un $G = \langle M, R, E \rangle$, et l'ensemble des taxons associés à l'ensemble des métabolites, sont trouvés par :

$$\begin{aligned} \text{products}(G) &= \{m \in M \mid \langle r, m \rangle \in E\} \\ \text{reactants}(G) &= \{m \in M \mid \langle m, r \rangle \in E\} \\ \text{taxa}(M) &= \{t \in T \mid \exists m \in M, \text{source}(m) = t\} \end{aligned}$$

Comme nous considérons que tous les métabolites produits sont disponibles comme dans (Belcour et al., 2020), la combinaison de tous les modèles dans une communauté est

simplement le produit Cartésien des unions des métabolites, des réactions et des relations bipartites : la *communauté assemblé* d'une communauté $\mathbf{G} = \{G_1, \dots, G_n\}$ est

$$\text{merge}(\mathbf{G}) = \langle \bigcup_{i=1}^n M_i, \bigcup_{i=1}^n R_i, \bigcup_{i=1}^n E_i \rangle.$$

A partir d'un nouvel ensemble de métabolites *graines* représentant la composition de l'environnement, le *portée* d'un GEM est l'ensemble de métabolites qui peuvent être produits par des réactions dans le réseau à partir des grains comme reactants d'entrée. Intuitivement, il s'agit d'une fermeture transitive de la relation de réaction bipartite ; à la suite de (Handorf et al., 2005), nous la définissons ci-dessous comme le point fixe d'une fonction scope_i . La portée d'une communauté étend le concept de portée pour considérer les métabolites extra-cellulaires produits par les membres de la communauté comme des graines potentielles. Pour des raisons pratiques, (Frioux et al., 2018) a établi une distinction entre la *portée individuelle* iscope et la *portée communautaire* cscope , mais notre étiquetage explicite des entités en fonction de leur organisme source nous permet de les unifier en un seul concept. Les métabolites échanges M entre les organismes auront le même *namedans* \mathcal{M} , ignorant leur *source*. Formellement, pour $\mathbf{G} = \langle M, R, E \rangle$ et des gaines S , nous définissons :

$$\text{scope}(\mathbf{G}, S) = \bigcup_{i=0}^{\infty} \text{scope}_i, \text{ where}$$

$$\text{scope}_0 = S$$

$$\text{scope}_{i+1} = \text{scope}_i \cup \text{products}(\{r \in R \mid \text{name}(\text{reactants}(r)) \subseteq \text{name}(\text{scope}_i)\}).$$

Nous observons que les fonctions de portée de (Belcour et al., 2020) pouvait être définie par $\text{iscope}(G, S) = \text{scope}(\{G\}, S)$ et $\text{cscope}(G_1, \dots, G_n) = \text{scope}(\text{merge}(\{G_1, \dots, G_n\}), S)$

4.3 L'approche basé sur le raisonnement pour approximer la coopération et la compétition dans des grandes communautés

4.3.1 Les métabolites échangés utilisés comme un indicateur de la coopération.

Nous stipulons que les métabolites échangés entre les organismes sont la façon la plus directe pour évaluer le potentiel de coopération dans une communauté. Un métabolite qu'un organisme consomme mais qui ne peut pas le produire est opportun pour la coopération : toutes chose étant égales, une communauté pouvant disposer d'un échangé supplémentaire de ce type à un plus grand potentiel pour la coopération. Formellement, pour une communauté $\mathbf{G} = \langle M, R, E \rangle$ et une portée S , les *échanges* d'un nom générique de métabolite \hat{m} dans \mathcal{M} par rapport à (\mathbf{G}, S) sont des paires de taxons producteurs et consommateurs P et C :

$$\begin{aligned} \text{exchange}_{\mathbf{G}, S}(\hat{m}) = & \{ \langle P, C \rangle \mid \exists G_C \in \mathbf{G}, \exists G_P \in \mathbf{G}, \exists m \in \text{reactants}(G_C), \\ & \text{name}(m) = \hat{m}, \\ & m \in \text{products}(G_P), \\ & m \notin \text{scope}(G_C, S), \\ & m \in \mathcal{P}_P(\text{scope}(\mathbf{G}, S)) \} \end{aligned}$$

4.3.2 La consommation de métabolites limitants utilisé pour approximer la compétition.

La compétition peut être évaluée en considérant les métabolites qui sont consommés par plus d'un organisme. Notons qu'un métabolite échangé est un métabolite générique \hat{m} pour lequel $\text{exchange}(\hat{m}) \neq \emptyset$. Nous définissons les *substrats limitants* dans une communauté en comptant, pour chaque métabolite échangé \hat{m} , le nombre d'organisme impliqué dans les échanges. Pour une communauté $\mathbf{G} = \langle M, R, E \rangle$ et une portée S ,

$$\begin{aligned} \text{limiting}(\mathbf{G}, S) = \{ & \text{name}(m) \mid \exists m \in \text{reactants}(\text{scope}(\mathbf{G}, S)), \\ & |\text{taxa}(m)| > 1 \wedge \\ & (\text{name}(m) \in S \vee \text{exchange}_{\mathbf{G}, S}(\text{name}(m)) \neq \emptyset) \}. \end{aligned}$$

4.3.3 Description des interactions potentielles dans les communautés microbienne au moyen du raisonnement.

Comme dans (Frioux et al., 2018), les règles inférées pour les concepts décrivent au dessus sont implémentées comme un ensemble de règles en ASP. Solutionner ces règles génèrent des ensembles de réponses contenant les ensembles de métabolites qui représentent le potentiel de coopération et de compétition dans la communauté. Les réponses sont définis en utilisant un vocabulaire contrôlé qui peut être interprété par les utilisateurs et utilisatrices biologistes. Dans la section suivante nous nous reposerons sur ces réponses inférées pour calculer des scores qui peuvent être utilisés pour comparer le score de coopération et de compétition en communautés.

4.4 Calcul des potentiels de coopérations et de compétitions

4.4.1 Les propriétés voulues des critères d'optimisation

Les ensembles de réponses calculés dans la section §4.3 disent que les échanges entre les organismes sont basés sur la caractérisation de toute la communauté plutôt que sur une paire d'organisme. Dans le but d'utiliser ces réponses comme critères d'optimisation, nous concevons des résumés des scores calculés à partir des ensembles de réponses qui permet de trier les communautés de tailles comparables du plus petit au plus grand potentiel de coopération et de compétition. Ces critères devraient permettre de classer les communautés comme attendu, mais aussi de résister aux pièges des cas artificiels, tels que le remplissage de la communauté par une quantité arbitraire d'aliments croisés redondants.

4.4.2 Calcule du potentiel de coopération CooP

Le score de potentiel de coopération CooP est basé sur le comptage de taxon distinct impliqué dans chaque métabolite échangé, comme définit dans §4.3, et illustré dans la Figure 4.1A. Comme les échanges métaboliques n'ont pas de priorités, nous avons construit un compromis entre la surestimation, i.e. tous les consommateurs (resp. producteurs) contribuent de façon égale, une sous-estimation, i.e. un consommateur (resp. producteur)

contribue seul. Les taxons impliqués dans un échange sont identifiés en groupant les producteurs et les consommateurs du métabolite correspondant :

$$\begin{aligned} P(\hat{m}) &= \{t \in \mathcal{T} \mid \langle t, C \rangle \in \text{exchange}_{\mathbf{G}, S}(\hat{m})\} \\ C(\hat{m}) &= \{t \in \mathcal{T} \mid \langle P, t \rangle \in \text{exchange}_{\mathbf{G}, S}(\hat{m})\}. \end{aligned}$$

Le nombre distinct de taxon peut être ainsi calculé, $|P(\hat{m})|$ and $|C(\hat{m})|$.

Pour chaque métabolite échangé, le compromis attribut un poids w de 1 pour le premier producteur (resp. consommateur) puis ajoute un bonus sous la forme d'une décroissance exponentielle pour chaque nouveau producteur (resp. consommateur) putatif ; formellement, $w(k) = \sum_{i=0}^k 2^{-i}$.

CooP est la somme des contributions des métabolites individuels :

$$\text{CooP} = \sum_{\hat{m} \in \mathcal{M}} w(|P(\hat{m})|) + w(|C(\hat{m})|).$$

Décroître le poids de nouveaux producteurs et de consommateurs pour chaque alimentation croisés à pour but de pénaliser la redondance d'une alimentation croisée à travers un nouveau métabolite échangé, et suppose que le métabolite en question ne peut être disponible pour tous les membres de la communauté. Le vocabulaire contrôlé donne deux sous score de **CooP**, **CooP-consommateurs** et **CooP-producteurs** (Figure 4.1A) qui mesure seulement la contribution des producteurs et des consommateurs lors de l'alimentation croisés :

$$\begin{aligned} \text{CooP_producteurs} &= \sum_{\hat{m} \in \mathcal{M}} w(|P(\hat{m})|) \\ \text{CooP_consommateurs} &= \sum_{\hat{m} \in \mathcal{M}} w(|C(\hat{m})|) \end{aligned}$$

4.4.3 Calcul du score du potentiel de compétition **ComP**

Le score de compétition **ComP**(Figure 4.1A) prend en compte les consommateurs de tous les substrats **limiting**et divise la cardinalité des substrats **limiting**par la taille de la communauté. Formellement, pour une communauté \mathbf{G} et une portée S ,

$$\text{ComP} = \frac{|\text{limiting}(\mathbf{G}, S)|}{|\mathbf{G}|}.$$

4.4.4 Illustration sur l'exemple jouet

Pour illustrer l'approche, Figure 4.1B montre l'application sur deux communautés jouets de trois and quatre réseaux métaboliques (incluant le premier trois). Les deux communautés partagent les mêmes trois graines comme nutriments. Les métabolites productible avec et sans l'alimentation croisée, et ceux non productibles sont illustrés dans la figure. Le nombre de producteurs et de consommateurs impliqués dans les échanges métaboliques, et les consommateurs des substrats limitants sont utilisés pour calculer les scores. Alors que la consommation de métabolites du milieu nutritionnel est relativement similaire dans la deux communauté, la présence du quatrième réseaux métabolique crée quatre nouvelles opportunités pour l'alimentation croisée, qui résulte d'une augmentation du potentiel de coopération.

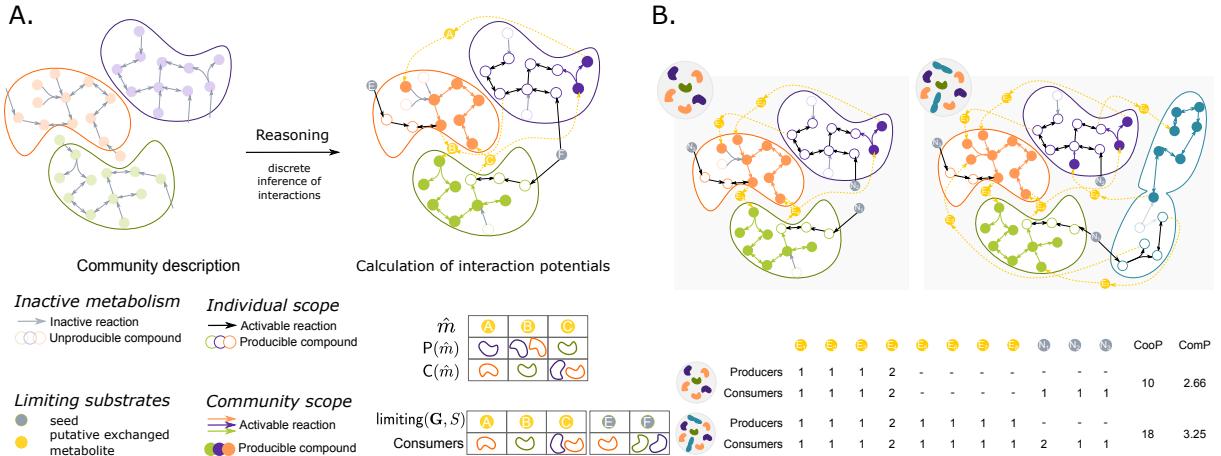


FIGURE 4.1 – Overview of the method and illustration on toy examples. (A.) Identifying exchanges and computing scores. The community is represented by a collection of metabolic networks and the description of the environment. Reasoning in Answer Set Programming permits modelling a network expansion-based model of producibility together with the putative exchanges of metabolites, limiting substrates, metabolite producibility and consumption by species. These are used to compute the cooperation potential (CoOP) and the competition potential (ComP) of the community. (B.) Illustration on two toy communities. Producible metabolites from the initial environment and from interactions with other species are used to compute CoOP and ComP.

4.4.5 Implémentation

Le modèle basé sur le raisonnement approximant la coopération et la compétition est implémenté en ASP. Ce modèle et le calcule respectif des interactions potentielles ont été évalué à travers le développement d'un package python appelé CoCoMiCo, pour *Competition and Cooperation in Microbial Communities*. Cette outil peut être installé en utilisant Pypi et son code source est disponible à <https://gitlab.inria.fr/ccmc/CoCoMiCo>.

4.5 Analyses comparatives

4.5.1 Données et génération de communautés artificielles

Les génomes de référence d'espèce cultivable associés à différentes écosystèmes sont collectés : 857 génomes de référence du sol de la base de données RefSoil (Choi et al., 2017), 1520 génomes de références du microbiote de l'intestin humain (Zou et al., 2019), 189 génomes du microbiote de la feuille (numéro accession PRJNA297956) et 1868 génomes isolés de la racine (PRJNA297942) de *Arabidopsis thaliana* (Bai et al., 2015). Tous les génomes ont été annoté en utilisant Prodigal (Hyatt et al., 2010) v.2.6.3 et eggNOG-mapper (Cantalapiedra et al., 2021) v.2.1.6 sur la base de donnée eggNOG (Huerta-Cepas et al., 2019) v.5.0.2. Les réseaux métaboliques ont par la suite été reconstruit en utilisant Pathway-tools v.25.5 (Karp et al., 2022) et mpwt v.0.7 (Belcour et al., 2020). Un milieu de culture générique consistant en des macro et micro nutriment basiques ont été utilisé pour toute les simulations, complété par un faible ensemble de métabolites courant (voir les données associées).

Cinquante communautés artificielles de taille différente allant 5 à 150 GEMs (5,10,20,30,50,75,100,150) fut créé pour chacun des écosystèmes de l'intestin, de la feuille, racine et du sol. En plus, cinquante communautés de mêmes tailles ont été créée en mixant les GEMs associés à des différents écosystèmes dans des proportions équivalentes.

4.5.2 Potentiels de coopération et de compétition à travers les différents écosystèmes

Les potentiels de coopération et de compétitions sont calculé pour toutes les communautés de GEMs artificielles (Fig. 4.2A) démontrant que les scores varient selon les écosystèmes et selon la taille de la communauté. En particulier, nous observons que les écosystèmes avec des environnements ouverts -feuille, racine, sol- montre de plus grande valeur de compétition que celui de l'intestin humain, le plus grand étant obtenu par le sol, le seul écosystème où aucun hôte est associé dans notre étude. Les distribution des potentiels de coopération et de compétition entre les écosystèmes pour toutes les tailles de communautés sont disponibles en Supp. Fig.S5. Alors que le potentiel de coopération tend à augmenter avec la taille de la communauté, le potentiel de compétition quant à lui tend à atteindre un plateau écosystème spécifique pour les tailles de communauté supérieur à 50. Seul le sol, l'environnement non lié à un hôte a montré un plus grand potentiel de coopération et de compétition par rapport aux autres habits associés à un hôte. Regarding le nombre putatif de métabolites échangés, le dernier écosystèmes montre en effet un plus grand nombre de métabolites échangeables que les trois autres (1024 ± 18 against 588 ± 33 , 693 ± 10 , 678 ± 18 pour l'intestin, la feuille et la racine pour une taille de communauté = 150). Tous ensemble, bien que les communautés de petites tailles montrent des potentiels d'interactions similaires sans tenir compte de l'écosystème, les plus grandes communautés, ressemblant plus à la taille d'écosystème naturel, sont plus distinguable. Ces observations peuvent être relaté à la diversité taxonomique des GEMs composant les communautés (Supp. fig.S4), et, les écosystèmes avec le plus de diversité taxonomique, le sol et l'intestin, montrent tous les deux respectivement le plus haut et le plus bas scores de compétition, suggérant suggérant que la taxonomie seul ne peut expliquer les interactions potentielles. D'autre part, les écosystèmes de la feuille et la racine sont similaires et ont beaucoup de taxon en commun, ce qui explique le faible écart dans leur valeurs d'interaction.

Une observation notable est que la taille de la communauté est fortement associé aux deux potentiels, mais particulièrement à la coopération (ComP : Spearman $\rho = 0.74$, $P < 2.2e-16$, CooP : Spearman $\rho = 0.80$, $P < 2.2e-16$). Ce dernier est attendu comme le nombre de voies métaboliques augmente avec l'ajout de nouveaux membre. Nous avons testé l'impact d'un écosystème sur les potentiels de compétition et de coopération en communautés de tailles similaires (Fig. 4.2B and Supp. Fig.S5). Tous ensemble, le potentiel de compétition discrimine toutes les paires d'écosystèmes, [ML – alors que seul la paire feuille-racine ne diffère pas au regard des potentiels de coopérations]. Néanmoins, la comparaison par taille de communauté montre que la coopération discrimine mieux les communautés de petites tailles. Les écosystème de la feuille et la racine sont les plus difficile à distinguer, mais les différences significatives apparaissent toujours pour des communautés de plus de cinquante GEMs.

Nous avons simulé une cinquième communauté dénoté comme *mix*, composé d'une combinaison de GEMs des quatre écosystèmes, dans le but de construire une communauté microbienne non réaliste. Nous avons observé que les communautés de l'écosystème *mix* possèdent des potentiels de compétition et de coopération significativement distinguable des autres (Fig. 4.2B) (tests Wilcoxon , P-values adjustés : $P < 0.01$) avec des scores de coopérations et de compétitions moyens respectifs de 4663 ± 115 (contre 3943 ± 65 pour l'écosystème du sol) and 330 ± 9 (versus 263 ± 5) pour les communautés de taille 150. De même, le nombre de métabolites échangeables pour la même taille de communauté montre un haut potentiel pour la communauté mix (1211 ± 31) comparé à celui du sol

(1025 ± 18). Cette observation est aussi valable pour chaque taille de communauté testée où la communauté non réaliste obtient des valeurs de potentiels de coopération et de compétition plus élevé que les quatre autres (tests Wilcoxon, P-valeurs ajustés : $P < 0.001$) (Supp. Fig. S6). Pour une communauté de taille 150, les valeurs de coopération (resp. compétition) de l'écosystème *mix* atteignent en moyenne 4663 ± 115 (330 ± 9) comparé au sol 3943 ± 65 (263 ± 5). Dans l'ensemble, ces résultats démontrent que l'on peut détecter que le métabolisme et les interactions associés à un assemblage aléatoire d'espèces diffère substantiellement du potentiel métabolique prédit d'apparaître dans des communautés réelles.

4.5.3 Impact de la similarité des GEMs sur les potentiels de compétition et de coopération

Nous avons par la suite demandé si une simple comparaison du contenu des GEMs pouvait donner des informations similaires aux potentiels de compétition et de coopération. En d'autres termes, Nous avons évalué l'association entre les potentiels interactions, se reposant sur le modèle de métabolites productibles et consommables, et sur la similarité du contenu des réseaux métaboliques à l'échelle des réactions. Les potentiels de compétition et de coopération ont été calculé pour cinquante paires de GEMs (Fig. 4.2C). Pour chaque paire de GEMs G_1 et G_2 , l'indice de similarité de Jaccard fut calculé selon l'équation 4.5.3. Nous avons observé aucune association significative entre la similarité des réseaux métaboliques et leurs potentiels (résultats des tests de corrélation en Table supplémentaire S10). Cette observation a été valide pour les quatre écosystèmes, et de ce fait, confirmant qu'un modèle discret du métabolisme communautaire prenant en compte les conditions environnementales donne un nouveau niveau d'information lors que la caractérisation de communauté microbienne.

$$Jaccard(G_1 = \langle M_1, R_1, E_1 \rangle, G_2 = \langle M_2, R_2, E_2 \rangle) = \frac{|R_1 \cap R_2|}{|R_1 \cup R_2|}$$

4.5.4 Impact des changements de compositions bactériens de la communauté sur les potentiels

La composition des communautés bactériennes change au cours du temps à la fois en terme d'abondance et en terme de la présence/ absence d'espèce. Bien que notre approche ne prend pas en compte l'abondance des microbes, des changements discrets peuvent néanmoins être considérés en altérant l'ensemble de GEMs d'une communauté. Nous avons évalué l'impact d'ajouter ou de retirer une espèce, pour les communautés de taille 3 à 150 associées aux GEMs de l'intestin.

Les résultats sont résumés en Figure 4.2D où les distributions des potentiels de coopération et de compétition dans la communauté original, la communauté après ajout d'un membre et la communauté après retrait d'une espèce sont référencées pour différentes taille de la communauté. Les comparaisons des potentiels obtenu après ajout et retrait d'espèces ont été réalisé aux respects de ceux de la communauté originel. Nous avons observé que [ML – la perturbation de la communauté a affecté les potentiels de coopération et de compétition jusque pour une communauté de taile $n = 10$. Cela indique que les substrats limitants, leur nombre de consommateur, sur lequel se repose le potentiel de compétition, ainsi que les métabolites échangeables, sur lequel s'appuie le potentiel de coopération, sont robustes face à une perturbation de la composition bactérienne.]

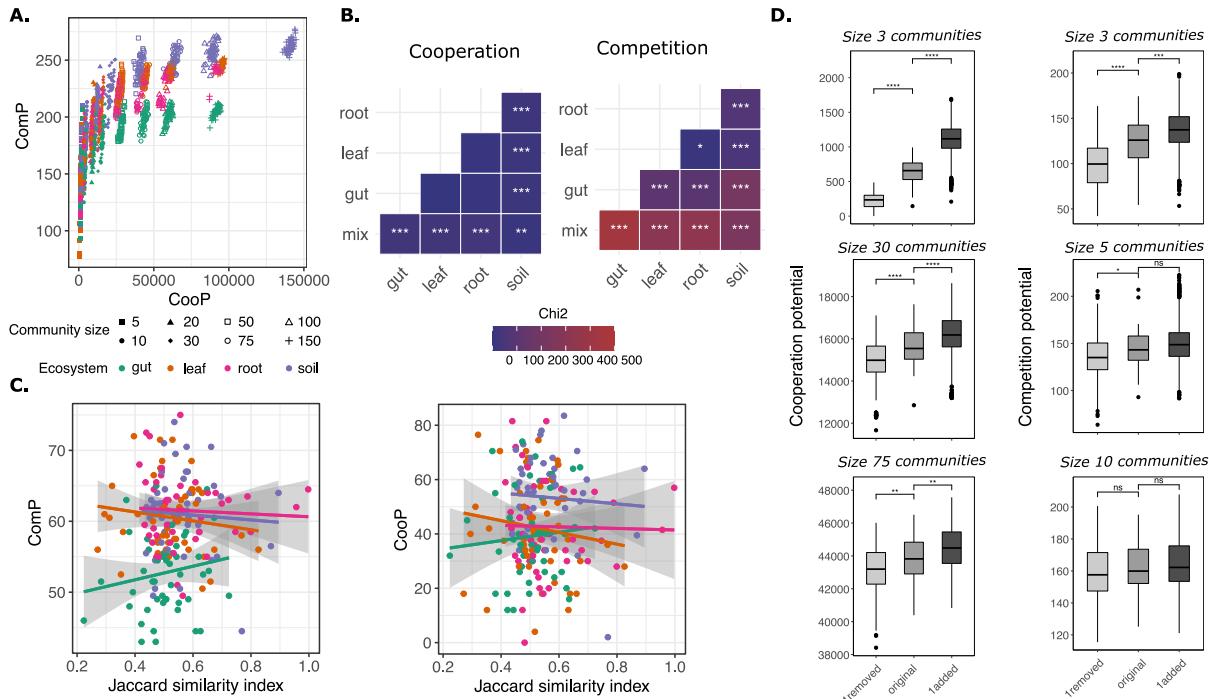


FIGURE 4.2 – Cooperation and competition potentials in diverse ecosystems. (A.) Distribution of cooperation and competition potentials in simulated communities composed of 5 to 150 cultivable bacteria isolated in the human gut, the leaf, roots or soil. (B.) Results of Kruskal-Wallis tests (χ^2 , adjusted P -value symbols) comparing cooperation and competition potentials across communities of panel A and artificial communities mixing bacteria from all ecosystems (“mix”). (*) Adjusted P -values (Benjamini-Hochberg) : $P < 0.05$, (**) Adjusted P -values (Benjamini-Hochberg) : $P < 0.01$, (***) Adjusted P -values (Benjamini-Hochberg) : $P < 0.001$. (C.) Jaccard similarity indices and competition/cooperation potentials of pairs of GSMMs from the four ecosystems. Lines are linear models drawn for each ecosystem. (D.) Effect on the cooperation and competition potentials of removing one (“1 removed”) or adding one (“1 added”) community member, according to the community size.

4.5.5 Performance

La modélisation basé sur le raisonnement permet l’analyse de données massives à grande échelle (Frioux et al., 2018). En utilisant un ordinateur personnel de 16 GiB de mémoire RAM et la puce M1 d’Apple, le temps de calcul varie de moins de deux secondes pour une communauté de taille trois à moins de trois minutes pour une communauté de taille cent cinquante. Cela démontre l’efficacité de calcul et le potentiel de passage à l’échelle du modèle par raisonnement et permet d’être utilisé comme une méthode de criblage systématique pour des grandes collections de communautés microbiennes de tailles réelles.

4.6 Comparaison à des méthodes numériques

Les méthodes numériques pour mesurer les interactions entre les membres d’une communauté microbienne se reposent sur des méthodes d’optimisation par contraintes de la distribution des flux métaboliques au sein d’un GEM. Ces méthodes sont numériquement précis mais coûtent chère en temps de calcul, et sont seulement appliquées à des communautés avec des GEMs de bonnes qualités. Notre approche par raisonnement est quant à lui, une approximation qui peut être appliquée pour des grandes communautés. Pour comparer notre méthode au standard OR fourni par les méthodes numériques, nous avons sélectionné SMETANA (Zelezniak et al., 2015) et MiCOM (Diener et al., 2020).

4.6.1 Méthodes

SMETANA calcule deux scores : Le potentiel d’interaction métabolique (MIP) calculant le nombre maximum de nutriments essentiels que la communauté peut fournir avec les échanges métaboliques entre les espèces, et le chevauchement de ressource métabolique (MRO), qui estime le chevauchement entre les pré-requis minimum pour la croissance de chaque espèce. Ces scores sont une abstraction respective de la coopération et de la compétition ; ils sont indépendants de leur habitat et reflète les limites théoriques des interactions (Machado et al., 2021). Les comparaisons avec SMETANA (v1.2.0) fut réalisés sur 250 communautés consistant de trois à trente GEMs reconstitués avec CaveMe (Machado et al., 2018) sur une collection de génomes bactériens de référence du NCBI RefSeq (release 84)¹. Le chevauchement de ressource métabolique (MRO) et le potentiel d’interaction métabolique (MIP) obtenu avec SMETANA sont comparés avec les potentiels respectifs de ComP et CooP.

MiCOM est une approche computationnelle originellement conçu spécifiquement pour la modélisation du microbiome de l’intestin humain, sur lequel il prédit le taux de croissance et de les échanges métaboliques. Les comparaison avec MiCOMM (v.0.32.0) ont été réalisé sur 188 communautés réelle de métagénomes (de taille entre seize et quatre-vingt dix huit espèces) présenté dans (Diener et al., 2020), et en utilisant un ensemble de modèle métabolique de la base de données VMH (Noronha et al., 2018). Avant l’assemblage de communauté, les réactions d’import des GEMs sont bloqués dans le but d’assurer que seul les graines seront utilisés comme nutriment pour le calcul de ComP et de CooP. Le milieu de culture *vmh_high_fiber_agora.qza* a été utilisé pour les simulations de MiCOM. MiCOM ne fournit pas directement les potentiels de coopération et de compétition, ainsi nous avons utilisés la fonction `knockout_taxa` dans MiCOM pour mesurer l’impact de la perte de la perte d’un taxon sur la croissance des autres membre de la communauté. La documentation de MiCOM dit que une diminution (resp. une augmentation) du taux de croissance après la suppression indique une sorte de compétition (resp. sorte de coopération) entre les deux espèces. Pour obtenir de potentiel de compétition (resp. coopération) nous avons sommé tous les scores d’impact positif (resp. négatif) entre les paires d’espèces d’une communauté.

4.6.2 Comparaison des approches par contraintes et celle par le raisonnement

La figure 4.3A décrit les associations entre les deux scores SMETANA et l’approximation fait par l’approche par raisonnement calculé sur 250 communautés microbiennes. De façon intéressante, une forte corrélation est observée entre les deux scores de coopération (Spearman’s $\rho = 0.8$, $P < 2.2e-16$) alors qu’aucune association significative apparait entre les scores de compétition (Spearman’s $\rho = 0.074$, $P = 0.3$). Les deux scores SMETANA sont des estimateurs des métriques d’interaction à leurs limites théoriques : ils ne s’appuient pas sur la composition nutritionnelle de l’environnement pour les prédictions, contrairement aux prédictions CooP et ComP. Cela suggère que les estimations de la compétition prenant en compte l’environnement diffèrent plus que leurs limites théoriques que la coopération.

Nous avons ensuite comparé les sortent de potentiels de la coopération et de la compétition de MiCOM à CooP et ComP dans 181 communautés 4.3B. Les deux scores de

1. https://github.com/cdanielmachado/embl_gems

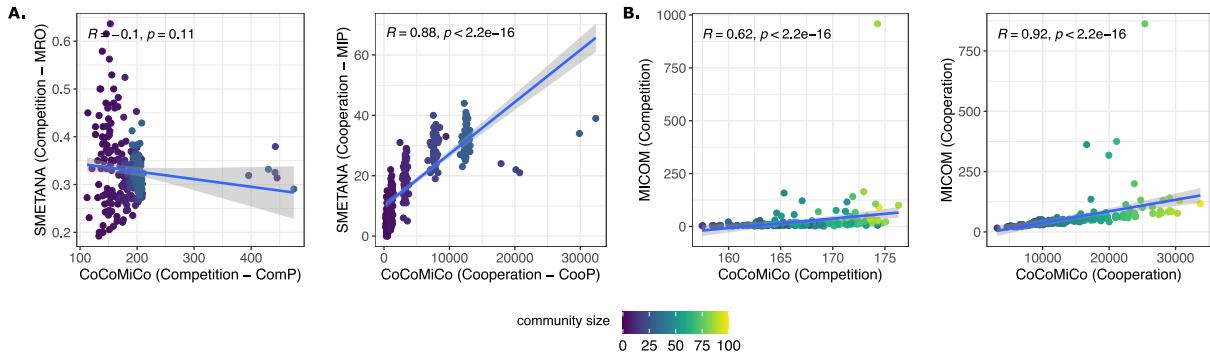


FIGURE 4.3 – Comparison of our approach predictions to SMETANA and MiCOM. (A.) Comparison to SMETANA on 250 communities of size ranging from 3 to 30. ComP is compared to SMETANA metabolic resource overlap (MRO), and CooP to SMETANA metabolic interaction potential (MIP). (B.) Comparison to MiCOM’s sums of pairwise positive and negative growth impact scores on 188 real metagenomic communities (size 16 to 98 species).

compétition et de coopération sont corrélés entre les deux outils (Spearman’s $\rho = 0.64$, $P < 2.2e-16$, et Spearman’s $\rho = 0.9$, $P < 2.2e-16$) suggérant que l’approche par raisonnement est une approximation pertinente des méthodes numériques.

Nous avons également observé que les scores SMETANAN et MiCOM sont positivement corrélés avec la taille de la communauté (SMETANA MIP : Spearman’s $\rho = 0.8$, SMETANA MRO : Spearman’s $\rho = 0.074$, $P = 0.3$, $P < 2.2e-16$, MiCOM competition : $\rho = 0.64$, $P < 2.2e-16$, MiCOM cooperation : $\rho = 0.9$, $P < 2.2e-16$). Dans l’ensemble, en dépit des différences méthodologiques majeures entre notre approche par raisonnement et les approches sur contraintes dans SMETANA et MiCOM, nous avons observé des associations significantes entre les prédictions faites par les outils, indiquant que notre méthode est pertinente pour comparer les potentiels de compétition et de coopérations entre les communautés.

4.7 Application à des données réelles

Dans le but d’améliorer l’utilisation de notre approche discrète, nous l’avons utilisé avec SMETANA et MiCOM, aux communautés bactériennes associées au Kéfir décrit dans (Blasche et al., 2021). Dans ce travail, les auteurs révèlent des interactions entre des paires de micro-organismes, dans le milieu de lait liquide ($n = 23$ souches bactériennes) et dans le milieu du lait solide ($n = 31$ souches incluant des levures). Les analyses par paires ont révélé un réseau d’interaction dense et majoritairement négatif parmi les paires d’espèce poussant sur le milieu solide. Nous avons testé les interactions potentielles dans les deux communautés au complet avec les trois outils, et évalué l’effet de la méthode de reconstruction des GEMs sur les prédictions.

4.7.1 Méthodes

Les génomes des souches bactériennes isolé du Kéfir dans (Blasche et al., 2021) ont été retrouvé et leurs GEMs reconstruits avec gapseq v1.2 (Zimmermann et al., 2021) et Pathway-tools v.25.5 (Karp et al., 2022) en utilisant mpwt v.0.7 (Belcour et al., 2020). Les modèles CarveMe (Machado et al., 2018) pré-construit par les auteurs fut également retrouvé². Deux communautés sont testées : une consistant en vingt-trois bactéries qui ont poussé deux-à-deux dans le milieu du lait liquide dans (Blasche et al., 2021), et

2. https://github.com/cdanielmachado/kefir_paper

	community	Pathway Tools Cooperation	Pathway Tools Competition	CarveMe Cooperation	CarveMe Competition	gapseq Cooperation	gapseq Competition
CoCoMiCo	Milk-liquid	1090.68 (311)	129.48	2183 (625)	150.91	2039.34 (571)	234.83
	Milk-solid	1098.85 (302)	130.70	2651.19 (758)	163	2039.39 (547)	244.53
SMETANA	Milk-liquid	-	-	22 (59)	0.329	NA	0.682
	Milk-solid	-	-	22 (66)	0.348	NA	0.693
MICOM	Milk-liquid	-	-	24.75 (99)	2.75	22.435 (95)	22.432
	Milk-solid	-	-	34.80 (109)	3.8	31.476 (103)	31.454

TABLE 4.1 – **CoCoMiCo, SMETANA, and MICOM applied to the kefir experimental data from (?)**. Co-operation and competition potentials are obtained for the two communities grown either on liquid milk medium ($n = 23$ strains) or solid milk medium ($n = 31$ strains) (?) using metabolic networks reconstructed with Pathway Tools, CarveMe and gapseq, and analysed with MICOM, SMETANA and CoCoMiCo. As metabolic networks obtained with Pathway Tools do not possess a biomass reaction by default, only CoCoMiCo can assess their competition and cooperation potentials. SMETANA produced no results for the gapseq reconstructed GSMNs, indicating that at least one member failed to grow in the community model [ML – Faire un tri sur l’identifiant du métabolite, cad, en enlevant le préfixe et le suffixe (M_ et _e ou c)].

une consistant en trente et une souches intialement mis en culture deux-à-deux dans le milieu solide contenant exactement les mêmes nutriments du lait. Nous avons prédits les interactions au sein des deux communautés en utilisant SMETANA (v.1.2.0), MiCOM (v.0.32.0) et CoCoMiCo ([ML – version à changer]0.2.0). à partir des nutriments du lait. Le score unique de coopération et de compétition pour MiCOM ont été retrouvé comme décrit plus haut. Pour MiCOM (grow workflow), SMETANA et notre approche, le nombre de métabolites échangeables ont été plus retrouvé.

4.7.2 Application aux données expérimentales et impact de la méthode de reconstruction

Les potentiels d’interactions obtenus avec les trois outils et reconstruits avec les trois méthodes de reconstruction sont présentes dans la Table 4.1. Une première observation est qu’il n’était pas possible d’appliquer SMETANA ou MiCOM sur les GEMs reconstruits avec Pathway-tools, comme le protocole que nous utilisons ne donne pas de réaction de biomasse pour les réseaux. Tous les outils prédisent une augmentation du potentiel de coopération dans les plus grandes communautés, ce qui peut être expliqué en partie par l’association positive des scores de coopération et la taille de la communauté. Le nombre absolu de composés échangeables qui ont été prédits par l’approche de raisonnement a identifié plus de métabolites échangeable uniques dans les plus petites communautés reconstruit avec Pathway-tools et gapseq. Cela peut-être relié au réseau d’interaction coopérative dense observé dans les cultures par paires pour cette communauté (Blasche et al., 2021). Le nombre de composés possiblement échangés prédit par CoCoMiCo surpasse largement les nombres prédits par les simulations numériques, ce qui est attendu puisque la méthode conservative identifie toutes les interactions potentielles, sans une optimisation numérique de la biomasse, qui pourrait contraindre les échanges possibles. Les prédictions de la compétition sont plus élevé pour les communauté plus grandes, ce qui est conforme avec les résultats des interactions deux-à-deux des expériences en laboratoire, bien que les différences sont souvent négligeables.

Ces résultats montrent la versatilité de notre approche, qui peut être appliquée à des communautés où les GEMs n'ont pas de réactions de biomasse. Les résultats souligne également l'importance du choix de la méthode de reconstructions des GEMs : le fait de s'assurer que chaque modèle individuel peut simuler la croissance sur un milieu typique conduit à des prédictions similaires de scores d'interaction, alors que les GSMN non curés, obtenus avec Pathway Tools, présentent des potentiels dissemblables.

4.8 Discussion

Nous proposons une approche basé sur le raisonnement inférant le potentiel de coopération et de compétition dans des communautés microbiennes modélisé par les réseaux métaboliques à l'échelle du génome. Des travaux précédent sur la sélection de communauté (Belcour et al., 2020 ; Frioux et al., 2018), inspiré de l'approche de modélisation métabolique discret de (Ebenhöh et al., 2004), ont utilisé le raisonnement sur une large base de connaissance encodant les modèles métaboliques. Le travail que nous présentons ici est une extension de ce dernier en enrichissant la représentation de la connaissance et en ajoutant des règles permettant l'inférence du potentiel de coopération et de compétition dans une communauté microbienne. Notre approche combine la rigueur et la précision de la modélisation logique avec un calcul rapide à l'échelle de grandes communautés.

Les interactions, identifiées comme des échanges métaboliques par notre méthode, peuvent être interprétés en détails ou utilisés comme une fonction à optimiser. Pour faciliter cette tâche, l'ensemble des échanges sont par la suite combiné en des scores numériques décrivant quelles sont les communautés avec le plus haut degré de coopération positive et compétition négative. Ces scores rangent ces communautés selon leur potentiel et permet d'éviter les pièges des communautés artificielles. Il est important de noter que les scores définis ici ne sont comparables que pour des communautés de taille similaire ; ils ne sont délibérément pas normalisés afin d'éviter les risques de comparaisons inappropriées.

L'évaluation *a priori* de la mesure dans laquelle une communauté microbienne est susceptible d'interagir, soit par la compétition pour les ressources, soit par l'alimentation croisée et le mutualisme, est particulièrement précieuse pour le criblage à haut débit des candidats lors de la conception des consortiums, pour modifier la composition des communautés afin de promouvoir un état d'intérêt, ou simplement pour mieux comprendre les règles de l'assemblage des microbes (Oliveira and Pamer, 2023). Cependant, la tâche n'est pas triviale, car les questions liées à la fonction d'objectif de fitness des communautés (Diener and Gibbons, 2023) ou à l'importance de la curation automatique du GEMs et du comblement des lacunes (Bernstein et al., 2019) persistent. L'utilisation d'une modélisation discrète sans fonction objective numérique réduit la nécessité d'une curation d'un GEM et convient à l'analyse d'organismes non cultivés. Elle peut donc constituer une première tentative peu coûteuse de modélisation des communautés microbiennes, une solution pour identifier les communautés qui pourraient le mieux bénéficier d'une curation supplémentaire, et un moyen de mettre en avant et de comparer les métabolites échangés et les substrats limitants entre les communautés.

En appliquant la méthode à une large collection de communautés simulées venant de divers écosystèmes, nous avons relevé des différences dans les potentiels de coopération et de compétition. En particulier, notre observation que les communautés du sol montrent de plus grand scores de compétition comparé aux communautés de l'intestin est corroboré par le travail de (Machado et al., 2021) qui observe plus de compétition dans les système vivant sans hôtes que ceux avec un hôte. Les comparaisons à grande échelle à SMETANA

(Zelezniak et al., 2015) et MiCOM (Diener et al., 2020) ont montré que notre méthode offre un pouvoir sélectif comparable à celui de ces approches numériques de référence, tout en s'adaptant à des communautés de milliers d'espèces en interaction. Cette évaluation souligne la nécessité de modéliser les interactions de plusieurs à plusieurs plutôt que par paire pour les membres de la communauté. Nous avons démontré que les potentiels d'interaction ne sont pas prédictifs par la seule topologie du réseau, mais qu'ils sont en corrélation avec les prédictions numériques basées sur les flux, en particulier dans le cas de la coopération.

Les scores d'interaction sont fortement associés à la taille de la communauté. Ceci est particulièrement vrai pour le potentiel de coopération, bien que nous ayons illustré que ce score peut discriminer entre les écosystèmes et qu'il est complémentaire de la compétition à cette fin. Cette association a été observée non seulement pour notre approche, mais aussi pour SMETANA (Zelezniak et al., 2015) et MICOM (Diener et al., 2020), ce qui confirme l'hypothèse selon laquelle l'effet de taille est un facteur déterminant des potentiels d'interaction lorsque l'on compare des communautés de tailles très différentes.

En comparant les potentiels d'interaction prédictifs à partir des GEMs construits avec trois méthodes de reconstruction, nous avons mis en évidence l'importance du choix entre une méthode qui assure la croissance individuelle de chaque espèce, mais qui ne peut guère être appliquée aux organismes non cultivables fréquemment traités en métagénomique, et une méthode qui ni ne cure d'avantage des GEMs ni ne déduit de réaction de biomasse. Ce choix incombe aux modélisateurs, mais nos résultats démontrent la difficulté de comparer les prédictions pour des communautés qui ne sont pas construites à l'aide des mêmes méthodes.

Notre travail améliore l'état de l'art en fournissant une méthode économique pour évaluer simultanément la coopération et la compétition entre tous les membres d'une communauté microbienne, sans nécessiter une modélisation numérique coûteuse de l'abondance relative des organismes dans ces communautés ou la spécification d'une fonction de biomasse. Parmi les outils testés ici, seul MiCOM tient compte de l'abondance. Nous considérons la prise en compte de l'abondance des espèces comme une fonctionnalité future de CoCoMiCo, modélisée non pas numériquement mais en utilisant une logique descriptive des niveaux et des trajectoires d'abondance. D'autres extensions de ce travail incluent l'utilisation des potentiels de coopération et de compétition pour sélectionner des communautés d'intérêt au sein d'une collection d'espèces, afin de guider l'assemblage des communautés dans des applications biotechnologiques ou sanitaires d'intérêt.

Chapitre 5

Enrichissement des modèles qualitatifs : Modèle de logique linéaire temporel et processus de sélection de communauté à façon

Sommaire

5.1	SelCom : Sélection de communauté basé sur des scores	72
5.2	LTL ; intégration temporelle	72

5.1 SelCom : Sélection de communauté basé sur des scores

l'idée est de pouvoir sélectionner des communautés a façon selon des critères de taille, de composition bactérienne, de l'étendu métaboliques à atteindre, du choix de l'optimisation et enfin, de l'objectif à optimiser. Cette sélection vise dans un premier temps à élaborer des hypothèses testables en laboratoire, et d'identifier des différences métaboliques entre les différentes communautés. Pour effectuer cette sélection, nous avons utiliser l'ASP, diminutif de Answer set programming, connu pour résoudre des problèmes combinatoires important [ML – REF]. L'asp décrit le problème initial en une liste de fait et le résout en écrivant des contraintes logiques. Puis, le solveur logique propose des modèles qui satisfont l'ensemble des contraintes précédents [ML – figure ASP]. Dans le cas de la sélection de communauté bactérienne, la liste minimale des faits nécessaire à la résolution de ce problème est constituée de l'ensemble des réactions et métabolites des réseaux métaboliques et de l'ensemble des nutriments disponibles dans le milieu extra-cellulaire.

Chaque critère de sélection est pensé pour accepter un vocabulaire précis répondant à une demande biologique. Concernant la taille des communautés sélectionnées, l'utilisateur ou l'utilisatrice peut fixer une taille précise , de borner partiellement ou complètement les tailles possibles ou encore de ne pas mettre de contraintes. En plus, la ou le biologiste peut filtrer la composition bactérienne de sa communauté via des connecteurs logiques : ET, NOT, COMBINAISON. La ou le biologiste est ainsi libre de remplir partiellement ou non l'ensemble de ces contraintes. Avec le connecteurs logique "ET", le programme interprète comme une obligation la présence du rang taxonomique que l'utilisateur ou l'utilisatrice indiquera. "NOT" fonctionne comme le "ET" mais pour des bactéries que les communautés sélectionnées ne doivent pas avoir. Enfin, "COMBINAISON" est une contrainte plus forte et précise si la combinaison entre deux rangs taxonomique d'une espèce doivent être retrouvés ou à l'inverse, ne doivent jamais être retrouvés parmi les communautés sélectionnées. Ces contraintes sont expliqués sous forme un fichier tabulé.

Le logiciel convertie ainsi le fichier YAML et le fichier de contraintes TSV en une liste de fait et de contraintes logiques.

5.2 LTL ; intégration temporelle

Chapitre 6

Approche de modélisation hybride, [*ML – pas vraiment dans le plan*]

Sommaire

6.0.1	Approches hybrides découplées et couplées	74
-------	---	----

6.0.1 Approches hybrides découplées et couplées

L'approche de modélisation hybride développé dans cette thèse à été développé sous de multiples angles. La première est la création d'une approche hybride découpé, avec les projets TANGO et CoCoMiCo. Chacun des deux permettent de caractériser une communauté bactérienne en mettant en avant les interactions entre bactéries. TANGO et CoCoMiCo apportent une information supplémentaire et complémentaire sur la communauté bactérienne. Ces deux approches sont complémentaires du fait du raisonnement utilisé dans le modèle logique, le criblage permet de mettre en avant l'ensemble de interactions métaboliques possibles en fonction de l'environnement nutritionnel et peut ainsi servir de support pour inférer les éléments à suivre dynamiquement dans le modèle dynamique (modèle hybride couplé). La modèle dynamique quand à lui, permet à partir d'une base de fait établie, issu de la connaissance biologique, de la littérature ou bien d'un logiciel de criblage, de décrire avec précision des mécanismes de régulation et entre autre, des interactions bactériennes.

En plus de la fiabilité des éléments criblées à vérifier, réduire leur nombre ainsi que obtenir un population microbienne minimale est nécessaire afin d'utiliser des systèmes dynamiques tels que celui de TANGO. En effet, plus le nombre de paramètres à suivre augmente plus le temps de calcul croit. Plusieurs questions subviennent alors quand aux critères de sélection permettant de mettre en place un filtre efficace. Dans (Frioux et al., 2018), les auteurs proposent de sélectionner une communauté bactérienne basée en minimisant la taille de la communauté. En fonctionnant ainsi, minimiser la taille ne permet pas de garantir que la communauté finale est la meilleure pour répondre à un objectif. En effet, il permet de résoudre la combinatoire du nombre de communautés bactérienne que le programme résout.

Chapitre 7

Annexes

.1 Figures supplémentaires

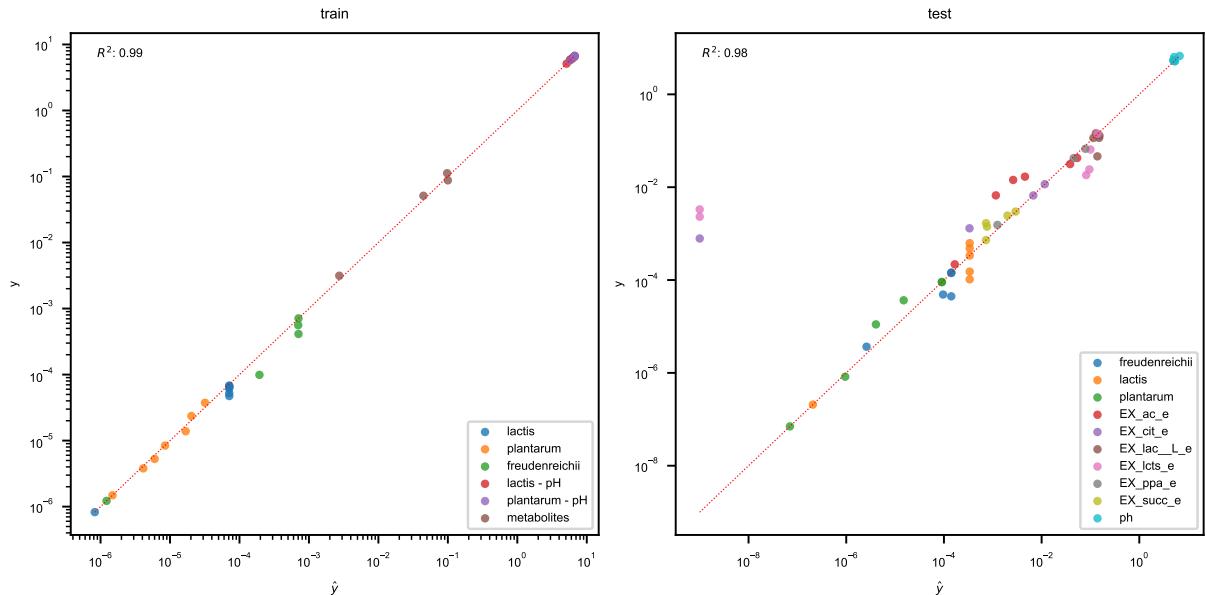


FIGURE S1 – **Model goodness of fit.** All the model outputs (\hat{y}) are plotted versus their respective data value (y) in the training (i.e. the mono-culture data, left panel) and in the testing set (i.e. the co-culture data, right panel). Each point is colored according to the data type and the bissector is plotted (red dotted line). A global coefficient of determination is computed with the formula $R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$, where \bar{y} is the average value of y . We can observe a very good agreement of the model in the training set, and a globally good agreement in the testing set, with 3 outliers on lactose and citrate.

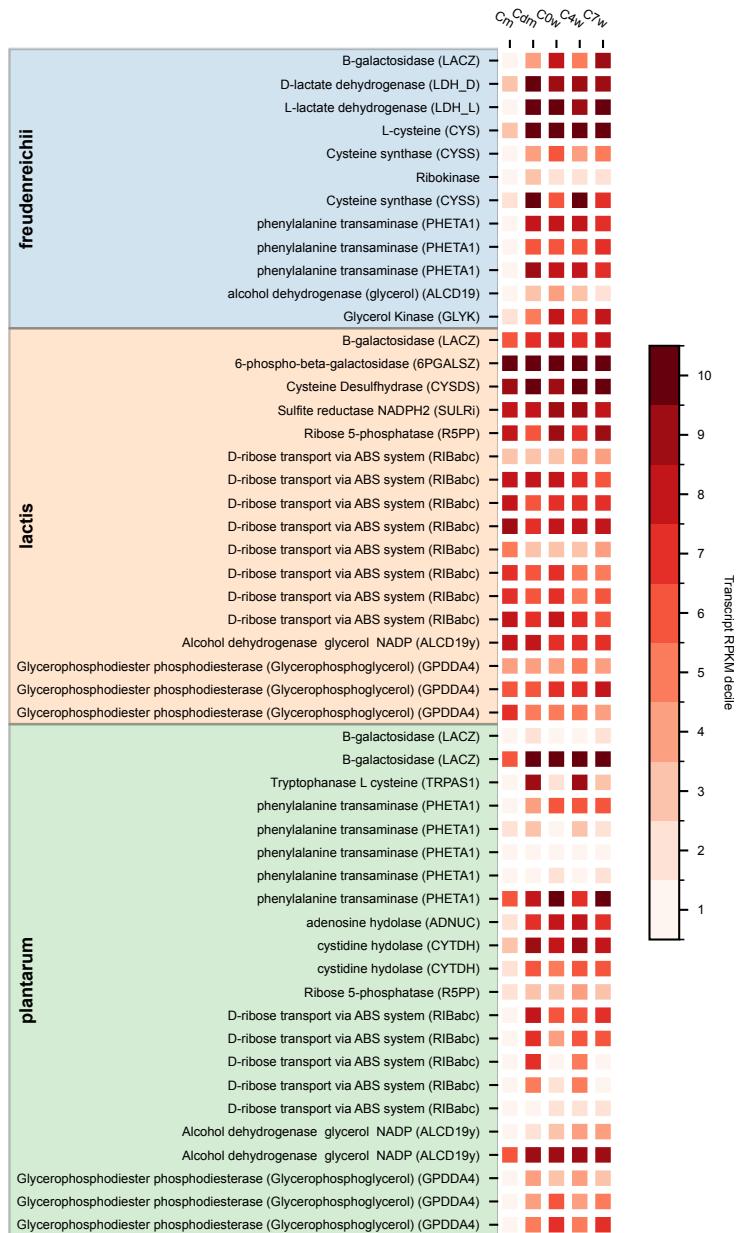


FIGURE S2 – Gene expression with metatranscriptomic data. This heatmap displays the decile of the RPKM of gene transcript counts at a given time point (columns) and in a given micro-organism (the colored areas indicate to which bacteria the gene belongs). Then, the decile were color coded from highest deciles (dark red, highest expression in this micro-organism at this time point) to lowest decile (light red, low expression). The displayed genes were manually selected from the list of possible interactions raised by SMETANA : for a selected interaction metabolite, genes involved in production (in the donor), and in consumption (in the receiver) were retained for analysis. Selected interaction metabolites were H₂S, ribose and phenylalanine. Additionally, genes involved in lactose consumption were added to this list. Abbreviations : C_m, molding (t = 19.5 hours); C_{d,m}, demolding (t = 40 hours); C_{0,w}, start of ripening (t = 60 hours); C_{4,w}, Fourth week ripening (t = 732 hours); C_{7,w}, seventh week ripening (t = 1236 hours).

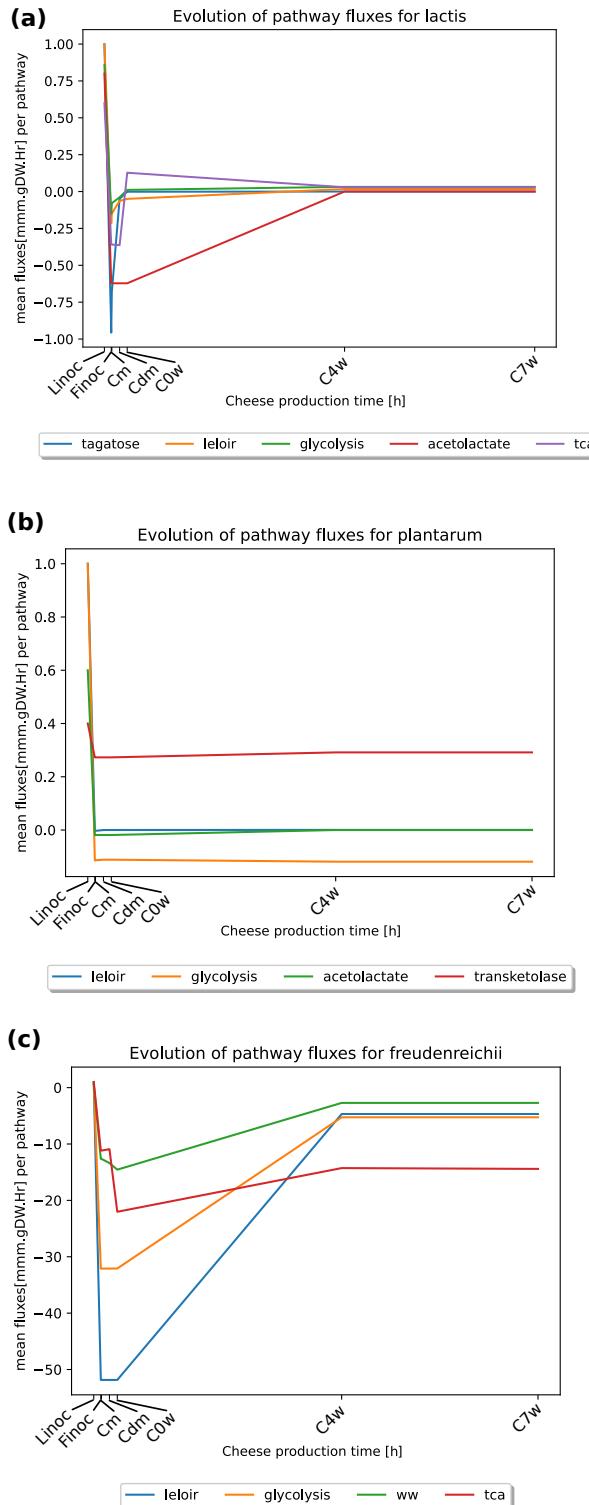


FIGURE S3 – Metabolic pathway switches during cheese production. For each bacterium we ran a dFBA simulation and reaction fluxes per each pathways of interest in Figure 3.1 were retrieved at the inoculum of LAB (Linoc, $t=0$ hours), the inoculum of *P. freudenreichii* (Finoc, $t=18$ hours) the molding (Cm, $t=19.5$ hours), demolding (Cdm, $t=40$ hours), start of ripening (C0w, $t=60$ hours), fourth week ripening (C4w, $t=732$ hours) and seventh week ripening state (C7w, $t=1236$ hours). We normalized fluxes by the inoculum flux value and the average flux for each pathways is therefore represented.

.2 Matériel supplémentaire

.2.1 Pure culture data

.2.1.1 Strain pure culture data

GEM	Time	Bacterial density	pH
<i>P. freudenreichii</i>	0	9.405e-07	6.76
	8	1.221e-06	6.76
	31	9.9e-05	6.55
	48	0.000561	6.22
	56	0.0004125	6.1
	130	0.0007095	5.7
<i>L. plantarum</i>	0	1.485e-06	6.7
	5	3.795e-06	6.68
	7	5.28e-06	6.66
	9	8.415e-06	6.645
	14	1.386e-05	6.555
	16	2.3595e-05	6.54
	79	3.762e-05	6.705
<i>L. lactis</i>	0	8.25e-07	6.7
	5	5.247e-05	6.49
	7	6.831e-05	6.315
	9	6.666e-05	6.075
	14	6.468e-05	5.935
	16	4.752e-05	8.865
	79	6.171e-05	5.115

TABLE S1 – Pure culture data used for the dynamic calibration of each strain

.2.1.2 Acids dosage for *P. freudenreichii*

Compounds	Concentration T_0	Concentration $T_f = 89$ h	Std
Lactate	16.5	7.88	0.08
Acetate	0	3.07	0.02
Succinate	0	0.371	0.063
Propionate	0	88.31	0.02

TABLE S2 – Acids dosage data in g/L for *P. freudenreichii*.

2.2 Raffinement manuel des GEMs

Pour assurer la production extracellulaire des composés, nous avons ajouté des modifications sur les bornes intracellulaires de réactions. Ces modification sont récapitulé dans la Table S3.

GEM	Reaction	Bound	Value	Justification
<i>L. lactis</i>	ACTD2	lower upper	0 0	Allows the activation of the acetolactate pathway
	ACLDC	lower upper	0 2	Allows the activation of the acetolactate pathway
	ACLS	lower upper	2 1000	Allows the activation of the acetolactate pathway
	LACZ	lower upper	0.0006 2	Force the consumption of lactose
<i>L. plantarum</i>	PFL	lower upper	0 0	Permit the production of both isomers of lactate from lactose.
	LCTSt3ipp	lower upper	0 0	Leads the flux of lactose into the Glycolysis pathway
	ACTD2	lower upper	0 0	Allows the activation of the acetolactate pathway
	GALUi	lower upper	-30 10	Permit the production of both isomers of lactate from lactose.
	ACLS	lower upper	2 1000	Allows the activation of the acetolactate pathway
<i>P. freudereichii</i>	6PGALSZ	lower upper	0 0	Tagatose pathway was used
	XYLI2	lower upper	0 0	Production of fructose from glucose instead of glucose-6-phosphate. Glycolysis was not activated
	POX2	lower upper	0 0	Regulate flux of acetate
	PPAKr	lower upper	0 0	Inhibit acetate over production
	LACZ	lower upper	0 5	Allow consumption of lactose
	PTAr	lower upper	-1000 8	Allow the appropriate ratio of acetate/propionate
	2131pyrpp	lower upper	8 1000	Allow the appropriate ratio of acetate/propionate

TABLE S3 – Manual refinement of the GEMs and corresponding justifications.

2.3 Données expérimentales communautaires

GEM	steps	Time (h)	Bacterial population (\log_{10} CFU/g)
<i>P. freudenreichii</i>	Linoc	0	6.1
	Finoc	18	6.1
	Cm	19.5	7.16
	C0w	60	8.11
	C4w	732	8.56
	C7w	1236	8.59
<i>L. plantarum</i>	Linoc	0	5.2
	Finoc	18	5.4
	Cm	19.5	6.49
	C0w	60	7.92
	C4w	732	8.47
	C7w	1236	8.47
<i>L. lactis</i>	Linoc	0	5.7
	Finoc	18	7.5
	Cm	19.5	8.78
	C0w	60	9.06
	C4w	732	8.95
	C7w	1236	8.45

TABLE S4 – Co-culture growth data used for testing community model predictions

2.4 Exchanged metabolites highlighted by computational approaches

2.4.1 MICOM

Bigg ID	Metacyc ID	Name	Metacyc Ontology	Export fluxes	Import
acald_e	ACETALD	acetaldéhyde	Aldehydes-Or-Ketones, Aldehydes	Pf; Ll	Lp
akg_e	2-KETOGLUTARATE	2-oxoglutarate	Acids, Organic-Acids	Pf	Lp
ala_D_e	D-ALANINE	D-ALANINE	All-Amino-Acids, Acids, Amino-Acids, Organic-Acids	Pf; Lp	Ll
ala_L_e	L-ALPHA-ALANINE	L-ALPHA-ALANINE	All-Amino-Acids, Acids, Amino-Acids, Organic-Acids	Pf; Lp	Ll
co2_e	CARBON-DIOXIDE	CARBON-DIOXIDE	Others, Others	Pf; Ll	Lp
coa_e	CO-A	co enzyme A	Groups, Others	Pf; Ll	Lp
fe2_e	FE+2	Fe2+	Ions, Inorganic-Ions, Cations	Lp	Pf;
gly_e	GLY	Glycine	All-Amino-Acids, Acids, Amino-Acids, Organic-Acids	Lp	Pf;
glyclt_e	GLYCOLATE	GLYCOLATE	Acids, Organic-Acids	Pf	Lp
gthox_e	OXIDIZED-GLUTATHIONE	glutathione disulfide	ORGANOSULFUR, All-Glutathiones	Lp	Ll
gthrd_e	GLUTATHIONE	glutathione	ORGANOSULFUR, All-Glutathiones, Thiols	Ll	Lp
h2o2_e	HYDROGEN-PEROXIDE	hydrogen peroxide	Peroxides, Others	Ll	Lp
h2o_e	WATER	H2O	Pseudo-Compounds, Others	Pf; Ll	Lp
h2s_e	HS	hydrogen sulfide	Ions, Anions, Inorganic-Ions	Ll	Lp;
ile_L_e	ILE	isoleucine	All-Amino-Acids, Acids, Amino-Acids, Organic-Acids	Lp	Pf;
lac_D_e	D-LACTATE	d-lactate	Acids, Organic-Acids	Pf	Ll;
lac_L_e	L-LACTATE	l-lactate	Acids, Organic-Acids	Ll	Lp;
mal_L_e	MAL	malate	Acids, Organic-Acids	Pf; Lp	Ll
nh4_e	AMMONIUM	ammonium	Ions, Inorganic-Ions, Cations	Lp	Pf;
phe_L_e	PHE	phenylalanine	Aromatics, All-Amino-Acids, Acids, Amino-Acids, Organic-aromatic-compounds	Pf	Lp
pro_L_e	PRO	proline	All-Amino-Acids, Acids, Amino-Acids, Organic-Acids	Ll	Pf;
ser_D_e	D-SERINE	serine	All-Amino-Acids, Acids, Amino-Acids, Organic-Acids	Ll	Lp;
ser_L_e	SER	serine	All-Amino-Acids, Acids, Amino-Acids, Organic-Acids	Ll	Lp;
succ_e	SUC	succinate	Acids, Organic-Acids	Ll	Pf;
xan_e	XANTHINE	xanthine	Others, Others	Ll	Lp;

TABLE S5 – Exchangeable metabolites highlighted by MICOM on the cheese bacterial community. Abbreviations : Ll, *L. lactis*; Lp, *L. plantarum*, Pf, *P. freudenreichii*

2.4.2 SMETANA

Bigg ID	Metacyc ID	Name	Metacyc Ontology	Export fluxes	Import fl
acald_e	ACETALD	acetaldéhyde	Aldehydes-Or-Ketones, Aldehydes	Pf Pf Lp Ll	Ll Lp Ll Lp
akg_e	2-KETOGLUTARATE	2-oxoglutarate	Acids, Organic-Acids	Lp Pf	Pf Lp
fum_e	FUM	fumarate	Acids, Organic-Acids	Pf	Lp
glyc_e	GLYCEROL	GLYCEROL	Alcohols, All-Carbohydrates, Sugar-alcohols, Carbohydrates	Lp Li Ll	Pf Pf Lp
glyclt_e	GLYCOLATE	GLYCOLATE	Acids, Organic-Acids	Lp	Pf
h2s_e	HS	hydrogen sulfide	Ions, Anions, Inorganic-Ions	Lp Pf Ll	Pf Lp Lp
lac_D_e	D-LACTATE	d-lactate	Acids, Organic-Acids	Pf Ll	Lp Lp
lac_L_e	L-LACTATE	l-lactate	Acids, Organic-Acids	Ll	Lp
mal_L_e	MAL	malate	Acids, Organic-Acids	Pf Lp Pf Ll	Ll Ll Lp Lp
phe_L_e	PHE	phenylalanine	Aromatics, All-Amino-Acids, Acids, Amino-Acids Organic-Acids, Organic-aromatic-compounds	Lp	Pf
rib_D_e	D-Ribopyranose	D-Ribopyranose	Aldehydes-Or-Ketones, All-Carbohydrates, Aldehydes, Carbohydrates	Ll Lp Pf Ll	Pf Pf Lp Lp
ser_D_e	D-SERINE	serine	All-Amino-Acids, Acids, Amino-Acids, Organic-Acids	Ll Lp Pf Lp Pf Ll	Pf Pf Ll Ll Lp Lp
succ_e	SUC	succinate	Acids, Organic-Acids	Ll Lp	Pf Pf
xan_e	XANTHINE	xanthine	Others, Others	Ll Lp	Pf

TABLE S6 – Exchangeable metabolites highlighted by SMETANA on the cheese bacterial community. Abbreviations : Ll, *L. lactis*; Lp, *L. plantarum*, Pf, *P. freudenreichii*

2.4.3 Metage2Metabo

Bigg ID	Metacyc ID	Name	Metacyc Ontology
galt1p benzcoa xtsn	D-galactose-1-phosphate BENZOYLCOA XANTHOSINE	a D-galactopyranose 1-phosphate benzoyl-CoA XANTHOSINE	All-Carbohydrates,Carbohydrates Esters,Thioesters Organic-heterocyclic-compound,All-Nucleosides,Nitrogen-Molecular-Entities,Organic-heteromonocyclic-compounds,Organic-Heteropolycyclic-Compounds,Organonitrogen-Compounds,Nucleosides,Aldehydes-Or-Ketones,All-Carbohydrates,Aldehydes,Carbohydrates
glyald 3hcoa 3hhdcoa 3hdcoa 3odcoa 3hhcoa udepp 3oooca	GLYCERALD ∅ CPD0-2232 CPD0-2244 CPD0-2123 OH-HEXANOYL-COA UNDECAPRENYL-P CPD0-2106	glyceraldehyde ∅ (S)-3-hydroxyhexadecanoyl-CoA (S)-3-hydroxydecanoyl-CoA 3-oxodecanoyl-CoA (S)-3-hydroxyhexanoyl-CoA all-trans-undecaprenyl phosphate 3-oxooctanoyl-CoA	Organic-heterocyclic-compound,All-Nucleosides,Nitrogen-Molecular-Entities,Organic-heteromonocyclic-compounds,Organic-Heteropolycyclic-Compounds,Organonitrogen-Compounds,Nucleosides,Aldehydes-Or-Ketones,All-Carbohydrates,Aldehydes,Carbohydrates ∅ Thioesters,Esters Thioesters,Esters Thioesters,Esters Thioesters,Esters Lipids,Polyisoprenoids Thioesters,Esters

TABLE S7 – Metabolites predicted to be producible through metabolic cross-feeding only by Metage2Metabo

2.5 Bornes dynamiques sur les réactions d'échanges

Dans cette section, nous récapitulons et justifions les fonctions de régulation modulant les métabolites échangés. Pour chaque GEM, les réactions d'échanges sont listées et leurs

bornes inférieurs et supérieurs sont spécifiées.

2.5.0.1 *L. lactis*

- EX_lcts_e :

- inférieur : $\max\left(-\frac{m_{lcts_e}}{\Delta t * \sum_{i \in \mathcal{M}(lcts_e)} b_i}, -\mu_{max,lcts} * 10^{(-k_{lac} * \phi_{undiss})} - \mu_{min,lcts}\right)$ (1)

- supérieur : 1000

Justification : Quand le lactose est épuisé, le premier terme est activé, modélisant le partage uniforme du lactose disponible parmi l'ensemble de bactéries $\mathcal{M}(lac_L)$ métabolisant le lactose. Autrement, une régulation par le pH est ajoutée : la quantité de molécules d'acide lactique non dissocié est calculé à partir de la concentration de lactate au moyen de la fonction ϕ_{undiss} (voir (3.7) pour l'expression de cette fonction) ; puis, avec une croissance exponentielle à partir de la valeur $-\mu_{max,lcts}^{lact} - \mu_{min,lcts}^{lact}$ à la valeur $-\mu_{min,lcts}^{lact}$ quand la concentration d'acide lactique non dissocié augmente, avec un taux exponentiel k_{lac} . Les valeurs de $\mu_{max,lcts}^{lact}$, $\mu_{min,lcts}^{lact}$ et k_{lac} sont inférées des expériences de co-cultures.

- EX_lac_L_e, EX_lac_D_e :

- inférieur : 0

- supérieur : $\min\left(\frac{m_{lcts_e}}{\Delta t * \sum_{i \in \mathcal{M}(lcts_e)} b_i}, \mu_{max,lcts} * 10^{(-k_{lac} * \phi_{undiss})} + \mu_{min,lcts}\right) * 4$ (2)

Justification : Comme la production de lactate est reliée à la disponibilité du lactose, une modulation dynamique pour l'export de lactate imitant l'import de lactose (voir eq. (??)), avec un coefficient stoechiométrique de 4.

- EX_ac_e :

- inférieur : 0

- supérieur : $\max\left(-\frac{m_{lcts_e}}{\Delta t * \sum_{i \in \mathcal{M}(lcts_e)} b_i}, v_{i,ace}^{exp}\right)$

Justification : La production d'acétate est régulée par la disponibilité de lactose quand celui-ci est épuisé, et est exporté autrement selon la limite d'export physiologique intrinsèque $v_{i,ace}^{exp} = 0.5$

- EX_diact_e, EX_btd_RR_e, EX_cit_e :

$$\max\left(-\frac{m_j}{\Delta t * \sum_{i \in \mathcal{M}(j)} b_i}, v_{i,j}^{int}\right)$$

- inférieur : pour $i = L. lactis$ et $j \in \{EX_diact_e, EX_btd_RR_e, EX_cit_e, EX_ac_e\}$

- supérieur : 1000

Justification : Limitation dynamique de consommation usuelle (voir eq. (3.6)) Quand le substrat est pas limité, l'import est contraint par la limite physiologique intrinsèque modélisée par $v_{i,j}^{int} = -8$ pour $j = L. lactis$ et $j \in \{EX_diact_e, EX_btd_RR_e\}$ et $v_{i,j}^{int} = -1$ pour $j = EX_cit_e$, et autrement, par la disponibilité du nutriment. La disponibilité du substrat est uniformément partagé par les micro-organismes consommateurs.

En plus, les bornes inférieures de *EX_glc_D_e* (glucose), *EX_coa_e* (coenzyme A) and *EX_starch1200_e* (amidon) sont mis à 0, puisque ces composés n'apparaissent pas dans la composition du lait.

2.5.0.2 *L. plantarum*

- EX_lcts_e :

- inférieur : $\max\left(-\frac{m_{lcts_e}}{\Delta t * \sum_{i \in \mathcal{M}(lcts_e)} b_i}, -\mu_{max,lcts} * 10^{(-k_{lac} * \phi_{undiss})} - \mu_{min,lcts}\right)$ (3)

- supérieur : 1000

Justification : Quand le lactose est épuisé, le premier terme est activé, modélisant le partage uniforme du lactose disponible parmi l'ensemble de bactéries $\mathcal{M}(lac_L)$ métabolisant le lactose. Autrement, une régulation par le pH est ajoutée : la quantité de molécules d'acide lactique non dissocié est calculé à partir de la concentration de lactate au moyen de la fonction ϕ_{undiss} (voir (3.7) pour l'expression de cette fonction) ; puis, avec une croissance exponentielle à partir de la valeur $-\mu_{max,lcts}^{plant} - \mu_{min,lcts}^{plant}$ à la valeur $-\mu_{min,lcts}^{plant}$ quand la concentration d'acide lactique non dissocié augmente, avec un taux exponentiel k_{lac} . Les valeurs de $\mu_{max,lcts}^{plant}$, $\mu_{min,lcts}^{plant}$ et k_{lac} sont inférées des expériences de co-cultures.

— EX_lac_L_e, EX_lac_D_e :

- inférieur : 0

$$\text{supérieur : } \min\left(\frac{m_{lctse}}{\Delta t * \sum_{i \in \mathcal{M}(lctse)} b_i}, \mu_{max,lcts} * 10^{(-k_{lac} * \phi_{undiss})} + \mu_{min,lcts}\right) * 4 \quad (4)$$

Justification : Comme la production de lactate est reliée à la disponibilité du lactose, une modulation dynamique pour l'export de lactate imitant l'import de lactose (voir eq. (??)), avec un coefficient stoechiométrique de 4.

— EX_ac_e :

- inférieur : 0

$$\text{supérieur : } \max\left(-\frac{m_{lctse}}{\Delta t * \sum_{i \in \mathcal{M}(lctse)} b_i}, v_{i,ace}^{exp}\right)$$

Justification : La production d'acetate est régulée par la disponibilité de lactose quand celui ci est épuisé, et est exporté autrement selon la limite d'export physiologique intrinsèque $v_{i,ace}^{exp} = 1.0$

— EX_btd_RR_e :

$$\max\left(-\frac{m_j}{\Delta t * \sum_{i \in \mathcal{M}(j)} b_i}, v_{i,j}^{int}\right),$$

pour $j = L. plantarum$ et $j \in \{EX_btd_RR_e\}$

- supérieur : 1000

Justification :

Limitation dynamique de consommation usuelle (voir eq. (3.6)) Quand le substrat est pas limité, l'import est contraint par la limite physiologique intrinsèque modélisé par $v_{i,j}^{int} = -5$ pour $i = L. plantarum$ et $j \in \{EX_ac_e, EX_btd_RR_e\}$ et autrement, par la disponibilité du nutriment. La disponibilité du substrat est uniformement partagé par les micro-organismes consommateurs.

En plus, les bornes inférieure de EX_glc_D_e (glucose), EX_gal_e (galactose), EX_coa_e (coenzyme A) and EX_dha_e (dihydroxyacetone) sont mis à 0, puisque ces composés n'apparaissent pas dans la composition du lait.

2.5.0.3 *P. freudenreichii*

— EX_ser_L_e :

- inférieur : -12

- supérieur : 1000

Justification : Consommation statique de serine

— EX_cit_e :

- inférieur : -12

- supérieur : 1000

Justification : Consommation statique de citrate

— EX_glu_L_e :

- inférieur : -12

- supérieur : 1000

Justification : Consommation statique de glutamate

— EX_ala_L_e :

- inférieur : -12

- supérieur : 1000

Justification : Consommation statique d'alanine

parameter	<i>L. lactis</i>	<i>L. plantarum</i>	<i>P. freudenreichii</i>
λ	\emptyset	0.351	1.496
k_{lac}	255.591	300	\emptyset
oxygene $\mu_{max,lactose}$	30.666	7.968	\emptyset

TABLE S8 – list of optimized parameters obtained after optimization and used in the fitted models.

— EX_lac__L_e,EX_lac__D_e :

$$\text{- inférieur : } \max\left(-\frac{m_{lac_{L,e}}}{\Delta t * \sum_{i \in \mathcal{M}(lac_L)} b_i}, F_{lac} * K_{freud} * \frac{b_{freud}}{B_{p,freud}}\right) \quad (5)$$

— supérieur : 1000

Justification : Quand le lactate est épuisé, le premier terme est activé, modélisant le partage uniforme parmi l'ensemble des bactéries $\mathcal{M}(lac_L)$, métabolisant L_lactate (resp. D_lactate). Autrement, un flux F_{lac} estimé des données de métabolomiques en culture pure (voir la section ??) est appliqué après une régulation par la fraction $\frac{b_{freud}}{B_{p,freud}}$, où $B_{p,freud}$ est la valeur plateau de la concentration de la biomasse en pure culture, et b_{freud} et la densité de biomasse actuelle, et le facteur K_{freud} , qui est inféré des données de croissance en culture pure. Cette modulation modélise un ordre de grandeur différent du rendement du métabolisme en culture pure ou en co-culture.

— EX_lcts_e :

$$\text{- inférieur : } \max\left(-\frac{m_{lcts_e}}{\Delta t * \sum_{i \in \mathcal{M}(lcts_e)} b_i}, F_{lcts}\right)$$

— supérieur : 1000

Justification : Quand le lactose est épuisé, le premier terme est activé, modélisant le partage uniforme parmi l'ensemble des bactéries $\mathcal{M}(lcts_e)$, métabolisant le lactose. Autrement, un flux constant F_{lcts} estimé des données de métabolomiques en culture pure (voir la section ??) est appliqué.

— EX_ac_e, EX_ppa_e,EX_succ_e :

— inférieure : 0

$$\text{- supérieur : } F_i * K_{freud} * \frac{b_{freud}}{B_{p,freud}}, \quad \text{pour } i \in \{EX_ac_e, EX_ppa_e, EX_succ_e\}$$

Justification : Le taux maximal de production est définie par F_i , pour $i \in \{EX_ac_e, EX_ppa_e, EX_succ_e\}$, qui est estimé des données de métabolomique de culture pure (voir la section ??). Le même term de modulation que 5 est appliqué, modélisant un un ordre de grandeur différent du rendement du métabolisme en culture pure ou en co-culture.

En plus, les bornes inférieure de *EX_starch1200_e* (amidon) and *EX_dha_e* (di-hydroxyacetone) ont été mise à 0, puisque que ces composés n'apparaissent pas dans la composition du lait.

2.6 Bornes dynamiques sur les réactions intracellulaires

Quand

2.7 Estimation des bornes sur la production de métabolite pour *P. freudenreichii*

2.8 Valeurs de paramètres

La valeur des paramètres optimisés sont récapitulé dans la Table S8. Les valeurs de paramètres restant sont récapitulé dans la Table S9.

parameter	description	value [unit]
c_1	pH linear dependence on lactate	-11.1937 [-]
c_2	intercept of pH dependence on lactate	2.8346 [-]
dt_{com}	integration time step in co-culture	0.5 [h]
Tf_{com}	Final experiment time in co-culture	1236 [h]
dt_{mono}	integration time step in co-culture	0.2 [h]
Tf_{mono}	Final experiment time in co-culture for <i>P. freudenreichii</i> , <i>L. lactis</i> and <i>L. plantarum</i>	122,80, 80[h]
$citrate_{concentration}$	Inoculum citrate	2.23 [g/kg]
$acetate_{concentration}$	Inoculum acetate	0.01 [g/kg]
$diacetylc_{concentration}$	Inoculum diacetyl	0.0 [g/kg]
$succinate_{concentration}$	Inoculum succinate	0.085750 [g/kg]
$lactose_{concentration}$	Inoculum lactose	48 [g/kg]
$lactate_{concentration}$	Inoculum lactate	0.0 [g/kg]
$thresh_{qs,cocult}$	Carrying capacity for <i>P. freudenreichii</i> , <i>L. lactis</i> and <i>L. plantarum</i> in co-culture	-3.85, -3.47, -4.04 [\log_{10} g kg $^{-1}$]
$thresh_{qs,monocult}$	Carrying capacity for <i>P. freudenreichii</i> , <i>L. lactis</i> and <i>L. plantarum</i> in mono-culture	-3.15, -4.15, -4.50 [\log_{10} g kg $^{-1}$]
$inoculum_com$	Bacteria inoculation for <i>P. freudenreichii</i> , <i>L. lactis</i> , <i>L. plantarum</i>	6.1,5.7,5.2 [\log_{10} CFU/g]
$oxygene$	Value of the oxygene for <i>P. freudenreichii</i> , <i>L. lactis</i> , <i>L. plantarum</i>	2,7,11 [mmol.gDW $^{-1}$.hr $^{-1}$]
$intrinsuc_{flux}$	Value of the intrinsic flux for <i>P. freudenreichii</i> , <i>L. lactis</i> , <i>L. plantarum</i>	-10,-8,-5 [mmol.gDW $^{-1}$.hr $^{-1}$]

TABLE S9 – Parameter values used for simulations

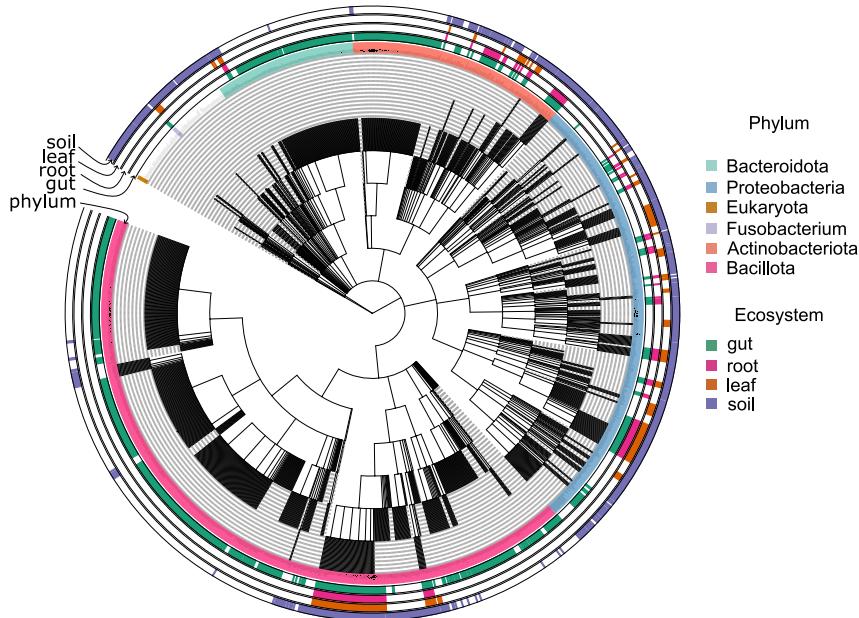


FIGURE S4 – Taxonomy of the genomes associated to the GSMNs of the four ecosystems. The inner colour circle depicts the phylum of the genome. The four additional circles indicate the presence of the taxon among the genomes of the corresponding ecosystem. The figure was created with iTOL (?).

.1 Taxonomie des génomes utilisé pour reconstruire les communautés simulés

2752 génomes de références isolés des microbiotes de l'intestin humain, feuille, racine, et du sol ont été utilisé pour la reconstruction des réseaux métaboliques à l'échelle du génomes (GEMs) et l'assemblage de communautés simulés pour chacun des quatre habitats. La figure supplémentaire S4 montre la diversité taxonomique de ces génomes.

.2 Prédiction des potentiels de coopération et de compétition au sein des communautés simulées dans les divers écosystèmes.

.2.1 Absence d'association entre la topologie du réseau et les interactions potentielles

Le contenu réactionnel de paire de GEMs de l'écosystème de l'intestin a été comparé aux interactions potentielles prédit par CoCoMiCo. La table supplémentaire Table S10 présente les résultats des corrélations porté entre les indices de similarité de Jaccard et les deux métriques d'interaction, illustrant une absence d'association entre eux.

.2.2 Potentiels de coopération et de compétition à travers différentes taille de communautés.

Le manuscrit principale présente la comparaison globale du potentiel de compétition et de coopération à travers les communautés associés à divers écosystèmes. Ces communautés sont composés de 5 à 150 GEMs. La figure supplémentaire S5 montre les résultats du

TABLE S10 – Correlation test results (Spearman) between Jaccard similarity distances and interaction potential scores

ecosystem	n	Competition			Cooperation		
		S	rho estimate	p. value	S	rho estimate	p. value
gut	49	21563.51	-0.035	0.807	20595.99	0.011	0.940
leaf	49	24018.37	-0.153	0.288	20513.94	0.015	0.918
root	49	23431.14	-0.125	0.387	23248.35	-0.116	0.421
soil	49	24797.20	-0.191	0.185	17236.74	0.172	0.231

potentiel d’interaction pour chaque taille de communauté. Les scores de compétition varie peu entre les écosystèmes pour des petites tailles de communautés, plus la communauté est grande, plus on observe des différences de potentiels de compétition entre les divers habitats. La même chose est applicable pour le score de coopération.

.2.3 Les potentials de coopération et de compétition dans un écosystème non réel.

La figure supplémentaire S6 compare les potentiels de coopération et de compétition de l’écosystème "mix",non réel, avec ceux des quatre autres habitats. Les résultats montre que pour toutes les tailles de communautés, les deux potentiels d’interaction de l’écosystème "mix" sont plus haut que les autres.

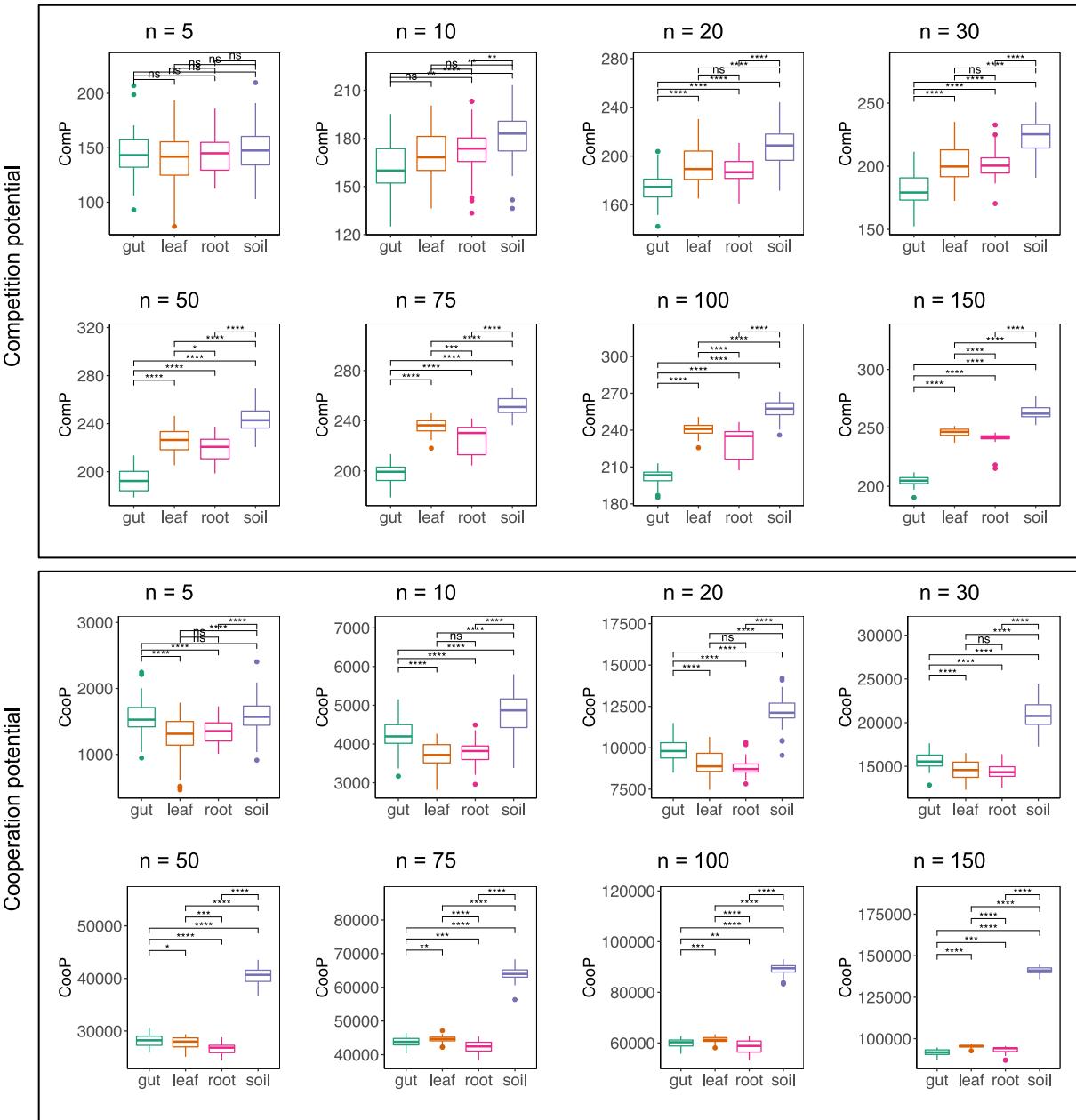


FIGURE S5 – Cooperation and competition potentials in microbial communities of different size across diverse ecosystems. Adjusted P-values (Benjamini Hochberg) of pairwise Wilcoxon tests are depicted on the plots for ecosystem comparison. "n = 10" : number of GSMNs considered in the community.

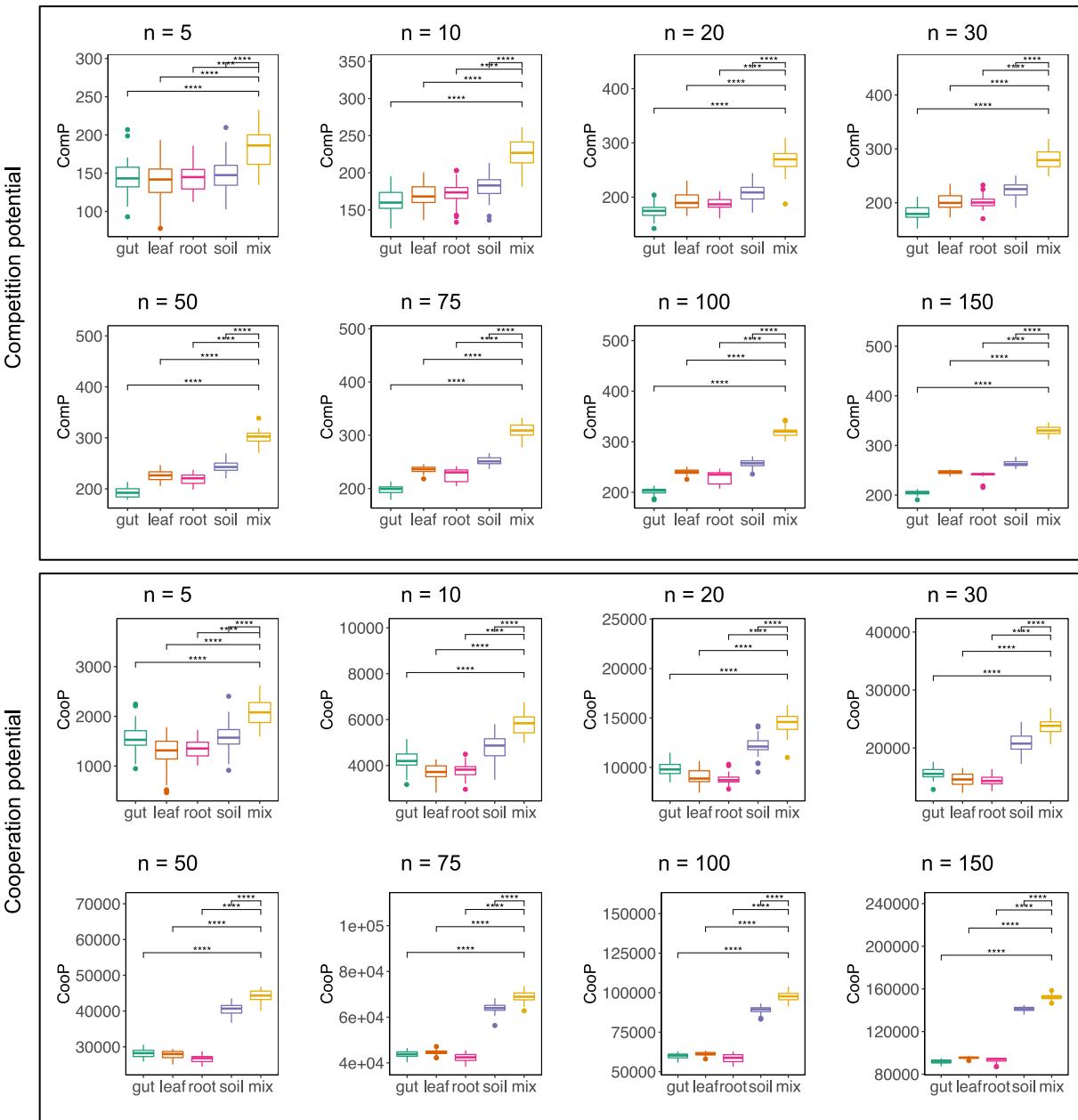


FIGURE S6 – Comparison of cooperation and competition potentials in microbial communities of different size across diverse ecosystems with those of a fifth ecosystem ('mix') consisting in a mixture of GSMNs from the four habitats. Adjusted P-values (Benjamini Hochberg) of pairwise Wilcoxon tests are depicted on the plots for ecosystem comparison. "n = 10" : number of GSMNs considered in the community.

Bibliographie

- [1] Abedi, E. and Hashemi, S. M. B. (2020). Lactic acid production producing microorganisms and substrates sources-state of art. *Helijon*, 6(10):e04974.
- [2] Adams, M. and Mitchell, R. (2002). Fermentation and pathogen control : A risk assessment approach. *International Journal of Food Microbiology*, 79(1-2):75–83.
- [3] Agren, R., Bordel, S., Mardinoglu, A., Pornputtapong, N., Nookaew, I., and Nielsen, J. (2012). Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Computational Biology*, 8(5).
- [4] Aite, M., Chevallier, M., Frioux, C., Trottier, C., Got, J., Cortés, M. P., Mendoza, S. N., Carrier, G., Dameron, O., Guillaudeux, N., Latorre, M., Loira, N., Markov, G. V., Maass, A., and Siegel, A. (2018). Traceability, reproducibility and wiki-exploration for à-la-carte reconstructions of genome-scale metabolic models. *PLoS Computational Biology*, 14(5):1–25.
- [5] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410.
- [6] Ankrah, N. Y. D., Bernstein, D. B., Biggs, M., Carey, M., Engevik, M., García-Jiménez, B., Lakshmanan, M., Pacheco, A. R., Sulheim, S., and Medlock, G. L. (2021). Enhancing Microbiome Research through Genome-Scale Metabolic Modeling. *mSystems*, 6(6):e00599–21.
- [7] Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., Dehal, P., Ware, D., Perez, F., Canon, S., Sneddon, M. W., Henderson, M. L., Riehl, W. J., Murphy-Olson, D., Chan, S. Y., Kamimura, R. T., Kumari, S., Drake, M. M., Brettin, T. S., Glass, E. M., Chivian, D., Gunter, D., Weston, D. J., Allen, B. H., Baumohl, J., Best, A. A., Bowen, B., Brenner, S. E., Bun, C. C., Chandonia, J. M., Chia, J. M., Colasanti, R., Conrad, N., Davis, J. J., Davison, B. H., Dejongh, M., Devoid, S., Dietrich, E., Dubchak, I., Edirisinghe, J. N., Fang, G., Faria, J. P., Frybarger, P. M., Gerlach, W., Gerstein, M., Greiner, A., Gurtowski, J., Haun, H. L., He, F., Jain, R., Joachimiak, M. P., Keegan, K. P., Kondo, S., Kumar, V., Land, M. L., Meyer, F., Mills, M., Novichkov, P. S., Oh, T., Olsen, G. J., Olson, R., Parrello, B., Pasternak, S., Pearson, E., Poon, S. S., Price, G. A., Ramakrishnan, S., Ranjan, P., Ronald, P. C., Schatz, M. C., Seaver, S. M., Shukla, M., Sutormin, R. A., Syed, M. H., Thomason, J., Tintle, N. L., Wang, D., Xia, F., Yoo, H., Yoo, S., and Yu, D. (2018). KBase : The United States department of energy systems biology knowledgebase. *Nature Biotechnology*, 36(7):566–569.

- [8] Assar, R., Montecino, M. A., Maass, A., and Sherman, D. J. (2014). Modeling acclimatization by hybrid systems : Condition changes alter biological system behavior models. *BioSystems*, 121:43–53.
- [9] Ay, F., Dang, M., and Kahveci, T. (2012). Metabolic network alignment in large scale by network compression. *BMC Bioinformatics*, 13(SUPPL.3):1–19.
- [10] Aziz, R. K., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O. (2008). The RAST Server : Rapid annotations using subsystems technology. *BMC Genomics*, 9:1–15.
- [11] Bai, Y., Müller, D. B., Srinivas, G., Garrido-Oter, R., Potthoff, E., Rott, M., Domrowski, N., Münch, P. C., Spaepen, S., Remus-Emsermann, M., Hüttel, B., McHardy, A. C., Vorholt, J. A., and Schulze-Lefert, P. (2015). Functional overlap of the Arabidopsis leaf and root microbiota. *Nature*, 528(7582):364–369.
- [12] Baldini, F., Heinken, A., Heirendt, L., Magnusdottir, S., Fleming, R. M., and Thiele, I. (2019). The microbiome modeling toolbox : From microbial interactions to personalized microbial communities. *Bioinformatics*, 35(13):2332–2334.
- [13] Barabási, A. L. and Oltvai, Z. N. (2004). Network biology : Understanding the cell’s functional organization. *Nature Reviews Genetics*, 5(2):101–113.
- [14] Becker, S. A. and Palsson, B. O. (2008). Context-Specific Metabolic Networks Are Consistent with Experiments. *PLoS Computational Biology*, 4(5):e1000082.
- [15] Belcour, A., Frioux, C., Aite, M., Bretaudau, A., Hildebrand, F., and Siegel, A. (2020). Metage2metabo, microbiota-scale metabolic complementarity for the identification of key species. *eLife*.
- [16] Bernstein, D. B., Dewhirst, F. E., and Segrè, D. (2019). Metabolic network percolation quantifies biosynthetic capabilities across the human oral microbiome. *eLife*, 8:e39733.
- [17] Berry, D. and Widder, S. (2014). Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Frontiers in Microbiology*, 5(MAY):1–14.
- [18] Beveridge, O. S., Petchey, O. L., and Humphries, S. (2010). Direct and indirect effects of temperature on the population dynamics and ecosystem functioning of aquatic microbial ecosystems. *Journal of Animal Ecology*, 79(6):1324–1331.
- [19] Biggs, M. B., Medlock, G. L., Kolling, G. L., and Papin, J. A. (2015). Metabolic network modeling of microbial communities. *Wiley Interdisciplinary Reviews : Systems Biology and Medicine*, 7(5):317–334.
- [20] Blasche, S., Kim, Y., Mars, R. A. T., Machado, D., Maansson, M., Kafkia, E., Milanes, A., Zeller, G., Teusink, B., Nielsen, J., Benes, V., Neves, R., Sauer, U., and Patil, K. R. (2021). Metabolic cooperation and spatiotemporal niche partitioning in a kefir microbial community. *Nature Microbiology*, 6(2):196–208.

- [21] Boele, J. and Olivier, B. G. (2012). MetabolicNetwork_Modeling_Server FAME , the Flux Analysis and Modeling Environment. *BMC Systems Biology*, 6(8).
- [22] Borghei, R., Mobasher, M., and Sobati, T. (2021). Genome Scale Constraint-Based Metabolic Reconstruction of Propionibacterium Freudenreichii DSM 20271. *Preprint from Research square*, pages 1–29.
- [23] Brook, B. S. and Waters, S. L. (2008). Mathematical challenges in integrative physiology. *Journal of Mathematical Biology*, 56(6):893–896.
- [24] Butland, G., Babu, M., Díaz-Mejía, J. J., Bohdana, F., Phanse, S., Gold, B., Yang, W., Li, J., Gagarinova, A. G., Pogoutse, O., Mori, H., Wanner, B. L., Lo, H., Wasniewski, J., Christopoulos, C., Ali, M., Venn, P., Safavi-Naini, A., Sourour, N., Carron, S., Choi, J. Y., Laigle, L., Nazarians-Armavil, A., Deshpande, A., Joe, S., Dat-senko, K. A., Yamamoto, N., Andrews, B. J., Boone, C., Ding, H., Sheikh, B., Moreno-Hagelsieb, G., Greenblatt, J. F., and Emili, A. (2008). eSGA : E. coli synthetic genetic array analysis. *Nature Methods*, 5(9):789–795.
- [25] Cabalar, P. telingo = ASP + Time. pages 1–13.
- [26] Cai, J., Tan, T., and Chan, S. H. (2020). Predicting Nash equilibria for microbial metabolic interactions. *Bioinformatics*, 36(24):5649–5655.
- [27] Cal, M. and Pasteur, L. (1947). BRITISH.
- [28] Cantalapiedra, C. P., Hernandez-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2 : Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, 38(12):5825–5829.
- [29] Cao, W., Aubert, J., Maillard, M. B., Boissel, F., Leduc, A., Thomas, J. L., Deutsch, S. M., Camier, B., Kerjouh, A., Parayre, S., Harel-Oger, M., Garric, G., Thierry, A., and Falentin, H. (2021). Fine-Tuning of Process Parameters Modulates Specific Metabolic Bacterial Activities and Aroma Compound Production in Semi-Hard Cheese. *Journal of Agricultural and Food Chemistry*, 69(30):8511–8529.
- [30] Carr, R. and Borenstein, E. (2012). NetSeed : a network-based reverse-ecology tool for calculating the metabolic interface of an organism with its environment. 28(5):734–735.
- [31] Carroll, N. M., Ross, R. P., Kelly, S. M., Price, N. C., Sheehan, D., and Cogan, T. M. (1999). Characterization of recombinant acetolactate synthase from Leuconostoc lactis NCW1. *Enzyme and Microbial Technology*, 25(1-2):61–67.
- [32] Caspi, R., Billington, R., Fulcher, C. A., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Midford, P. E., Ong, Q., Ong, W. K., Paley, S., Subhraveti, P., and Karp, P. D. (2018). The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Research*, 46(D1):D633–D639.
- [33] Cattani, C. and Ciancio, A. (2012). Separable transition density in the hybrid model for tumor-immune system competition. *Computational and Mathematical Methods in Medicine*, 2012.

- [34] Chan, S. H. J., Simons, M. N., and Maranas, C. D. (2017). SteadyCom : Predicting microbial abundances while ensuring community stability. *PLoS Computational Biology*, 13(5):1–25.
- [35] Chang, A., Jeske, L., Ulbrich, S., Hofmann, J., Koblitz, J., Schomburg, I., Neumann-Schaal, M., Jahn, D., and Schomburg, D. (2021). BRENDa, the ELIXIR core data resource in 2021 : New developments and updates. *Nucleic Acids Research*, 49(D1):D498–D508.
- [36] Chiu, H. C., Levy, R., and Borenstein, E. (2014). Emergent Biosynthetic Capacity in Simple Microbial Communities. *PLoS Computational Biology*, 10(7).
- [37] Choi, J., Yang, F., Stepanauskas, R., Cardenas, E., Garoutte, A., Williams, R., Flatter, J., Tiedje, J. M., Hofmockel, K. S., Gelder, B., and Howe, A. (2017). Strategies to improve reference databases for soil microbiomes. *ISME Journal*, 11(4):829–834.
- [38] Colarusso, A. V., Goodchild-Michelman, I., Rayle, M., and Zomorrodi, A. R. (2021). Computational modeling of metabolism in microbial communities on a genome-scale. *Current Opinion in Systems Biology*, 26:46–57.
- [39] Corblin, F., Bordeaux, L., Hamadi, Y., Fanchon, E., and Trilling, L. (2006). A SAT-based approach to decipher Gene Regulatory Networks. *Gene*, pages 1088–1091.
- [40] Costa, R. S., Hartmann, A., and Vinga, S. (2016). Kinetic modeling of cell metabolism for microbial production. *Journal of Biotechnology*, 219:126–141.
- [41] Crow, V. L. (1986). Metabolism of aspartate by propionibacterium freudenreichii subsp. shermanii : Effect on lactate fermentation. *Applied and Environmental Microbiology*, 52(2):359–365.
- [42] Cruz, D. A. and Kemp, M. L. (2022). Hybrid computational modeling methods for systems biology. *Progress in Biomedical Engineering*, 4(1).
- [43] Dank, A., van Mastrigt, O., Boeren, S., Lillevang, S. K., Abbe, T., and Smid, E. J. (2021). Propionibacterium freudenreichii thrives in microaerobic conditions by complete oxidation of lactate to CO₂. *Environmental Microbiology*, 23(6):3116–3129.
- [44] Deborde, C. and Boyaval, P. (2000). Interactions between pyruvate and lactate metabolism in Propionibacterium freudenreichii subsp. shermanii : In vivo ¹³C nuclear magnetic resonance studies. *Applied and Environmental Microbiology*, 66(5):2012–2020.
- [45] Delcher, A. L., Bratke, K. A., Powers, E. C., and Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics*, 23(6):673–679.
- [46] Demir, E., Cary, M. P., Paley, S., Fukuda, K., Lemke, C., Vastrik, I., Wu, G., D'Eustachio, P., Schaefer, C., Luciano, J., Schacherer, F., Martinez-Flores, I., Hu, Z., Jimenez-Jacinto, V., Joshi-Tope, G., Kandasamy, K., Lopez-Fuentes, A. C., Mi, H., Pichler, E., Rodchenkov, I., Splendiani, A., Tkachev, S., Zucker, J., Gopinath, G., Rajasimha, H., Ramakrishnan, R., Shah, I., Syed, M., Anwar, N., Babur, Ö., Blinov, M., Brauner, E., Corwin, D., Donaldson, S., Gibbons, F., Goldberg, R., Hornbeck, P., Luna, A., Murray-Rust, P., Neumann, E., Reubenacker, O., Samwald, M., Van Iersel, M., Wimalaratne, S., Allen, K., Braun, B., Whirl-Carrillo, M., Cheung, K. H., Dahlquist, K., Finney, A., Gillespie, M., Glass, E., Gong, L., Haw, R., Honig, M., Hubaut, O., Kane, D., Krupa,

- S., Kutmon, M., Leonard, J., Marks, D., Merberg, D., Petri, V., Pico, A., Ravenscroft, D., Ren, L., Shah, N., Sunshine, M., Tang, R., Whaley, R., Letovksy, S., Buetow, K. H., Rzhetsky, A., Schachter, V., Sobral, B. S., Dogrusoz, U., McWeeney, S., Aladjem, M., Birney, E., Collado-Vides, J., Goto, S., Hucka, M., Novère, N. L., Maltsev, N., Pandey, A., Thomas, P., Wingender, E., Karp, P. D., Sander, C., and Bader, G. D. (2010). The BioPAX community standard for pathway data sharing. *Nature Biotechnology*, 28(9):935–942.
- [47] Dias, O., Rocha, M., Ferreira, E. C., and Rocha, I. (2015). Reconstructing genome-scale metabolic models with merlin. *Nucleic acids research*, 43(8):3899–3910.
- [48] Diener, C. and Gibbons, S. M. (2023). More is Different : Metabolic Modeling of Diverse Microbial Communities. *mSystems*, pages e01270–22.
- [49] Diener, C., Gibbons, S. M., and Resendis-Antonio, O. (2020). MICOM : Metagenome-Scale Modeling To Infer Metabolic Interactions in the Gut Microbiota. *mSystems*, 5(1).
- [50] Ebenhöh, O., Handorf, T., and Heinrich, R. (2004). Structural analysis of expanding metabolic networks. *Genome informatics. International Conference on Genome Informatics*, 15(1):35–45.
- [51] Ebrahim, A., Lerman, J. A., Palsson, B. O., and Hyduke, D. R. (2013). COBRAPy : COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*, 7(1):74.
- [52] F., V. P. (1838). Notice sur la loi que la population suit dans son accroissement. *Correspondence mathematique et physique*, 10:113–129.
- [53] Falentin, H., Deutsch, S.-M., Jan, G., Loux, V., Thierry, A., Parayre, S., Maillard, M.-B., Dherbécourt, J., Cousin, F. J., Jardin, J., Siguier, P., Couloux, A., Barbe, V., Vacherie, B., Wincker, P., Gibrat, J.-F., Gaillardin, C., and Lortal, S. (2010). The Complete Genome of Propionibacterium freudenreichii CIRM-BIA1T, a Hardy Actinobacterium with Food and Probiotic Applications. *PLoS ONE*, 5(7):e11748.
- [54] Fang, X., Lloyd, C. J., and Palsson, B. O. (2020). Reconstructing organisms in silico : genome-scale models and their emerging applications. *Nature Reviews Microbiology*, pages 1–13.
- [55] Faust, K. and Raes, J. (2012). Microbial interactions : From networks to models. *Nature Reviews Microbiology*, 10(8):538–550.
- [56] Fazzino, L., Anisman, J., Chacón, J. M., Heineman, R. H., and Harcombe, W. R. (2020). Lytic bacteriophage have diverse indirect effects in a synthetic cross-feeding community. *ISME Journal*, 14(1):123–134.
- [57] Feist, A. M. and Palsson, B. O. (2010). The biomass objective function. *Current Opinion in Microbiology*, 13(3):344–349. Ecology and industrial microbiology * Special section : Systems biology.
- [58] Fischbach, M. A. and Krogan, N. J. (2010). The next frontier of systems biology : Higher-order and interspecies interactions. *Genome Biology*, 11(5).

- [59] Fisher, J. and Piterman, N. (2010). The executable pathway to biological networks. *Briefings in Functional Genomics and Proteomics*, 9(1):79–92.
- [60] Fournier, P., Pellan, L., Barroso-Bergadà, D., Bohan, D. A., Candresse, T., Delmotte, F., Dufour, M.-C., Lauvergeat, V., Marrec, C. L., Marais, A., Martins, G., Masneuf-Pomarède, I., Rey, P., Sherman, D., This, P., Frioux, C., Labarthe, S., and Vacher, C. (2022). The functional microbiome of grapevine throughout plant evolutionary history and lifetime. *Advances in Ecological Research*.
- [61] Freilich, S., Zarecki, R., Eilam, O., Segal, E. S., Henry, C. S., Kupiec, M., Gophna, U., Sharan, R., and Ruppin, E. (2011). Competitive and cooperative metabolic interactions in bacterial communities. *Nature Communications*, 2(1).
- [62] Frioux, C., Fremy, E., Trottier, C., and Siegel, A. (2018). Scalable and exhaustive screening of metabolic functions carried out by microbial consortia. *Bioinformatics*, 34(17):i934–i943.
- [63] Frioux, C., Schaub, T., Schellhorn, S., Siegel, A., and Wanko, P. (2019). Hybrid metabolic network completion. *Theory and Practice of Logic Programming*, 19(1):83–108.
- [64] Galimberti, A., Bruno, A., Agostinetto, G., Casiraghi, M., Guzzetti, L., and Labra, M. (2021). Fermented food products in the era of globalization : tradition meets biotechnology innovations. *Current Opinion in Biotechnology*, 70:36–41.
- [65] García-Jiménez, B., García, J. L., and Nogales, J. (2018). FLYCOP : Metabolic modeling-based analysis and engineering microbial communities. *Bioinformatics*, 34(17):i954–i963.
- [66] Gebser, M., Kaminski, R., Kaufmann, B., and Schaub, T. (2019). Multi-shot ASP solving with clingo. *Theory and Practice of Logic Programming*, 19(1):27–82.
- [67] Goelzer, A., Muntel, J., Chubukov, V., Jules, M., Prestel, E., Nölker, R., Mariadasou, M., Aymerich, S., Hecker, M., Noirot, P., Becher, D., and Fromion, V. (2015). Quantitative prediction of genome-wide resource allocation in bacteria. *Metabolic Engineering*, 32:232–243.
- [68] Gomez, J. A. and Barton, P. I. (2018). Dynamic flux balance analysis using DFBAlab. *Methods in Molecular Biology*, 1716:353–370.
- [69] Gomez, J. A., Höffner, K., and Barton, P. I. (2014). Dfbalab : a fast and reliable matlab code for dynamic flux balance analysis. *BMC bioinformatics*, 15(1):1–10.
- [70] Handorf, T., Ebenhöh, O., and Heinrich, R. (2005). Expanding metabolic networks : Scopes of compounds, robustness, and evolution. *Journal of Molecular Evolution*, 61(4):498–512.
- [71] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362.

- [72] Heiden, J. V., Wernisch, L., Gilbert, D., and Wodak, S. J. (2002). Graph-Based Analysis of Metabolic Networks.
- [73] Henry, C. S., Bernstein, H. C., Weisenhorn, P., Taylor, R. C., Lee, J. Y., Zuckerman, J., and Song, H. S. (2016). Microbial Community Metabolic Modeling : A Community Data-Driven Network Reconstruction. *Journal of Cellular Physiology*, 231(11):2339–2345.
- [74] Hoorman, J. J. (2011). The Role of Soil Bacteria. *The Ohio State University Extension*, pages 1–4.
- [75] Hucka, M., Bergmann, F. T., Chaouiya, C., Dräger, A., Hoops, S., Keating, S. M., König, M., Novère, N. L., Myers, C. J., Olivier, B. G., Sahle, S., Schaff, J. C., Sheriff, R., Smith, L. P., Waltemath, D., Wilkinson, D. J., and Zhang, F. (2019). *The Systems Biology Markup Language (SBML) : Language Specification for Level 3 Version 2 Core Release 2*, volume 16.
- [76] Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernandez-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., Von Mering, C., and Bork, P. (2019). EggNOG 5.0 : A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47(D1):D309–D314.
- [77] Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal : prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11(1):119.
- [78] Janhunen, T., Kaminski, R., Ostrowski, M., Schellhorn, S., Wanko, P., and Schaub, T. (2017). Clingo goes linear constraints over reals and integers. *Theory and Practice of Logic Programming*, 17(5-6):872–888.
- [79] Jenior, M. L., Moutinho, T. J., Dougherty, B. V., and Papin, J. A. (2020). Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments. *PLoS Computational Biology*, 16(4).
- [80] Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., and Barabasi, A. L. (2011). The large-scale organization of metabolic networks. *The Structure and Dynamics of Networks*, 9781400841356(1990):211–214.
- [81] Kanehisa, M. and Goto, S. (2000). KEGG : Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1):27–30.
- [82] Karlsen, E., Schulz, C., and Almaas, E. (2018). Automated generation of genome-scale metabolic draft reconstructions based on KEGG. *BMC Bioinformatics*, 19(1):1–11.
- [83] Karp, P. D., Billington, R., Caspi, R., Fulcher, C. A., Latendresse, M., Kothari, A., Keseler, I. M., Krummenacker, M., Midford, P. E., Ong, Q., Ong, W. K., Paley, S. M., and Subhraveti, P. (2018). The BioCyc collection of microbial genomes and metabolic pathways. *Briefings in Bioinformatics*, 20(4):1085–1093.

- [84] Karp, P. D., Paley, S., Krummenacker, M., Kothari, A., Wannemuehler, M. J., and Phillips, G. J. (2022). Pathway Tools Management of Pathway/Genome Data for Microbial Communities. *Frontiers in Bioinformatics*, 2(April):1–11.
- [85] Khandelwal, R. A., Olivier, B. G., Röling, W. F., Teusink, B., and Bruggeman, F. J. (2013). Community Flux Balance Analysis for Microbial Consortia at Balanced Growth. *PLoS ONE*, 8(5).
- [86] King, Z. A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N. E., and Palsson, B. O. (2015). Escher : A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. *PLoS Computational Biology*.
- [87] King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. O., and Lewis, N. E. (2016). BiGG Models : A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*.
- [88] Kleerebezem, M., Boekhorst, J., Van Kranenburg, R., Molenaar, D., Kuipers, O. P., Leer, R., Tarchini, R., Peters, S. A., Sandbrink, H. M., Fiers, M. W., Stiekema, W., Klein Lankhorst, R. M., Bron, P. A., Hoffer, S. M., Nierop Groot, M. N., Kerkhoven, R., De Vries, M., Ursing, B., De Vos, W. M., and Siezen, R. J. (2003). Complete genome sequence of *Lactobacillus plantarum* WCFS1. *Proceedings of the National Academy of Sciences of the United States of America*, 100(4):1990–1995.
- [89] Klitgord, N. and Segrè, D. (2010). Environments that induce synthetic microbial ecosystems. *PLOS Computational Biology*, 6(11):1–17.
- [90] Koch, S., Benndorf, D., Fronk, K., Reichl, U., and Klamt, S. (2016). Predicting compositions of microbial communities from stoichiometric models with applications for the biogas process. *Biotechnology for Biofuels*, 9(1):1–16.
- [91] Kreimer, A., Doron-Faigenboim, A., Borenstein, E., and Freilich, S. (2012). NetCmpt : A network-based tool for calculating the metabolic competition between bacterial species. *Bioinformatics*.
- [92] Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. (2009). Langmead b, trapnell c, pop m, salzberg sl.. ultrafast and memory-efficient alignment of short dna sequences to the human genome. *genome biol* 10 : R25. *Genome biology*, 10:R25.
- [93] LeChevallier, M. W. (2013). Water Treatment and Pathogen Control : Process Efficiency in Achieving Safe Drinking-water. *Water Intelligence Online*, 12.
- [94] Lee, C. K., Laughlin, D. C., Bottos, E. M., Caruso, T., Joy, K., Barrett, J. E., Brabyn, L., Nielsen, U. N., Adams, B. J., Wall, D. H., Hopkins, D. W., Pointing, S. B., McDonald, I. R., Cowan, D. A., Banks, J. C., Stichbury, G. A., Jones, I., Zawar-Reza, P., Katurji, M., Hogg, I. D., Sparrow, A. D., Storey, B. C., Allan Green, T. G., and Cary, S. C. (2019). Biotic interactions are an unexpected yet critical control on the complexity of an abiotically driven polar ecosystem. *Communications Biology*, 2(1).
- [95] Lemke, N., Herédia, F., Barcellos, C. K., dos Reis, A. N., and Mombach, J. C. (2004). Essentiality and damage in metabolic networks. *Bioinformatics*, 20(1):115–119.
- [96] Levesque, H. J. (1986). Knowledge representation and reasoning. *Annual Review of Computer Science*, 1(1):255–287.

- [97] Levy, R. and Borenstein, E. Reverse Ecology : From Systems to Environments and Back. pages 329–345.
- [98] Levy, R. and Borenstein, E. (2013). Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proceedings of the National Academy of Sciences*, 110(31):12804 LP – 12809.
- [99] Levy, R., Carr, R., Kreimer, A., Freilich, S., and Borenstein, E. (2015). NetCooperate : A network-based tool for inferring host-microbe and microbe-microbe cooperation. *BMC Bioinformatics*.
- [100] Lewis, N. E., Hixson, K. K., Conrad, T. M., Lerman, J. A., Charusanti, P., Polpitiya, A. D., Adkins, J. N., Schramm, G., Purvine, S. O., Lopez-Ferrer, D., Weitz, K. K., Eils, R., König, R., Smith, R. D., and Palsson, B. (2010). Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Molecular Systems Biology*, 6(390).
- [101] Liao, Y. C., Tsai, M. H., Chen, F. C., and Hsiung, C. A. (2012). GEMSiRV : A software platform for GEome-scale metabolic model simulation, reconstruction and visualization. *Bioinformatics*, 28(13):1752–1758.
- [102] Lieven, C., Beber, M. E., Olivier, B. G., Bergmann, F. T., Ataman, M., Babaei, P., Bartell, J. A., Blank, L. M., Chauhan, S., Correia, K., Diener, C., Dräger, A., Ebert, B. E., Edirisinghe, J. N., Faria, J. P., Feist, A. M., Fengos, G., Fleming, R. M. T., García-Jiménez, B., Hatzimanikatis, V., Helvoirt, W. v., Henry, C. S., Hermjakob, H., Herrgård, M. J., Kaafarani, A., Kim, H. U., King, Z., Klamt, S., Klipp, E., Koehorst, J. J., König, M., Lakshmanan, M., Lee, D.-Y., Lee, S. Y., Lee, S., Lewis, N. E., Liu, F., Ma, H., Machado, D., Mahadevan, R., Maia, P., Mardinoglu, A., Medlock, G. L., Monk, J. M., Nielsen, J., Nielsen, L. K., Nogales, J., Nookaew, I., Palsson, B. O., Papin, J. A., Patil, K. R., Poolman, M., Price, N. D., Resendis-Antonio, O., Richelle, A., Rocha, I., Sánchez, B. J., Schaap, P. J., Sheriff, R. S. M., Shoaie, S., Sonnenschein, N., Teusink, B., Vilaça, P., Vik, J. O., Wodke, J. A. H., Xavier, J. C., Yuan, Q., Zakhartsev, M., and Zhang, C. (2020). MEMOTE for standardized genome-scale metabolic model testing. *Nature Biotechnology*, 38(3):272–276.
- [103] Lifschitz, V. (2008). What is answer set programming ? *Proceedings of the National Conference on Artificial Intelligence*, 3:1594–1597.
- [104] Lloyd, C. J., King, Z. A., Sandberg, T. E., Hefner, Y., Olson, C. A., Phaneuf, P. V., O'Brien, E. J., Sanders, J. G., Salido, R. A., Sanders, K., Brennan, C., Humphrey, G., Knight, R., and Feist, A. M. (2019). The genetic basis for adaptation of modeldesigned syntrophic co-cultures. *PLoS Computational Biology*, 15(3):1–28.
- [105] Loux, V., Mariadassou, M., Almeida, S., Chiapello, H., Hammani, A., Buratti, J., Gendrault, A., Barbe, V., Aury, J. M., Deutsch, S. M., Parayre, S., Madec, M. N., Chuat, V., Jan, G., Peterlongo, P., Azevedo, V., Le Loir, Y., and Falentin, H. (2015). Mutations and genomic islands can explain the strain dependency of sugar utilization in 21 strains of *Propionibacterium freudenreichii*. *BMC Genomics*, 16(1):1–19.
- [106] Ma, H. and Zeng, A.-p. (2003). Genome Data and Analysis of Their Global. *Bioinformatics*, 19(2):270–277.

- [107] Machado, D., Andrejev, S., Tramontano, M., and Patil, K. R. (2018). Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Research*.
- [108] Machado, D., Maistrenko, O. M., Andrejev, S., Kim, Y., Bork, P., Patil, K. R., and Patil, K. R. (2021). *Polarization of microbial communities between competitive and cooperative metabolism*, volume 5.
- [109] Mahadevan, R., Edwards, J. S., and Doyle, F. J. (2002a). Dynamic Flux Balance Analysis of Diauxic Growth in Escherichia coli. *Biophysical Journal*, 83(3):1331–1340.
- [110] Mahadevan, R., Edwards, J. S., and Doyle, F. J. (2002b). Dynamic Flux Balance Analysis of Diauxic Growth in Escherichia coli. *Biophysical Journal*, 83(3):1331–1340.
- [111] Mahadevan, R. and Schilling, C. H. (2003). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering*.
- [112] Mahout, M., Carlson, R. P., and Peres, S. (2020). Answer set programming for computing constraints-based elementary flux modes : Application to escherichia coli core metabolism. *Processes*, 8(12):1–17.
- [113] Makhlof, A. (2006). *Méthodologie pour l'optimisation dynamique multicritère d'un procédé discontinu alimenté : application à la production bactérienne d'arômes laitiers*. PhD thesis, Institut National Polytechnique de Lorraine.
- [114] Manna, M., Han, G., Seo, Y.-S., and Park, I. (2021). Evolution of Food Fermentation Processes and the Use of Multi-Omics in Deciphering the Roles of the Microbiota. *Foods*, 10(11):2861.
- [115] McKinney, W. et al. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX.
- [116] M??digue, C., Bocs, S., Labarre, L., Math ??, C., and Vallenet, D. (2002). Bio-informatique (1) l'annotation in silico des s ??quences g ??nomiques. *Medecine/Sciences*, 18(2):237–250.
- [117] Mendes-Soares, H., Mundy, M., Soares, L. M., and Chia, N. (2016). MMinte : An application for predicting metabolic interactions among the microbial species in a community. *BMC Bioinformatics*, 17(1):1–10.
- [118] Mendoza, S. N., Olivier, B. G., Molenaar, D., and Teusink, B. (2019). A systematic assessment of current genome-scale metabolic reconstruction tools. *Genome Biology*, 20(1):1–20.
- [119] Miyoshi, A., Rochat, T., and Gratadoux, J.-j. (2003). Miyoshi. *Genetics and Molecular Research*, 2(4):348–359.
- [120] Morin, M. A., Morrison, A. J., Harms, M. J., and Dutton, R. J. (2022). Higher-order interactions shape microbial interactions as microbial community complexity increases. *Scientific Reports*, 12(1):1–16.
- [121] Nachtweide, S. and Stanke, M. (2019). Chapter 8. 1962:139–160.

- [122] Nair, R. R., Vasse, M., Wielgoss, S., Sun, L., Yu, Y. T. N., and Velicer, G. J. (2019). Bacterial predator-prey coevolution accelerates genome evolution and selects on virulence-associated prey defences. *Nature Communications*, 10(1):1–10.
- [123] Noronha, A., Modamio, J., Jarosz, Y., Guerard, E., Sompairac, N., Preciat, G., Daniélsdóttir, A. D., Krecke, M., Merten, D., Haraldsdóttir, H. S., Heinken, A., Heirendt, L., Magnúsdóttir, S., Ravcheev, D. A., Sahoo, S., Gawron, P., Friscioni, L., Garcia, B., Prendergast, M., Puente, A., Rodrigues, M., Roy, A., Rouquaya, M., Wiltgen, L., agare, A., John, E., Krueger, M., Kuperstein, I., Zinovyev, A., Schneider, R., Fleming, R. M. T., and Thiele, I. (2018). The Virtual Metabolic Human database : integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Research*, 47(D1):D614–D624.
- [124] Oliveira, R. A. and Pamer, E. G. (2023). Assembling symbiotic bacterial species into live therapeutic consortia that reconstitute microbiome functions. *Cell Host & Microbe*, 31(4):472–484.
- [125] Olivier, B. (2018). SystemsBioinformatics/cbmpy-metadraft : MetaDraft is now available.
- [126] Orth, J. D., Thiele, I., and Palsson, B. O. (2010). What is flux balance analysis ?
- [127] Pacheco, A. R., Moel, M., and Segrè, D. (2019). Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems. *Nature Communications*, 10(1):103.
- [128] Pacheco, A. R. and Segrè, D. (2021). An evolutionary algorithm for designing microbial communities via environmental modification. *Journal of the Royal Society Interface*, 18(179).
- [129] Palles, T., Beresford, T., Condon, S., and Cogan, T. M. (1998). Citrate metabolism in *Lactobacillus casei* and *Lactobacillus plantarum*. *Journal of Applied Microbiology*, 85(1):147–154.
- [130] Palsson, B. O. and Varma, A. (1994). Metabolic Flux Balancing : Basic Concepts, Scientific and Practical Use. *Biotechnology*, 12(October):994–998.
- [131] Passi, A., Tibocha-Bonilla, J. D., Kumar, M., Tec-Campos, D., Zengler, K., and Zuniga, C. (2022). Genome-scale metabolic modeling enables in-depth understanding of big data. *Metabolites*, 12(1).
- [132] Pitkänen, E., Jouhten, P., Hou, J., Syed, M. F., Blomberg, P., Kludas, J., Oja, M., Holm, L., Penttilä, M., Rousu, J., and Arvas, M. (2014). Comparative Genome-Scale Reconstruction of Gapless Metabolic Networks for Present and Ancestral Species. *PLoS Computational Biology*, 10(2).
- [133] Popp, D. and Centler, F. (2020). μ BialSim : Constraint-Based Dynamic Simulation of Complex Microbiomes. *Frontiers in Bioengineering and Biotechnology*, 8(June).
- [134] Putri, G. H., Anders, S., Pyl, P. T., Pimanda, J. E., and Zanini, F. (2022). Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics*, 38(10):2943–2945.

- [135] Quatravaux, S., Remize, F., Bryckaert, E., Colavizza, D., and Guzzo, J. (2006). Examination of Lactobacillus plantarum lactate metabolism side effects in relation to the modulation of aeration parameters. *Journal of Applied Microbiology*, 101(4):903–912.
- [136] R, K., R, Z., and T, L. (2000). Pathway analysis in metabolic databases via differential metabolic display (DMD). *Bioinformatics*, 16(9):825–836.
- [137] R Core Team (2021). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [138] Rau, M. H. and Zeidan, A. A. (2018). Constraint-based modeling in microbial food biotechnology. *Biochemical Society Transactions*, 46(2):249–260.
- [139] Robinson, J. L., Kocab, P., Wang, H., Cholley, P. E., Cook, D., Nilsson, A., Anton, M., Ferreira, R., Domenzain, I., Billa, V., Limeta, A., Hedin, A., Gustafsson, J., Kerckhoven, E. J., Svensson, L. T., Palsson, B. O., Mardinoglu, A., Hansson, L., Uhlén, M., and Nielsen, J. (2020). An atlas of human metabolism. *Science Signaling*, 13(624):1–12.
- [140] Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2009). edgeR : a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140.
- [141] Roissart, H. d. (DL 1994). *Bactéries lactiques : aspects fondamentaux et technologiques / H. de Roissart, F. M. Luquet, coordonnateurs*. Lorica, Uriage.
- [142] Römer, M., Eichner, J., Dräger, A., Wrzodek, C., Wrzodek, F., and Zell, A. (2016). ZBIT bioinformatics toolbox : A web-platform for systems biology and expression data analysis. *PLoS ONE*.
- [143] Rousk, K., Michelsen, A., and Rousk, J. (2016). Microbial control of soil organic matter mineralization responses to labile carbon in subarctic climate change treatments. *Global Change Biology*, 22(12):4150–4161.
- [144] Schroeder, W. L. and Saha, R. (2020). Introducing an Optimization- and explicit Runge-Kutta- based Approach to Perform Dynamic Flux Balance Analysis. *Scientific Reports*, 10(1):1–28.
- [145] SCHUSTER, S. and HILGETAG, C. (1994). On elementary flux modes in biochemical reaction systems at steady state. *Journal of Biological Systems*, 02(02):165–182.
- [146] Schwab, J. D., Kühlwein, S. D., Ikonomi, N., Kühl, M., and Kestler, H. A. (2020). Concepts in Boolean network modeling : What do they all mean ? *Computational and Structural Biotechnology Journal*, 18:571–582.
- [147] Seaver, S. M., Liu, F., Zhang, Q., Jeffryes, J., Faria, J. P., Edirisinghe, J. N., Mundy, M., Chia, N., Noor, E., Beber, M. E., Best, A. A., DeJongh, M., Kimbrel, J. A., D'haeseleer, P., McCorkle, S. R., Bolton, J. R., Pearson, E., Canon, S., Wood-Charlson, E. M., Cottingham, R. W., Arkin, A. P., and Henry, C. S. (2021). The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Research*.

- [148] Seemann, T. (2014). Prokka : rapid prokaryotic genome annotation. *Bioinformatics*, 30(14):2068–2069.
- [149] Smid, E. J. and Kleerebezem, M. (2014). Production of aroma compounds in lactic fermentations. *Annual Review of Food Science and Technology*, 5(1):313–326.
- [150] Somerville, V., Grigaitis, P., Battjes, J., Moro, F., and Teusink, B. (2021). Use and limitations of genome-scale metabolic models in food microbiology. *Current Opinion in Food Science*.
- [151] Stein, R. R., Bucci, V., Toussaint, N. C., Buffie, C. G., Rätsch, G., Pamer, E. G., Sander, C., and Xavier, J. B. (2013). Ecological Modeling from Time-Series Inference : Insight into Dynamics and Stability of Intestinal Microbiota. *PLoS Computational Biology*, 9(12):31–36.
- [152] Stéphanou, A. and Volpert, V. (2016). Hybrid Modelling in Biology : A Classification Review. *Mathematical Modelling of Natural Phenomena*, 11(1):37–38.
- [153] Steuer, R. (2007). Computational approaches to the topology, stability and dynamics of metabolic networks. *Phytochemistry*, 68(16-18):2139–2151.
- [154] Stobbe, M. D., Jansen, G. A., Moerland, P. D., and van Kampen, A. H. (2014). Knowledge representation in metabolic pathway databases. *Briefings in Bioinformatics*, 15(3):455–470.
- [155] Succurro, A. and Ebenhöh, O. (2018). Review and perspective on mathematical modeling of microbial ecosystems. *Biochemical Society Transactions*, 46(2):403–412.
- [156] Sweetlove, L. J. and Fernie, A. R. (2005). Regulation of metabolic networks : Understanding metabolic complexity in the systems biology era. *New Phytologist*, 168(1):9–24.
- [157] Swindell, S. R., Benson, K. H., Griffin, H. G., Renault, P., Ehrlich, S. D., and Gasson, M. J. (1996). Genetic manipulation of the pathway for diacetyl metabolism in *Lactococcus lactis*. *Applied and Environmental Microbiology*, 62(7):2641–2643.
- [158] Tamang, J. P., Shin, D.-H., Jung, S.-J., and Chae, S.-W. (2016a). Functional Properties of Microorganisms in Fermented Foods. *Frontiers in Microbiology*, 7:578.
- [159] Tamang, J. P., Watanabe, K., and Holzapfel, W. H. (2016b). Review : Diversity of microorganisms in global fermented foods and beverages. *Frontiers in Microbiology*, 7(MAR).
- [160] Thiele, I. and Palsson, B. O. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, 5(1):93–121.
- [161] Thierry, A., Deutsch, S. M., Falentin, H., Dalmasso, M., Cousin, F. J., and Jan, G. (2011). New insights into physiology and metabolism of *Propionibacterium freudenreichii*. *International Journal of Food Microbiology*, 149(1):19–27.
- [162] Thorleifsson, S. G. and Thiele, I. (2011). rBioNet : A COBRA toolbox extension for reconstructing high-quality biochemical networks. *Bioinformatics*, 27(14):2009–2010.

- [163] Tong, A. H., Evangelista, M., Parsons, A. B., Xu, H., Bader, G. D., Pagé, N., Robinson, M., Raghibizadeh, S., Hogue, C. W., Bussey, H., Andrews, B., Tyers, M., and Boone, C. (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science*, 294(5550):2364–2368.
- [164] Trinh, C. T., Wlaschin, A., and Srienc, F. (2009). Elementary mode analysis : A useful metabolic pathway analysis tool for characterizing cellular metabolism. *Applied Microbiology and Biotechnology*, 81(5):813–826.
- [165] Turgay, M., Bachmann, H.-P., Irmler, S., von Ah, U., Frölich-Wyder, M.-T., Fallentin, H., Deutsch, S.-M., Jan, G., and Thierry, A. (2020). Propionibacterium spp. and Acidipropionibacterium spp. In *Reference Module in Food Science*, Reference Module in Food Science. Elsevier.
- [166] Vallenet, D., Calteau, A., Dubois, M., Amours, P., Bazin, A., Beuvin, M., Burlot, L., Bussell, X., Fouteau, S., Gautreau, G., Lajus, A., Langlois, J., Planel, R., Roche, D., Rollin, J., Rouy, Z., Sabatet, V., and Médigue, C. (2020). MicroScope : An integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Research*, 48(D1):D579–D589.
- [167] Vallenet, D., Calteau, A., Dubois, M., Amours, P., Bazin, A., Beuvin, M., Burlot, L., Bussell, X., Fouteau, S., Gautreau, G., Lajus, A., Langlois, J., Planel, R., Roche, D., Rollin, J., Rouy, Z., Sabatet, V., and Médigue, C. (2019). MicroScope : an integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Research*.
- [168] Van Rooijen, R. J., Van Schalkwijk, S., and De Vos, W. M. (1991). Molecular cloning, characterization, and nucleotide sequence of the tagatose 6-phosphate pathway gene cluster of the lactose operon of *Lactococcus lactis*. *Journal of Biological Chemistry*, 266(11):7176–7181.
- [169] van Rosmalen, R., Smith, R., Martins dos Santos, V., Fleck, C., and Suarez-Diez, M. (2021). Model reduction of genome-scale metabolic models as a basis for targeted kinetic models. *Metabolic Engineering*, 64:74–84.
- [170] Van Rossum, G. and Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA.
- [171] Virdis, C., Sumby, K., Bartowsky, E., and Jiranek, V. (2021). Lactic Acid Bacteria in Wine : Technological Advances and Evaluation of Their Functional Role. *Frontiers in Microbiology*, 11(January):1–16.
- [172] Wagner, A. and Fell, D. A. (2001). The small world inside large metabolic networks. *Proceedings of the Royal Society B : Biological Sciences*, 268(1478):1803–1810.
- [173] Wang, H., Marcišauskas, S., Sánchez, B. J., Domenzain, I., Hermansson, D., Agren, R., Nielsen, J., and Kerkhoven, E. J. (2018). RAVEN 2.0 : A versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. *PLoS Computational Biology*, 14(10):1–17.

- [174] Wardle, D. A. and Peltzer, D. A. (2007). *Aboveground-Belowground Linkages, Ecosystem Development, and Ecosystem Restoration*, pages 45–68. Springer New York, New York, NY.
- [175] Weiss, A. S., Burrichter, A. G., Durai Raj, A. C., von Strempe, A., Meng, C., Kleigrewe, K., Münch, P. C., Rössler, L., Huber, C., Eisenreich, W., Jochum, L. M., Göing, S., Jung, K., Lincetto, C., Hübner, J., Marinos, G., Zimmermann, J., Kaleta, C., Sanchez, A., and Stecher, B. (2022). In vitro interaction network of a synthetic gut bacterial community. *ISME Journal*, 16(4):1095–1109.
- [176] Wes McKinney (2010). Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61.
- [177] Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017a). Completing bacterial genome assemblies with multiplex MinION sequencing. *Microbial Genomics*, 3(10):1–7.
- [178] Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017b). Unicycler : Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Computational Biology*, 13(6):1–22.
- [179] Widystuti, Y., , R., and Febrisiantosa, A. (2014). The Role of Lactic Acid Bacteria in Milk Fermentation. *Food and Nutrition Sciences*, 05(04):435–442.
- [180] Wilmes, P., Martin-Gallausiaux, C., Ostaszewski, M., Aho, V. T., Novikova, P. V., Laczy, C. C., and Schneider, J. G. (2022). The gut microbiome molecular complex in human health and disease. *Cell Host & Microbe*, 30(9):1201–1206.
- [181] Yadav, S. K., Soni, R., and Rajput, A. S. (2018). Role of Microbes in Organic Farming for Sustainable Agro-Ecosystem. pages 241–252.
- [182] Zelezniak, A., Andrejev, S., Ponomarova, O., Mende, D. R., Bork, P., and Patil, K. R. (2015). Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proceedings of the National Academy of Sciences of the United States of America*.
- [183] Zhang, Y. J., Li, S., Gan, R. Y., Zhou, T., Xu, D. P., and Li, H. B. (2015). Impacts of gut bacteria on human health and diseases. *International Journal of Molecular Sciences*, 16(4):7493–7519.
- [184] Zhuang, K., Izallalen, M., Mouser, P., Richter, H., Risso, C., Mahadevan, R., and Lovley, D. R. (2011). Genome-scale dynamic modeling of the competition between Rhodoferax and Geobacter in anoxic subsurface environments. *ISME Journal*, 5(2):305–316.
- [185] Zien, A., Küffner, R., Zimmer, R., and Lengauer, T. (2000). Analysis of gene expression data with pathway scores. *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology*, 8:407–417.
- [186] Zimmermann, J., Kaleta, C., and Waschina, S. (2021). gapseq : informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. *Genome Biology*, 22(1):81.

- [187] Zomorrodi, A. R. d-OptCom : Dynamic Multi-level and Multi-objective Metabolic Modeling of Microbial Communities.
- [188] Zomorrodi, A. R. and Maranas, C. D. (2012). OptCom : A multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLoS Computational Biology*, 8(2).
- [189] Zomorrodi, A. R. and Segrè, D. (2017). Genome-driven evolutionary game theory helps understand the rise of metabolic interdependencies in microbial communities. *Nature Communications*, 8(1):1–11.
- [190] Zou, Y., Xue, W., Luo, G., Deng, Z., Qin, P., Guo, R., Sun, H., Xia, Y., Liang, S., Dai, Y., Wan, D., Jiang, R., Su, L., Feng, Q., Jie, Z., Guo, T., Xia, Z., Liu, C., Yu, J., Lin, Y., Tang, S., Huo, G., Xu, X., Hou, Y., Liu, X., Wang, J., Yang, H., Kristiansen, K., Li, J., Jia, H., and Xiao, L. (2019). 1,520 Reference Genomes From Cultivated Human Gut Bacteria Enable Functional Microbiome Analyses. *Nature Biotechnology*, 37(2):179–185.
- [191] Zuñiga, C., Li, T., Guarnieri, M. T., Jenkins, J. P., Li, C. T., Bingol, K., Kim, Y. M., Betenbaugh, M. J., and Zengler, K. (2020). Synthetic microbial communities of heterotrophs and phototrophs facilitate sustainable growth. *Nature Communications*, 11(1).
- [192] Özcan, E., Seven, M., irin, B., Çakr, T., Nikerel, E., Teusink, B., and Öner, E. T. (2020). Dynamic coculture metabolic models reveal the fermentation dynamics, metabolic capacities and interplays of cheese starter cultures. *Biotechnology and Bioengineering*, 118(1):223–237.