



Reconstructions de réseaux métaboliques à partir de données de métagénomique shotgun Nanopore.

Moana Aulagner

► To cite this version:

Moana Aulagner. Reconstructions de réseaux métaboliques à partir de données de métagénomique shotgun Nanopore.. Bio-informatique [q-bio.QM]. 2022. hal-03879996

HAL Id: hal-03879996

<https://inria.hal.science/hal-03879996v1>

Submitted on 30 Nov 2022

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

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



Stage de Master 1 mention Bio-informatique

Université de Rennes 1

Moana AULAGNER

2021 - 2022

Reconstructions de réseaux métaboliques à
partir de données de métagénomique shotgun
Nanopore.

Centre de recherche Inria Rennes, équipe Dyliss

Encadrants : Anne SIEGEL, Jeanne GOT & Emeline ROUX

ENGAGEMENT DE NON PLAGIAT

Je, soussigné (e) **Moana Aulagner**.....
Etudiant (e) en **Master 1 Bio-Informatique**.....

Déclare être pleinement informé (e) que le plagiat de documents ou d'une partie de documents publiés sous toute forme de support (y compris l'internet), constitue une violation des droits d'auteur ainsi qu'une fraude caractérisée.

En conséquence, je m'engage à citer toutes les sources que j'ai utilisées pour la rédaction de ce document.

Signature



Remerciements

Ce stage a été réalisé à l'interface entre l'équipe Dyliss et l'équipe Genscale de l'INRIA de Rennes, en collaboration avec l'équipe EAT de l'Institut NuMeCan à l'INRAE de Saint-Gilles. Je tiens à exprimer ma reconnaissance à Anne SIEGEL (directrice de recherche au CNRS, Dyliss) et Emeline ROUX (Maître de conférence de l'université de Rennes 1, collaboratrice extérieure Genscale) pour m'avoir accueillie au sein de leurs équipes respectives. Ce stage m'a beaucoup apporté et m'a permis d'obtenir une première expérience dans le domaine des réseaux métaboliques qui représentent le centre de mon intérêt pour la bio-informatique. Je remercie Anne SIEGEL pour m'avoir aidé à garder une vision d'ensemble sur cette importante quantité de données. Je remercie également Emeline ROUX et Gaëlle BOUDRY (Directrice de recherche, INRAE) pour m'avoir partagé les données du projet HOLOPIG ainsi que de leur soutien sur les aspects et interprétations biologiques du projet. Je tiens à remercier Jeanne GOT (ingénieure en bio-informatique, Dyliss) pour son aide régulière et ses guides d'installation qui ont facilité mon utilisation des nombreux outils nécessaires durant ce stage. Merci à Samuel BLANQUART (chercheur, Dyliss) pour ses conseils qui m'ont aidé à mieux orienter et structurer mes recherches. Je tiens également à mentionner Arnaud BELCOUR (doctorant, Dyliss) et Jacky AME (stagiaire, Genscale) qui ont grandement facilité ma prise en main du sujet en prenant le temps de répondre mes multiples questions.

Sommaire

1	Introduction	1
1.1	Santé et microbiote intestinal	1
1.2	<i>Escherichia coli</i> , une bactérie d'intérêt	1
1.3	Rôle de la métagénomique	2
1.4	Métabolisme et réseaux métaboliques	2
1.5	Méthodes de reconstruction	3
1.6	Objectifs du travail	4
2	Matériel et méthodes	4
2.1	Obtention des données métagénomiques	4
2.2	Présentation des outils principaux	5
2.2.1	Pathway Tools, un outil de reconstruction	5
2.2.2	Padmet et Metage2Metabo, deux outils d'exploration	6
2.3	Reconstruction des réseaux métaboliques	6
2.3.1	Récupération des identifiants MetaCyc	6
2.3.2	Etapes de reconstruction	6
2.4	Premières analyses des réseaux métaboliques	7
2.4.1	Production globale	8
2.4.2	Production du succinate et du tryptophane	8
2.4.3	Dégradation du succinate et du tryptophane	8
3	Résultats	9
3.1	Souches identifiées	9
3.2	Récupération des identifiants MetaCyc	9
3.3	Reconstruction des réseaux	9
3.4	Production globale	10
3.5	Succinate	10
3.5.1	Production du succinate	10
3.5.2	Dégradation du succinate	11
3.6	Tryptophane	11
3.6.1	Production du tryptophane	11
3.6.2	Dégradation du tryptophane	12

4	Discussion	12
4.1	Souches identifiées	12
4.2	Récupération des identifiants MetaCyc	13
4.3	Evaluation globale	13
4.4	Production et dégradation du succinate et du tryptophane	13
4.4.1	Succinate	13
4.4.2	Tryptophane	14
4.5	Limites de la reconstruction automatisée	14
4.6	Perspectives	14
5	Conclusion	15
	Bibliographie	i
A	Présentation de la structure d'accueil	iv

1 Introduction

1.1 Santé et microbiote intestinal

Un microbiote est un ensemble de bactéries vivant en symbiose dans un environnement donné et formant une communauté où chaque organisme interagit avec les autres. Le développement du microbiote intestinal s'accélère juste après la naissance et dépend de nombreux facteurs environnementaux, nutritionnels, comportementaux et génétiques [9]. Les interactions entre le microbiote et son hôte se font via des métabolites qui peuvent impacter la santé de l'hôte de façon positive ou négative. De manière générale les micro-organismes dégradent des composés complexes en composés plus simples tels que les SCFA¹. Ces acides jouent un rôle important dans la protection de la barrière muqueuse en servant de source d'énergie aux cellules épithéliales [9]. Ils ont aussi un rôle de protection contre les pathogènes en participant notamment au développement du système immunitaire [16]. La communauté bactérienne est également capable de synthétiser des composés nécessaires à l'organisme comme des vitamines ou du cholestérol, voire même des composés neurochimiques comme le GABA² dont la mauvaise régulation peut être à l'origine de troubles psychologiques [9]. Ce travail se focalise sur deux composés d'intérêt pour leurs valeurs nutritionnelles et leurs rôles clés dans la santé humaine, le succinate et le tryptophane. Le succinate est un métabolite très présent dans les processus énergétiques et donc indispensable au bon fonctionnement des cellules [20]. Le tryptophane est à la base de la production de métabolites impactant directement les échanges entre le microbiote et l'hôte. Par exemple la sérotonine - un dérivé du tryptophane - est un neurotransmetteur [1]. Le microbiote intestinal a ainsi un impact direct sur la santé humaine et il a d'ailleurs été montré qu'une modification contrôlée de la composition du microbiote d'un hôte malade a entraîné des améliorations de la maladie [9].

1.2 *Escherichia coli*, une bactérie d'intérêt

E. coli est l'entérobactérie la plus abondante du microbiote intestinal et une des 10 bactéries les plus abondantes. C'est un organisme modèle, c'est-à-dire un organisme très étudié car facile à cultiver et ayant un génome de taille réduite. Il sert de représentant et de référence pour l'ensemble de l'étude des bactéries. *E. coli* est pionnière dans la colonisation du microbiote intestinal dû à son anaérobiose facultative. Elle crée ainsi petit à petit un environnement anaérobie idéal pour le reste des bactéries constituant le microbiote en début de vie [19, 5]. *E. coli* colonise la couche de mucus de l'épithélium et elle a un impact sur le développement des cellules épithéliales de

1. Short Chain Fatty Acids

2. Gamma Amino Butyrique Acid

l'intestin [13]. Elle est pour cela un point central de l'étude du microbiote.

1.3 Rôle de la métagénomique

Les bactéries se développent naturellement en communautés et dépendent souvent des autres membres pour survivre. L'étude d'un micro-organisme isolé est donc très complexe et souvent impossible, expliquant le petit nombre d'espèces bactériennes modèles malgré leur incroyable diversité. L'étude de la communauté microbienne dans son environnement naturel a notamment été rendue possible grâce au développement d'outils de métagénomique performants. La métagénomique permet l'identification des micro-organismes présents et d'approximer leurs proportions relatives via le séquençage du métagénome, c'est-à-dire le séquençage de l'ADN total produit par ce microbiote dans un environnement spécifique [6]. Afin d'analyser les échantillons d'une communauté, des techniques de séquençage et d'assemblage de génomes individuels ont été mises en place [7]. En s'affranchissant du besoin de cultiver chaque micro-organisme indépendamment pour pouvoir le séquencer, la métagénomique permet ainsi d'en étudier de nouveaux et de générer les premières données qui vont permettre l'étude des interactions complexes ayant lieu au sein de la communauté.

1.4 Métabolisme et réseaux métaboliques

Afin de caractériser les différentes fonctions de chaque organisme au sein de la communauté ou leurs interactions avec l'hôte il est possible d'utiliser les réseaux métaboliques. Le métabolisme fait référence aux transformations biochimiques qui se font sous l'action d'enzymes catalysant une réaction [12]. Un réseau biologique est une représentation abstraite et simplifiée d'un système biologique, rendant compte des éléments le composant et de leurs interactions. Un réseau métabolique rend compte de toutes ces interactions sous la forme d'un graphe représentant les différents liens entre métabolites et réactions (Figure 1). Il peut représenter différentes échelles du métabolisme, que ce soit au niveau cellulaire, d'un organisme, voire d'une communauté toute entière. Cette approche facilite grandement notre compréhension de la dynamique des processus biologiques. En effet chaque réseau peut être représenté visuellement sous la forme d'un graphe et plusieurs outils permettent d'en extraire les informations désirées [12]. Ces réseaux sont construits grâce à la combinaison des données génomiques avec celles de bases de données fonctionnelles [7]. En effet, les enzymes catalysant les réactions sont des protéines qui sont produites via des gènes. Il est donc possible, via l'utilisation de bases de données omiques, de reconstruire les réseaux métaboliques à partir de données génomiques. Pour une étude précise il est important de coupler ces résultats avec des données de méta-transcriptomiques, la présence d'un gène ne garantissant pas son expression. Le temps nécessaire pour la reconstruction com-

plète d'un réseau métabolique est très long, de quelques mois à plusieurs années, et le réseau évolue constamment au fil des ans, s'affinant et s'enrichissant via l'apport de nouvelles données expérimentales [17].

1.5 Méthodes de reconstruction

Un GSM³ permet de représenter toutes les réactions qu'un organisme peut produire à partir de son information génétique sous forme de réseau. Un protocole en 96 étapes a été mis en place pour la construction de GSM robustes et reproductibles [17]. Les premières étapes visent la construction d'un réseau brut, d'un "brouillon". La qualité de cette première reconstruction dépend grandement de celle des annotations du génome. En effet, si la qualité des annotations ou celle de la base de données sont erronées ou incomplètes, cela risque de générer un nombre important de faux positifs et de faux négatifs [17]. Ensuite plusieurs étapes importantes viennent améliorer et ajuster le modèle de manière itérative. Il existe deux types d'approches principales pour la reconstruction de réseaux métaboliques, ou deux catégories de modèles [6]. La première est une approche topologique basée sur l'utilisation de "sources", c'est à dire d'un milieu de culture qui va servir d'entrée à notre réseau et l'activer. Elle est qualitative.

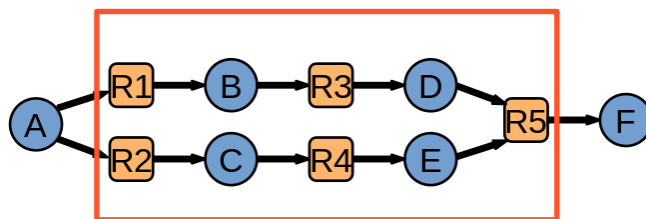


Figure 1. Schéma d'un réseau métabolique simple, contenu dans l'encadré orange, reliant les réactions R_i aux métabolites en bleu. Par exemple la réaction R_3 catalyse la transformation du composé B en composé D. La graine A va venir activer le réseau qui va alors produire en sortie le composé F. Si F est un composé que l'on cherche à obtenir en sortie de réseau, on l'appelle "cible".

La seconde est une approche de flux, qui permet l'analyse dynamique d'un réseau. C'est une approche quantitative bien plus précise, mais beaucoup plus complexe à mettre en place. En effet elle nécessite d'avoir accès à de nombreuses ressources empiriques et celles-ci sont encore très incomplètes. C'est encore plus le cas pour des approches métagénomiques. En effet si le niveau de précision nécessaire n'est pas toujours disponible pour les organismes modèles, il l'est encore moins pour les "non-modèles" très présents en métagénomique.

3. Genome Scale Metabolic model

1.6 Objectifs du travail

Au total, 34 échantillons de colon ont été récupérés par l'INRAE chez des porcelet de 7 jours traités ou non à la colistine, un antibiotique ciblant les bactéries Grams-. Le but de cette étude initiale était d'étudier le rôle des Grams- et d'*E. coli* au sein du microbiote. Pour cela les métabolites présents dans les deux groupes d'échantillons contrôles et traités ont été identifiés et quantifiés. Ils constituent nos données expérimentales. Les génomes de chacune des souches présentes dans ce microbiote ont été identifiés par des techniques de métagénomique et un seul échantillon, l'échantillon numéro 9, a servi à la reconstruction des réseaux métaboliques, le traitement métagénomique n'étant pas encore finalisé pour les autres échantillons. Ce travail se focalise sur une reconstruction pour une analyse topologique. La première étape est d'établir une version brute du réseau, en combinant de manière automatisée les données génomiques obtenues expérimentalement et les informations des bases de données GO⁴, protéiques et métabolomiques. La seconde étape de ce travail est un début d'analyse des prédictions des modèles reconstruits. Le processus repose sur la comparaison des métabolites prédits par le modèle avec les métabolites identifiés empiriquement. Il se focalisera dans un premier temps sur un sous-ensemble des métabolites obtenus expérimentalement, puis une analyse plus poussée sera réalisée en se focalisant sur deux composés d'intérêt, le succinate et le tryptophane.

2 Matériel et méthodes

2.1 Obtention des données métagénomiques

Un séquençage amplicon avait préalablement été réalisé à l'INRAE, où seule une partie des gènes codant pour l'ARN ribosomal 16S avait été séquencé. Cette partie de l'ADN est très conservée entre espèces et permet de les différencier. Cependant cette méthode permet d'identifier les bactéries au mieux à l'échelle du genre et de l'espèce et seul un core génome de l'espèce peut alors être utilisé pour reconstruire des réseaux, ce qui peut manquer de précision dans certains cas. Un nouveau séquençage a été réalisé via une technique de métagénomique shotgun. En ce qui concerne l'échantillon 9 qui a servi à la reconstruction, il s'agit d'un échantillon de colon de porcelet nouveau-né de 7 jours non traité duquel a été extrait l'ADN. Le séquençage a été réalisé en utilisant la technologie Nanopore – avec un kit de préparation de la librairie SQK-LSK109 - et un séquenceur MinION Mk1B. La technologie Nanopore se base sur la détection du signal électrique propre à chaque nucléotide et ne nécessite pas d'amplification PCR. Elle permet ainsi le séquençage de long reads ce qui facilite grandement l'identification des souches

4. Gene Ontology

aux étapes suivantes, tout en s'affranchissant du biais quantitatif dû à la PCR. Les reads obtenus ont été classés par genre ou par espèce avec Kraken 2 v.2.0.9, un système de classification taxonomique basé sur les k-mers. Ils ont été filtrés en ne conservant que ceux correspondant à des espèces présentes au-dessus du seuil arbitraire de 0,5% d'abondance des reads. Beaucoup de reads ont été perdus à cette étape, une majorité de cette perte provenant de reads de mauvaise qualité ou appartenant à l'hôte. Pour chaque groupe de reads classés en un fichier unique fastq par espèce, l'outil ORI v0.0.2 leur attribue ensuite un génome de référence. Il se base sur un échantillon de 4000 reads longs sélectionnés au hasard dans chaque groupe et crée des sous-groupes de reads [15]. Ce sont ces génomes formant une moyenne représentant au mieux nos reads qui vont être conservés pour la suite des analyses. En effet en métagénomique il est difficile de récupérer des génomes complets et cette approximation facilitera le travail de reconstruction de réseaux métaboliques par la suite. De plus, dans un souci de simplification, seuls les génomes déterminés par ORI avec une présence supérieure au seuil arbitraire de 30% dans chaque groupe ont été conservés pour la suite. Parmi cette sélection, un seul génome a été conservé pour les souches très proches génétiquement afin d'éviter les redondances inutiles. Cette estimation statistique a été obtenue grâce à Orthofinder v2.5.2 qui compare les protéines présentes chez chaque souche et estime un degré de similarité. Les annotations ont ensuite été harmonisées pour tous les génomes conservés grâce à EggnoGmapper v2.0.1 afin de permettre plus tard la comparaison des réseaux métaboliques de chaque organisme. Pour plus de détails sur la méthode d'obtention des génomes annotés à partir des échantillons de colon de porc-let il est possible de se référer aux rapports de Rania OUAZHAROU et de Jacky AME qui se sont chargés de ce travail dans le cadre de leurs stages de M2 du master bio-informatique de l'Université de Rennes 1.

2.2 Présentation des outils principaux

2.2.1 Pathway Tools, un outil de reconstruction

Un des outils centraux à cette première reconstruction automatique est Pathway Tools [11]. Pathway Tools est un outil permettant de reconstruire un réseau à partir de génomes annotés et de créer des bases de données de voies métaboliques à partir de ces reconstructions appelées PGDB⁵. L'algorithme de Pathway Tools prédit les voies métaboliques à partir des enzymes présentes dans le génome, en faisant le lien avec les informations contenues dans la base de données MetaCyc [10]. Pour cela, il utilise les annotations du génome uniquement et ne fait pas d'analyses de séquences ou de prédictions de fonctions.

5. Pathway to Genome Databases

2.2.2 Padmet et Metage2Metabo, deux outils d'exploration

Ces deux outils ont été développés à l'INRIA de Rennes. Ils ont pour but de faciliter l'exploration des réseaux à l'aide de diverses options permettant d'en extraire des informations essentielles.

Padmet est un format de fichier qui permet de stocker un réseau métabolique. C'est aussi une librairie qui permet de manipuler ces informations et de faire notamment des comparaisons de réseaux relatives à la présence/absence de réactions. [2].

Metage2metabo est un ensemble d'outils permettant de faire une analyse topologique des réseaux en manipulant des données au format SBML⁶. Il permet d'étudier la complémentarité métabolique au sein d'une communauté microbienne et d'identifier les micro-organismes essentiels pour l'activation d'un processus métabolique cible [3].

2.3 Reconstruction des réseaux métaboliques

2.3.1 Récupération des identifiants MetaCyc

Les métabolites produits par la communauté microbienne intestinale des 34 échantillons ont été caractérisés au sein de deux groupes par l'INRAE, en séparant les échantillons contrôle et traités. Ils ont été identifiés et quantifiés par spectrométrie de masse et RMN⁷. Ces analyses ont permis de déterminer la présence de 857 composés, ainsi que leur abondance relative entre échantillons contrôles et traités. Une récupération semi-automatique des identifiants MetaCyc a été initiée pour chaque métabolite des données expérimentales. Cette récupération s'est basée sur les identifiants CAS, HMDB, KEGG, PUBCHEM et les noms chimiques disponibles dans le fichier fourni par l'INRAE. Tous les identifiants des molécules n'étaient pas présents et plusieurs croisements ont été réalisés avec les données extraites des fichiers suivants :

- All_metabolites_HMDB.xml du 17/11/2021 (<https://hmdb.ca/downloads>)
- MetaCyc.25.0.padmet. (<https://biocyc.org/download.shtml>, téléchargement des fichiers plats, puis constitution du fichier Padmet grâce à l'outil Padmet)
- chem_xref.tsv v4.0 (<https://www.metanetx.org/mnxdoc/mnxref.html>)

Une recherche manuelle des identifiants MetaCyc a été réalisée afin de compléter les identifiants manquants d'un sous-groupe de métabolites qui sera utilisé dans la suite des analyses.

2.3.2 Etapes de reconstruction

La fusion des différentes informations des génomes annotés au format Genbank a été faite grâce à l'outil Emapper2gbk v0.1.0. Avec l'outil Pathway Tools v24.5 et le contenu de MetaCyc

6. System Biology Markup Language

7. Résonance Magnétique Nucléaire

(MetaCyc.25.0.padmet) les informations métaboliques des 54 souches présentes dans l'échantillon ont été récupérées. Cette étape a pu être automatisée grâce à l'implémentation de Mpwt v0.7.2 qui permet de lancer Pathway Tools en ligne de commande. Puis les informations extraites par Pathway Tools ont été combinées grâce à l'outil Padmet v5.0.1.

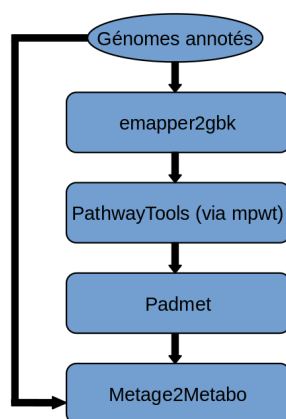


Figure 2. Pipeline récapitulant les outils utilisés pour la génération des réseaux à partir des génomes annotés. La reconstruction a été récemment automatisée et il est possible de réaliser les étapes intermédiaires automatiquement via M2M⁸ avec l'option "recon".

2.4 Premières analyses des réseaux métaboliques

Le but est de comparer différents réseaux métaboliques afin de tenter d'expliquer les variations expérimentales quantitatives entre le groupe contrôle et le groupe traité à la colistine grâce aux prédictions qualitatives des différents réseaux. Afin de réaliser la comparaison, les souches ont été réparties en 3 groupes :

- un groupe contenant toutes les souches : simulation de l'échantillon contrôle.
- un groupe contenant uniquement les Grams+ : simulation de l'échantillon traité.
- un groupe contenant uniquement les Grams- : identification des métabolites produits uniquement par les Grams-.

Plus précisément, cette comparaison a cherché à mettre en relation les variations des quantités de métabolites au sein des échantillons contrôles et traités des données expérimentales, avec la capacité des différents groupes à produire ou dégrader ces métabolites. Ce travail s'est focalisé d'abord sur un premier sous-groupe de métabolites en sélectionnant ceux dont la quantité varie au moins d'un facteur 2 entre les échantillons. Les lipides ont été écartés car leur métabolisme est très complexe et difficile à reconstruire de manière automatisée. Les molécules partiellement caractérisées ont été exclues de cette liste ainsi que les variations non significatives ($p\text{-value} > 0.05$). Puis une comparaison plus poussée de la production et de la dégradation de deux composés d'intérêts, le succinate et le tryptophane a été réalisée.

Afin de prédire les métabolites produits par les réseaux, il est nécessaire de définir un fichier SBML contenant les sources qui vont "démarrer" notre réseau en activant les premières réactions. Ce fichier contient une liste simplifiée des métabolites résultants de la dégradation du lait de truie ainsi que des co-facteurs. Ces co-facteurs sont essentiels pour l'activation systématique des cycles métaboliques. Les informations ont été déduites de l'article [14].

2.4.1 Production globale

Les fichiers Padmet ont été reformatés au format SBML grâce à l'option `sbmlGenerator` de l'outil Padmet. L'outil Metage2Metabo et l'option "cscope" ont été utilisés pour prédire la production des réseaux métaboliques. Cette opération a été réalisée pour chacun des 3 groupes. Les sources définies précédemment ont été utilisées et aucune cible n'a été précisée.

2.4.2 Production du succinate et du tryptophane

Les fichiers SBML des cibles ont été créés via l'utilisation de la plateforme AuReMe v2.4, grâce à l'option "compound_to_sbml" qui convertit une liste d'identifiants MetaCyc de métabolites au format TXT en SBML. L'outil Metage2Metabo a ensuite servi à déterminer les souches capables de produire ou non ces deux composés d'intérêts, via l'option "mincom". Cette étape a été réalisée deux fois, une fois avec le succinate comme cible et une seconde fois avec le tryptophane afin d'obtenir la liste des souches capables de produire ces composés. De plus, une analyse avec l'outil Miscoto v3.1.2 [8] avec l'option "scopes" a été utilisée afin d'identifier les souches capables de produire ces deux composés individuellement puis en incluant la coopération. L'analyse a été répétée 6 fois, en analysant la capacité des 3 groupes à produire les 2 composés cibles.

2.4.3 Dégradation du succinate et du tryptophane

Les outils actuels ne permettent pas d'étudier la dégradation des composés, l'obtention de ces résultats est donc plus complexe. La stratégie utilisée ici a été d'identifier les voies et réactions responsables de la dégradation de nos composés et d'identifier les organismes contenant ces voies ou réactions. Un premier travail d'identification a été réalisé via la base de données MetaCyc en utilisant les SmartTables afin d'extraire les listes des voies capables de dégrader nos composés, des réactions et taxonomies associées [4]. Les voies non répertoriées chez les procaryotes ont été écartées. L'outil Metage2metabo avec l'option "mincom" a été utilisé, avec comme cible la liste des composés de dégradation pour chaque voie identifiée. Une seconde validation de la capacité de dégradation des réseaux a été réalisée en utilisant l'outil Padmet, avec le module `padmet_compare`. Il a permis de comparer les 54 réseaux en rendant compte de la présence/absence des voies métaboliques et des réactions pour chaque souche. Les informa-

tions de ces fichiers de comparaison ont été croisées avec les listes de voies et de réactions de dégradation du succinate et du tryptophane. Seules les souches contenant au moins une réaction dégradant le composé d'intérêt ont été considérées comme capables de le dégrader.

3 Résultats

3.1 Souches identifiées

Un total de 22 genres et espèces différents correspondant à 54 génomes ont été identifiés avec une abondance $>0,5\%$ des reads, 17 sont des bactéries Grams- et 37 sont Grams+. Chez les Grams-, 6 souches *E. coli* ont été identifiées. Chez les Grams+, une majorité de lactobacilles est présente.

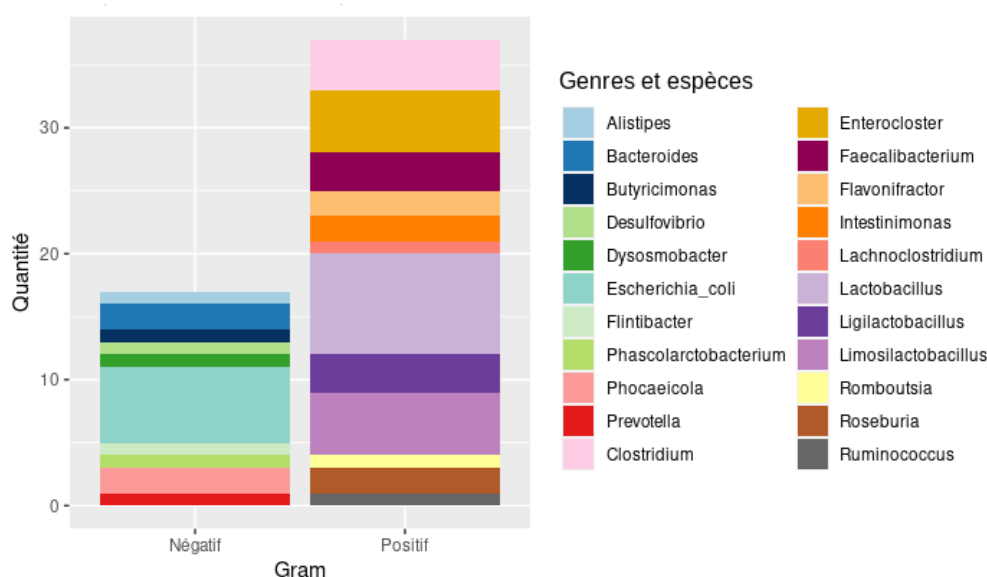


Figure 3. Représentation de la répartition des 54 principales souches identifiées par métagénomique, réparties en 22 genres et espèces. Le calcul de quantité pour chaque souche est relatif, il se base sur le nombre de paires de bases qui lui a été attribué lors de l'analyse métagénomique et la taille de son génome.

3.2 Récupération des identifiants MetaCyc

Les croisements d'informations entre les différents fichiers des bases de données ont permis de récupérer 498 identifiants MetaCyc sur les 857 composés issus des analyses RMN et MS des échantillons de colon.

3.3 Reconstruction des réseaux

Les 54 réseaux des 54 souches sélectionnées ont été créés puis ils ont été regroupés en 3 groupes (ou populations) définis précédemment.

	Métabolites	Réactions	Gènes	Voies
ALL	3586	3426	53317	753
GRAM +	3244	3065	20312	688
GRAM-	2852	2612	33005	531

Figure 4. Nombres de métabolites, réactions, gènes et voies métaboliques présentes au sein de réseaux des 3 populations. Les voies métaboliques incomplètes sont incluses si les réactions clés de ces voies sont présentes. "ALL" fait référence à la population contenant toutes les souches.

3.4 Production globale

Au total, 55 composés ayant un identifiant MetaCyc et une quantité variant d'un facteur d'au moins 2 entre les échantillons contrôles et traités ont été isolés. Ils constituent le sous-groupe de métabolites d'intérêt qui va servir à faire une première validation globale des prédictions des réseaux. Parmi les 55 métabolites identifiés expérimentalement, 22 sont prédits par le réseau des 54 souches et 4 sont prédits uniquement par le réseau de la population Gram-. Ces 4 métabolites sont représentés dans la figure ci-dessous.

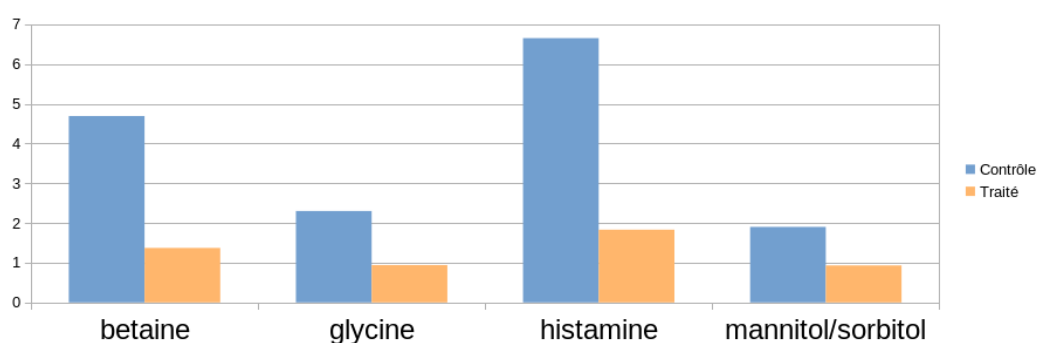


Figure 5. Quantités expérimentales relatives obtenues par spectrométrie de masse de 4 métabolites produits uniquement par la population Gram-. Le mannitol et le sorbitol n'ont pas pu être différenciés par l'analyse de SM.

3.5 Succinate

3.5.1 Production du succinate

En se basant sur les données expérimentales on observe une accumulation du succinate dans les échantillons traités.

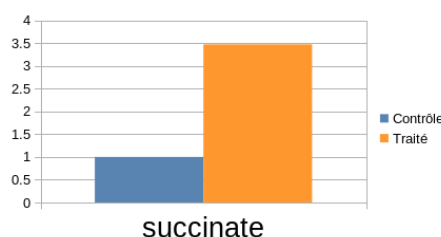


Figure 6. Proportions relatives du succinate dans les échantillons contrôles et traités.

Au total, 16 souches dont une majorité de Grams- sont capables de produire le succinate individuellement. L'ensemble des 54 souches est impliqué dans sa production via des échanges de métabolites, c'est à dire via coopération. Les résultats sont récapitulés dans la figure 7.

3.5.2 Dégradation du succinate

Une voie et une réaction de dégradation du succinate ont pu être identifiées chez les souches :

- Voie de dégradation en propanoate (PWY0-43)
- Réaction de dégradation du succinate en succinylCoA (SUCCCOASYN-RXN)

On observe que toutes les souches *E. coli* dégradent le succinate via ces 2 options. De plus, ces étapes de dégradation du succinate sont présentes chez une majorité de Grams-. Les résultats sont récapitulés dans la figure 7.

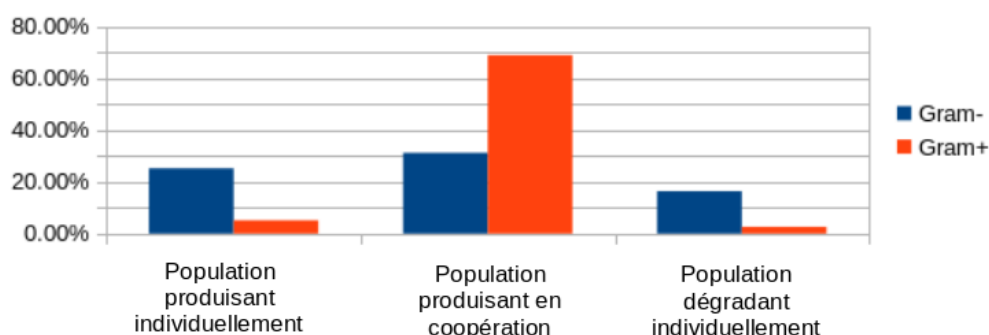


Figure 7. Proportions relatives en nombre de bases des populations Grams+ et Grams- capables de dégrader ou de produire le succinate.

3.6 Tryptophane

3.6.1 Production du tryptophane

En se basant sur les données expérimentales on observe une diminution de la quantité de tryptophane dans les échantillons traités.

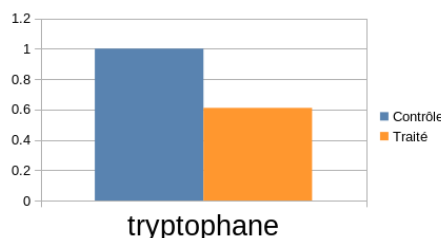


Figure 8. Proportions relatives du tryptophane dans les échantillons contrôles et traités.

Le tryptophane est produit individuellement par 21 souches dont une majorité de Grams-. 14 souches additionnelles participent à la production du tryptophane via une coopération. Les résultats sont récapitulés dans la figure 10.

3.6.2 Dégradation du tryptophane

Une seule voie de dégradation du tryptophane a pu être identifiée chez les souches :

— Voie de dégradation en pyruvate (TRYPDEG-PWY)

On observe qu'une majorité de Grams- dégradent le tryptophane en pyruvate. Les résultats sont récapitulés dans la figure 10. Pour 3 des voies étudiées, seule la seconde réaction de la voie était présente chez une majorité de Grams+.

	Conversion en kynurenine (9 voies)	Conversion via tryptamine (PWY-3181)	Conversion en indole-3-acétate (PWY-3161)
E.Coli	0	6	6
Gram – (sauf E.coli)	3	1	1
Gram +	11	0	10
Total	14	7	17

Figure 9. Tableau représentant 3 voies de dégradation du tryptophane partiellement présentes chez les souches étudiées. Concernant la conversion en kynurenine, il s'agit d'une "sous-voie" métabolique à l'origine de 9 autres voies plus complexes.

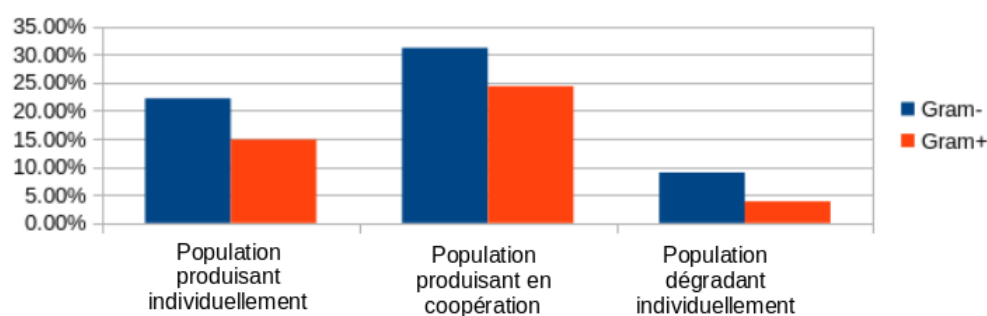


Figure 10. Proportions relatives en nombre de bases des populations Grams+ et Grams- capables de dégrader ou de produire le tryptophane.

4 Discussion

4.1 Souches identifiées

Chez les Grams-, 6 souches *E. coli* ont été identifiées et représentent plus d'un tiers de la population Gram-, ce qui concorde avec le fait qu'*E. coli* est un membre important du microbiote intestinal. Chez les Grams+, une majorité de lactobacilles sont présentes. Ces bactéries sont connues pour dégrader le lait et leur présence majoritaire concorde avec le régime alimentaire des porcelets. Cependant les bactéries responsables de la dégradation du lait chez l'homme sont majoritairement les bifidobactérium [18]. Le microbiote intestinal en début de vie du porcelet n'est peut être pas le modèle le plus adapté pour représenter celui d'un nouveau né humain.

4.2 Récupération des identifiants MetaCyc

Une des difficultés majeure et qui a été chronophage est la recherche des identifiants MetaCyc pour chacun des 857 métabolites. Les identifiants d'autres bases de données étaient disponibles pour certains métabolites, mais il n'est pas aisé de faire les liens nécessaires avec MetaCyc. De plus, en cas d'absence de tout identifiant de départ, la recherche par nom est rendue extrêmement ardue dû au grand nombre de synonymes existant. De plus, 29 des composés présentant la plus grande variation entre échantillons traités et contrôles n'ont pas pu être récupérés sur MetaCyc et ont dû être écartés, réduisant nos possibilités de validation.

4.3 Evaluation globale

Au total, 22 métabolites ont été prédits par les réseaux parmi les 55 obtenus expérimentalement, suggérant un manque d'informations dans nos réseaux. Une première explication de ce résultat est que les données expérimentales ont été obtenues à partir de 34 échantillons de colon et que les réseaux utilisés ont été obtenus à partir d'un seul de ces échantillons. Ces échantillons contiennent des souches additionnelles et différentes en partie de celles de l'échantillon utilisé dans ce travail. Il faudrait reconstruire les réseaux à partir de ces échantillons pour plus de précision une fois leurs traitements métagénomique terminés. Une seconde explication serait que nos réseaux sont incomplets, dû à un manque d'informations soit au niveau des annotations du génomes, soit au niveau des réactions et voies métaboliques de MetaCyc. Parmi ces métabolites 4 sont produits uniquement par la population Gram-. On observe une diminution de leur production dans les échantillons traités, c'est à dire en l'absence de Gram-. Ces résultats concordent, la diminution expérimentale pouvant être reliée à l'absence des Grams- qui étaient les seuls à pouvoir produire ces composés.

4.4 Production et dégradation du succinate et du tryptophane

Une importante proportion de Grams- est capable de produire les deux composés, incluant les 6 souches *E. coli* présentes dans notre échantillon. Ceci vient appuyer l'hypothèse de leur rôle clé au sein du microbiote, le succinate et le tryptophane étant deux composants importants du métabolisme.

4.4.1 Succinate

Toutes les souches semblent participer à la production directe ou indirecte du succinate, via coopération. Une majorité de Grams- est capable de dégrader le succinate, plusieurs le dégradant via une voie et une réaction actives en parallèle. On peut donc faire l'hypothèse qu'en l'absence des Grams- le succinate est moins dégradé mais que sa production continue et donc il

s'accumule, ce qui correspond à ce qui est observé expérimentalement.

4.4.2 Tryptophane

Une majorité de Grams- semble responsable de sa production. Cette première observation rejoint les résultats expérimentaux, où en absence de Grams- la quantité en tryptophane diminue. Cependant une majorité de Grams- dégrade également le tryptophane. Ainsi, en l'absence de Grams- le tryptophane est moins produit, mais il est aussi moins dégradé. Afin de pouvoir avancer une hypothèse plus d'informations sont nécessaires, comme par exemple la stœchiométrie des réactions et leurs cinétiques. De plus, le tryptophane est un acide aminé essentiel pouvant être en plus utilisé dans la synthèse de protéines. Ce mode de consommation du tryptophane n'est pas pris en compte dans l'étude des voies de dégradation. C'est pourquoi il est difficile dans ce cas de savoir à quoi attribuer une diminution expérimentale de la quantité de tryptophane. Elle peut être due à une baisse de sa production et/ou à une augmentation de son utilisation pour la synthèse de protéines.

4.5 Limites de la reconstruction automatisée

Dans ce travail nous avons reconstruit de manière automatisée 54 réseaux métaboliques en se basant sur les annotations de génomes sélectionnés par les analyses métagénomiques. Afin de simplifier le modèle, une partie seulement des souches sont représentées et les génomes sélectionnés sont ceux représentant au mieux les reads obtenus. Notamment le seuil arbitraire de 0.5% fixé lors de l'étude de métagénomique a pu éliminer de l'étude des organismes peut être essentiels au niveau du métabolisme. Les annotations récupérées et sur lesquelles repose la qualité de reconstruction de l'outil Pathway Tools n'ont pas été complétées et validées manuellement, ce qui aurait été beaucoup trop long. De plus, pour la création de réseaux plus fidèles à la réalité, il aurait fallu compléter les données de métagénomique avec des données de métatranscriptomique afin de n'intégrer dans nos réseaux que les gènes exprimés. Ces multiples approximations peuvent être à l'origine de faux positifs et de faux négatifs dans les prédictions de nos réseaux et il est important de les garder à l'esprit. De plus, le milieu de culture (les sources) est dans ce cas difficile à déterminer précisément et il a un impact direct sur la sortie du réseau (les composés produits).

4.6 Perspectives

Nous étudions ici les interactions au sein d'une communauté bactérienne. Cependant ce travail s'est penché uniquement sur les capacités de dégradations individuelles. Un axe d'étude complémentaire pourrait servir à déterminer si certains organismes sont capables de participer à la dégradation de composés d'intérêts via une coopération. Un travail d'automatisation pourrait

être entrepris afin d'accélérer le travail réalisé manuellement pour l'étude de la dégradation du succinate et du tryptophane. L'objectif serait d'automatiser la récupération des voies et réactions associées à la dégradation d'un grand nombre de métabolites et de les croiser avec les informations des réseaux afin de valider ou non la présence de ces voies chez différentes souches. En ce qui concerne les 3 voies additionnelles de dégradation du tryptophane qui ne sont que partiellement présentes, il est également possible que ces réactions manquantes soient de faux négatifs. Une recherche plus approfondie de la présence ou non de ces réactions serait intéressante. De plus, d'après [1] ces voies sont 3 voies métaboliques majoritaires pour la dégradation du tryptophane dans les intestins, contrôlées directement ou indirectement par le microbiote intestinal. Dans une perspective plus large, le microbiote intestinal réalisant de nombreux échanges bidirectionnels avec son hôte, il serait intéressant dans l'avenir d'ajouter ces échanges à l'étude.

5 Conclusion

L'objectif de ce travail était d'évaluer la qualité des reconstructions automatiques basées sur des données de métagénomique. Les avancées techniques de séquençage, de métagénomique et de biologie computationnelle permettent d'obtenir des informations de plus en plus riches et précises. Dans ce travail, la métagénomique shotgun a permis d'intégrer les organismes non modèles aux résultats. Une classification taxonomique poussée a permis d'identifier les souches principales présentes dans l'échantillon et de récupérer leurs génomes annotés. La biologie computationnelle a réussi à automatiser la reconstruction de réseaux métaboliques en exploitant la richesse des informations disponibles dans les nombreuses bases de données omiques. Via l'étude de la production et de la dégradation d'une sélection de métabolites, les données expérimentales quantitatives et les prédictions qualitatives des réseaux ont pu être mises en relation. La concordance entre ces deux types de données a permis d'émettre plusieurs hypothèses qui pourront servir de pistes de recherche dans des études futures et améliorer notre compréhension du microbiote et de ses interactions avec son hôte par exemple. Etre capable de sélectionner des communautés microbiennes minimales, exploitables et remplissant une fonction précise a d'importantes applications industrielles. La qualité des reconstructions automatiques de GSM est donc cruciale, la curation manuelle de centaines voire de milliers de génomes n'étant pas envisageable. Les limites principales à l'automatisation de la reconstruction sont la détermination des sources et les co-facteurs de chaque organisme, le pH intracellulaire, le sens des réactions ainsi que les données disponibles pour chaque organisme [17]. De plus, il est parfois difficile de valider la qualité des prédictions lorsque les données expérimentales sont difficiles à obtenir, ainsi que de combiner prédictions et expérimentations pour les gros microbiotes [7].

Bibliographie

- [1] Allison Agus, Julien Planchais et Harry Sokol. « Gut Microbiota Regulation of Tryptophan Metabolism in Health and Disease ». en. In : *Cell Host & Microbe* 23.6 (juin 2018), p. 716-724. issn : 19313128. doi : 10.1016/j.chom.2018.05.003. url : <https://linkinghub.elsevier.com/retrieve/pii/S1931312818302579> (visité le 14/06/2022).
- [2] Méziane Aite et al. « Traceability, reproducibility and wiki-exploration for “à-la-carte” reconstructions of genome-scale metabolic models ». en. In : *PLOS Computational Biology* 14.5 (mai 2018). Sous la dir. de Jens Nielsen, e1006146. issn : 1553-7358. doi : 10.1371/journal.pcbi.1006146. url : <https://dx.plos.org/10.1371/journal.pcbi.1006146> (visité le 14/06/2022).
- [3] Arnaud Belcour et al. « Metage2Metabo, microbiota-scale metabolic complementarity for the identification of key species ». en. In : *eLife* 9 (déc. 2020), e61968. issn : 2050-084X. doi : 10.7554/eLife.61968. url : <https://elifesciences.org/articles/61968> (visité le 14/06/2022).
- [4] Ron Caspi et al. « The MetaCyc database of metabolic pathways and enzymes ». en. In : *Nucleic Acids Research* 46.D1 (jan. 2018), p. D633-D639. issn : 0305-1048, 1362-4962. doi : 10.1093/nar/gkx935. url : <http://academic.oup.com/nar/article/46/D1/D633/4559117> (visité le 14/06/2022).
- [5] Unai Escribano-Vazquez et al. « The commensal Escherichia coli CEC15 reinforces intestinal defences in gnotobiotic mice and is protective in a chronic colitis mouse model ». en. In : *Scientific Reports* 9.1 (déc. 2019), p. 11431. issn : 2045-2322. doi : 10.1038/s41598-019-47611-9. url : <http://www.nature.com/articles/s41598-019-47611-9> (visité le 14/06/2022).
- [6] Clémence Frioux. « Investigating host-microbiota cooperation with gap-filling optimization problems ». en. Thèse de doct. Université Rennes 1, nov. 2018. url : <https://hal.inria.fr/tel-01945853> (visité le 14/06/2022).
- [7] Clémence Frioux et al. « From bag-of-genes to bag-of-genomes : metabolic modelling of communities in the era of metagenome-assembled genomes ». en. In : *Computational and Structural Biotechnology Journal* 18 (2020), p. 1722-1734. issn : 20010370. doi : 10.1016/j.csbj.2020.06.028. url : <https://linkinghub.elsevier.com/retrieve/pii/S2001037020303172> (visité le 14/06/2022).

- [8] Clémence Frioux et al. « Scalable and exhaustive screening of metabolic functions carried out by microbial consortia ». en. In : *Bioinformatics* 34.17 (sept. 2018), p. i934-i943. issn : 1367-4803, 1460-2059. doi : 10.1093/bioinformatics/bty588. url : <https://academic.oup.com/bioinformatics/article/34/17/i934/5093211> (visité le 14/06/2022).
- [9] Eman Zakaria Gomaa. « Human gut microbiota/microbiome in health and diseases : a review ». en. In : *Antonie van Leeuwenhoek* 113.12 (déc. 2020), p. 2019-2040. issn : 0003-6072, 1572-9699. doi : 10.1007/s10482-020-01474-7. url : <https://link.springer.com/10.1007/s10482-020-01474-7> (visité le 14/06/2022).
- [10] P. D. Karp. « The MetaCyc Database ». In : *Nucleic Acids Research* 30.1 (jan. 2002), p. 59-61. issn : 13624962. doi : 10.1093/nar/30.1.59. url : <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/30.1.59> (visité le 14/06/2022).
- [11] Peter D. Karp, Mario Latendresse et Ron Caspi. « The Pathway Tools Pathway Prediction Algorithm ». en. In : *Standards in Genomic Sciences* 5.3 (déc. 2011), p. 424-429. issn : 1944-3277. doi : 10.4056/sigs.1794338. url : <http://www.standardsingenomics.org/index.php/sigen/article/view/sigs.1794338> (visité le 14/06/2022).
- [12] Vincent Lacroix. « Identification de motifs dans les réseaux métaboliques : définitions, algorithmes, et application au métabolisme d'*Escherichia coli* ». fre. Thèse de doct. Paris : Edilivre, 2008.
- [13] Jonathan N. V. Martinson et Seth T. Walk. « *Escherichia coli* Residency in the Gut of Healthy Human Adults ». en. In : *EcoSal Plus* 9.1 (mars 2020). Sous la dir. d'Edward G. Dudley, ecosalplus.ESP-0003-2020. issn : 2324-6200. doi : 10.1128/ecosalplus.ESP-0003-2020. url : <https://journals.asm.org/doi/10.1128/ecosalplus.ESP-0003-2020> (visité le 14/06/2022).
- [14] Gianfranco Picone et al. « Metabolomics characterization of colostrum in three sow breeds and its influences on piglets' survival and litter growth rates ». en. In : *Journal of Animal Science and Biotechnology* 9.1 (déc. 2018), p. 23. issn : 2049-1891. doi : 10.1186/s40104-018-0237-1. url : <https://jasbsci.biomedcentral.com/articles/10.1186/s40104-018-0237-1> (visité le 14/06/2022).
- [15] Grégoire Siekaniec et al. « Identification of isolated or mixed strains from long reads : a challenge met on *Streptococcus thermophilus* using a MinION sequencer ». en. In : *Microbial Genomics* 7.11 (nov. 2021). issn : 2057-5858. doi : 10.1099/mgen.0.000654.

- url : <https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000654> (visité le 14/06/2022).
- [16] Masaru Tanaka et Jiro Nakayama. « Development of the gut microbiota in infancy and its impact on health in later life ». en. In : *Allergology International* 66.4 (oct. 2017), p. 515-522. issn : 13238930. doi : 10.1016/j.alit.2017.07.010. url : <https://linkinghub.elsevier.com/retrieve/pii/S1323893017301119> (visité le 14/06/2022).
- [17] Ines Thiele et Bernhard Ø Palsson. « A protocol for generating a high-quality genome-scale metabolic reconstruction ». en. In : *Nature Protocols* 5.1 (jan. 2010), p. 93-121. issn : 1754-2189, 1750-2799. doi : 10.1038/nprot.2009.203. url : <http://www.nature.com/articles/nprot.2009.203> (visité le 14/06/2022).
- [18] Pamela Thomson, Daniel A. Medina et Daniel Garrido. « Human milk oligosaccharides and infant gut bifidobacteria : Molecular strategies for their utilization ». en. In : *Food Microbiology* 75 (oct. 2018), p. 37-46. issn : 07400020. doi : 10.1016/j.fm.2017.09.001. url : <https://linkinghub.elsevier.com/retrieve/pii/S0740002017302198> (visité le 14/06/2022).
- [19] Julie Tomas et al. « Early colonizing *Escherichia coli* elicits remodeling of rat colonic epithelium shifting toward a new homeostatic state ». en. In : *The ISME Journal* 9.1 (jan. 2015), p. 46-58. issn : 1751-7362, 1751-7370. doi : 10.1038/ismej.2014.111. url : <http://www.nature.com/articles/ismej2014111> (visité le 14/06/2022).
- [20] Laszlo Tretter, Attila Patocs et Christos Chinopoulos. « Succinate, an intermediate in metabolism, signal transduction, ROS, hypoxia, and tumorigenesis ». en. In : *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1857.8 (août 2016), p. 1086-1101. issn : 00052728. doi : 10.1016/j.bbabi.2016.03.012. url : <https://linkinghub.elsevier.com/retrieve/pii/S0005272816300597> (visité le 14/06/2022).

A Présentation de la structure d'accueil

L'Inria, l'Institut National de Recherche en Informatique et en Automatique est spécialisé dans le développement d'outils numériques et de nouvelles technologies. Son but est de soutenir la recherche scientifique et de stimuler la création de nouvelles entreprises innovantes. Six centres de recherche sont présents en France, dont le centre Inria Rennes Bretagne Atlantique créé en 1980. Le centre a de nombreuses connections avec divers centres de formation comme l'université de Rennes 1, l'INSA et l'ENS. Il fait également parti d'un joint-venture appelé l'IRISA rassemblant 8 centres de recherche en étroite collaboration. Il possède de nombreux partenariats avec différents établissements de la région comme l'INRAE et le CNRS. Ses principaux axes de recherche sont la sécurité numérique, les interactions hommes-machines, ainsi que la biologie, santé et écologie numérique. Dyliss, une des 29 équipes-projets du centre rennais a pour objectif de structurer et combiner les informations des données génomiques et physiologiques, au niveau des séquences, des espèces ou d'un environnement. Pour cela l'équipe se base sur des méthodes d'intégration de données, d'apprentissage grammatical, de modélisation dynamique et des technologies de web sémantique.

Résumé

Le microbiome intestinal est un écosystème complexe, situé majoritairement dans le colon. Il est très diversifié et contient des centaines de micro-organismes. La biologie des systèmes est un domaine récent de la biologie qui cherche à étudier ces systèmes complexes. Elle va de pair avec la biologie computationnelle qui développe et fournit les outils nécessaires. Un réseau biologique est une représentation abstraite d'un système biologique, rendant compte des éléments le composant et de leurs interactions. Les progrès dans les techniques de séquençage ont permis d'obtenir une quantité d'informations génétiques grandissantes. En combinant cette nouvelle profusion de données au développement des techniques de métagénomique il devient alors possible d'identifier les organismes constituant une communauté et de générer les réseaux métaboliques correspondants. Cependant les données de protéomique et de métabolomique expérimentales sont encore assez limitées dès que l'on sort de l'étude d'organismes modèles. C'est pour cela qu'il est important de construire des stratégies de reconstruction de réseaux métaboliques robustes, capables de palier à ce manque de données. L'objectif global de cette étude est d'évaluer l'efficacité du séquençage Nanopore pour l'analyse métagénomique et la reconstruction de réseaux métaboliques. Ce travail se base sur des données de métagénomique obtenues à partir d'un échantillon de microbiote intestinal de porcelet. Une première étape consistera à générer automatiquement les réseaux métaboliques pour chaque souche identifiée. Dans un second temps les prédictions des réseaux obtenus seront comparées aux données expérimentales, en se focalisant sur deux composés d'intérêt qui sont le succinate et le tryptophane.

Mots clés : MinION, microbiote intestinal, porcelet, succinate, tryptophane

Abstract

Title : Draft reconstructions of metabolic networks from Nanopore and shotgun metagenomic data. The gut microbiome is a complex ecosystem, located mostly in the colon. It is highly diverse and contains hundreds of microorganisms. Systems biology is a recent field of biology that aims to study these complex systems. It goes hand in hand with computational biology which develops and provides the necessary tools. A biological network is an abstract representation of a biological system, accounting for its components and their interactions. Advances in sequencing techniques have made it possible to obtain a growing amount of genetic information. By combining this new profusion of data with the development of metagenomic techniques, it becomes possible to identify the organisms constituting a community and to generate the corresponding metabolic networks. However, experimental proteomics and metabolomics data are still rather limited when it comes to studies out of the model organisms' range. This is why it is important to build robust metabolic network reconstruction strategies that can overcome this lack of data. The overall objective of this study is to evaluate the efficiency of Nanopore sequencing for metagenomic

analysis and metabolic network reconstruction. This work is based on metagenomic data obtained from samples of piglet's gut microbiota. A first step will be to automatically generate the metabolic networks for each identified strain. In a second step, the predictions of the resulting networks will be compared to experimental data, focusing on two compounds of interest which are the succinate and the tryptophan.

Key words : MinION, gut microbiota, piglets, succinate, tryptophane