



HAL
open science

Combinaisons de modélisations en réseaux pour la réinterprétation des règles d'assemblages du microbiome végétal

Victoria Potdevin

► **To cite this version:**

Victoria Potdevin. Combinaisons de modélisations en réseaux pour la réinterprétation des règles d'assemblages du microbiome végétal . Ecologie, Environnement. 2017. dumas-01631009

HAL Id: dumas-01631009

<https://dumas.ccsd.cnrs.fr/dumas-01631009>

Submitted on 8 Nov 2017

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

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

Année universitaire : 2016-2017

Spécialité :

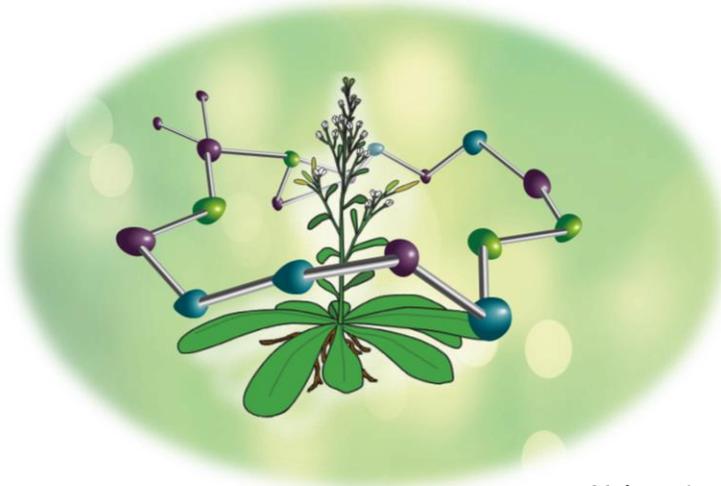
Modélisation en Ecologie (MODE)

Mémoire de fin d'études

- d'Ingénieur de l'Institut Supérieur des Sciences agronomiques, agroalimentaires, horticoles et du paysage
- de Master de l'Institut Supérieur des Sciences agronomiques, agroalimentaires, horticoles et du paysage
- d'un autre établissement (étudiant arrivé en M2)

Combinaisons de modélisations en réseaux pour la réinterprétation des règles d'assemblages du microbiome végétal.

Par : Victoria Potdevin



Maîtres de stage :

- Philippe Vandenkoornhuyse
- Anne Siegel
- Marie Chevallier

Soutenu à Rennes, le 19 juin 2017.

Devant le jury composé de :

Président : Cédric Wolf

Enseignant référent : Frédéric Hamelin

Autres membres du jury :

Nicolas Parisey (rapporteur)

Solenn Stoeckel (rapporteur)

Les analyses et les conclusions de ce travail d'étudiant n'engagent que la responsabilité de son auteur et non celle d'AGROCAMPUS OUEST

Fiche de confidentialité et de diffusion du mémoire

Confidentialité

Non Oui si oui : 1 an 5 ans 10 ans

Pendant toute la durée de confidentialité, aucune diffusion du mémoire n'est possible ⁽¹⁾.

Date et signature du maître de stage ⁽²⁾ : 12 juin 2017

A la fin de la période de confidentialité, sa diffusion est soumise aux règles ci-dessous (droits d'auteur et autorisation de diffusion par l'enseignant à renseigner).

Droits d'auteur

L'auteur ⁽³⁾ Nom Prénom POTDEVIN Victoria

autorise la diffusion de son travail (immédiatement ou à la fin de la période de confidentialité)

oui non

Si oui, il autorise

la diffusion papier du mémoire uniquement⁽⁴⁾

la diffusion papier du mémoire et la diffusion électronique du résumé

la diffusion papier et électronique du mémoire (joindre dans ce cas la fiche de conformité du mémoire numérique et le contrat de diffusion)

(Facultatif) accepte de placer son mémoire sous licence Creative commons CC-By-Nc-

Nd (voir Guide du mémoire Chap 1.4 page 6)

Date et signature de l'auteur : le 22 juin 2017,

Autorisation de diffusion par le responsable de spécialisation ou son représentant

L'enseignant juge le mémoire de qualité suffisante pour être diffusé (immédiatement ou à la fin de la période de confidentialité)

oui non

Si non, seul le titre du mémoire apparaîtra dans les bases de données.

Si oui, il autorise

la diffusion papier du mémoire uniquement⁽⁴⁾

la diffusion papier du mémoire et la diffusion électronique du résumé

la diffusion papier et électronique du mémoire

Date et signature de l'enseignant : le 22/06/17.

(1) L'administration, les enseignants et les différents services de documentation d'AGROCAMPUS OUEST s'engagent à respecter cette confidentialité.

(2) Signature et cachet de l'organisme

(3) Auteur = étudiant qui réalise son mémoire de fin d'études

(4) La référence bibliographique (= Nom de l'auteur, titre du mémoire, année de soutenance, diplôme, spécialité et spécialisation/Option) sera signalée dans les bases de données documentaires sans le résumé

« Je crois que l'avenir de l'humanité est dans le progrès de la raison par la science »
Le Docteur Pascal ; Émile Zola (1983)

En hommage à Marie Bernard, partie trop tôt.

Remerciements

Parce que chacun de vous avez, de près ou de moins près, participé au bon (extra!) déroulement de mon stage, je tiens à commencer mon rapport en vous remerciant.

Je dois te remercier Philippe Vandenkoornhuysen en tout premier, car c'est toi qui m'a permis de réaliser ce stage, et qui a su intégrer mon projet à la dynamique de l'équipe. Toujours disponible, tu as été mon guide ! Tu m'as aussi appris à goûter le vin, ce qui est particulièrement original et donne encore plus de saveur à mon nom de famille !

Marie Chevallier, je ne sais comment te remercier car tu as toujours été là pour me soutenir. En plus d'un aide technique, tu as su m'écouter et répondre à l'ensemble de mes interrogations sur le projet. Tu es douée d'une très grande pédagogie et sans toi, mon stage aurait une bien pâle couleur !

Je ne peux aussi commencer ce rapport sans te remercier Anne Siegel, car ton encadrement régulier m'a aidé à m'organiser, et tu as toujours su m'aider à prioriser les objectifs. Tu m'as aussi fait découvrir l'IRISA et ses équipes, toujours à la pointe de l'innovation.

Je vous remercie tous les trois, aussi pour m'avoir permis de participer à trois événements :

→ une formation de qualité dans l'environnement fascinant de Roscoff, sur la reconstruction métabolique. Je remercie au passage non seulement Marie et Anne mais aussi Camille Trottier, Meziane Aite, Clémence Frioux ainsi que Jeanne Got, la reine du bash Linux (et son bébé qui était déjà là, bien caché) !

→ une enrichissante journée Gen2Bio à Nantes avec Philippe, Marie et ma complice Philomène, durant laquelle Damien Eveillard présentait, lui qui m'a donné plusieurs outils et une aide certaine sur la modélisation en réseaux.

→ trois jours for-mi-da-bles à Paris au MNHN pour le Symposium international sur l'Holobionte, durant lequel j'ai eu l'occasion de découvrir les dernières avancées scientifiques concernant mon sujet d'étude, de rencontrer les plus « grands » dans ce domaine et même de manger à la table de Eugène Rosenberg, Kevin Theis, Paul Schulze-Lefert et Stéphane Hacquard en compagnie de Philippe, Marie, Nathan et Philomène ! Mes félicitations encore à Nathan Vannier, qui a reçu l'approbation divine de Pr. Rosenberg pour sa superbe présentation sur le concept de « méta-holobionte ». Merci Nathan, pour toutes tes aides techniques sur mon projet et ton humour !

L'intégralité des analyses que j'ai réalisé a été permise grâce à la collaboration avec le Max-Planck Institute (Cologne, Allemagne), depuis lequel Ruben Garrido-Oter m'a transmis ses données de thèse avec l'appui de Stéphane Hacquard et de Paul Schulze-Lefert. Danke schön, Sie schaukeln !

Je remercie l'intégralité de l'équipe Dynamo pour tous les bons moments passés, autour d'un café ou d'un verre, et en particulier je n'oublie pas Alix Mas et Lorine Béthencourt qui m'ont aussi beaucoup appris !

Je me permets de faire un petit tour de mes camarades stagiaires, qui ont aussi beaucoup marqué mon expérience, mes camarades de modélisation Philippe Le Noach et Célia Bodin, ma camarade agronome Philomène Brunellière mais aussi dans mon bureau, Inès mon soleil ivoirien, Cédric, Claire, Solène, Kyllian, Virginie et Yamina !

Table des matières

I) Introduction	p. 1
II) Matériel, méthodes et stratégie d'étude	p. 3
a) Les données d'étude	p. 3
b) Les techniques mobilisées et évaluation des mesures	p. 4
i) Normalisation de la table d'abondance par <i>downsampling</i>	p. 4
ii) Mise en évidence d'interactions potentielles entre OTUs par modélisation en réseaux	p. 4
1) Construction d'un réseau de co-occurrences	p. 4
2) Analyses basées sur les similarités entre OTUs	p. 5
3) Interactions potentielles mises en évidence par la matrice de précision	p. 7
iii) Analyse de complétion métabolique entre OTUs: modélisation d'un méta-réseau métabolique	p. 8
III) Résultats	p. 10
a) La richesse fonctionnelle des communautés établies n'est pas maximale	p. 10
b) Modélisation en réseaux des interactions potentielles entre OTUs	p. 12
c) Analyse de la complétion métabolique entre OTUs	p. 13
i) Choix d'un sous-réseau jouet	p. 13
ii) Approche globale	p. 14
iii) Approche ciblée: analyse de la production de phytohormones	p. 15
d) Elaboration d'une stratégie pour la modélisation de type "systems ecology"	p. 15
IV) Discussion	p. 17
a) Rétrospection sur les données d'étude	p. 17
b) Rétrospection sur la démarche et les résultats	p. 18
i) Les possibles raisons d'une richesse fonctionnelle plus faible	p. 18
1) La nature de l'annotation, un élément crucial	p. 18
2) La plante, un censeur potentiel	p. 19
3) La spécialisation de niche reste à explorer	p. 19
ii) Combinaison des approches en réseaux dans le contexte de l'émergence d'une 'systems ecology'	p. 20
1) Un choix impératif parmi les multiples réseaux d'interactions potentielles entre OTUs.	p. 20
2) Retours sur les résultats de la combinaison et appréciation de l'efficacité et de l'aspect innovant de la méthode	p. 20
iii) Un pipeline modulable, orientable vers une démarche d'optimisation théorique du microbiome	p. 21
c) Trois perspectives envisagées	p. 22
i) Intégrer la dimension spatiale	p. 22
ii) Intégrer la dimension temporelle	p. 22
iii) Moduler les conditions environnementales	p. 23
V) Conclusion	p. 23

*Glossaire

- Benchmark : *Anglicisme*. Test de performance.
- Bootstrap : Méthode statistique basée sur un grand nombre de simulations aléatoires.
- Endosphère : Région définie comme l'ensemble des tissus internes à la plante.
- Gnotobiotique : Du grec γνωτός *gnôtós* connu et βίος *bíos* vie.
Se dit d'un organisme dont le microbiote est parfaitement connu et contrôlé ou d'un environnement entièrement fermé, au sein duquel tous les organismes sont identifiés.
- Holobionte: Ensemble d'un hôte et de son microbiote.
- Hologénome: Ensemble du génome de l'hôte et de son microbiome.
- Microbiome : Réunion des génomes du microbiote de l'hôte.
- Microbiote : Ensemble des communautés microbiennes associées à un hôte.
- Neighbour joining : Méthode de regroupement agglomérante itérative qui consiste à joindre le couple de variables ayant la plus forte similarité, considérer ce couple comme une nouvelle variable et recommencer l'opération jusqu'à ce que tous les noeuds du réseau soient liés.
- Omique: *Néologisme*. De l'anglais "Omics".
Se dit de nouvelles technologies permettant une caractérisation globale et la quantification de certaines molécules qui interviennent sur la structure, le fonctionnement et la dynamique d'un organisme.
Les « omiques » les plus connus sont la génomique, la protéomique, la transcriptomique et la métabolomique.
- O.R.F. : De l'anglais 'Open Reading Frame'. Cadre de lecture ouvert ; séquence ADN délimitée par un codon START et un codon STOP susceptible d'encoder de l'ARN.
- Phyllosphère : Surface aérienne de la plante.
- Rhizosphère : Région du sol influencé par les racines de la plante notamment par rhizodéposition d'exsudats, du mucilage et de cellules mortes.

Abréviations:

AuReMe :	AUtomatic REconstruction of MEtabolic models (Aite <i>et al.</i> , <i>in prep.</i>)
ASP :	Answer Set Programming
BiGG :	BIOchemical Genetic and Genomic (King <i>et al.</i> , 2016)
BLAST :	Basic Local Alignment Search Tool
BQH :	Black Queen Hypothesis ; Hypothèse de la Reine Noire (Morris <i>et al.</i> , 2012)
CFU :	Colony Forming Unit
COG :	Cluster of Orthologous Groups of proteins (Tatusov <i>et al.</i> , 2000)
KeGG :	Kyoto Encyclopedia of Genes and Genomes (Kanehisa & Goto, 2000)
FBA :	Flux Balance Analysis
FVA :	Flux Value Analysis
LASSO:	Least Absolute Shrinkage and Selection Operator (Tibshirani, 1996)
SIBL :	Sequence Indexed Bacterial Library (Bai <i>et al.</i> , 2015)
MALDI-TOF :	Matrix Assisted Laser Desorption Ionisation Time-Of-Flight mass spectrometer
MAMPs :	Microbe-Associated Molecular Patterns (Mackey & McFall, 2006)
NCBI :	National Center for Biotechnology Information (U.S.A.)
ORFs :	Open Reading Frames*
OTU :	Operational Taxonomic Unit (Sokal & Sneath, 1963)
PADMeT :	Python library for hAndling metaData of METabolism (Aite <i>et al.</i> , <i>in prep.</i>)
PAMGO :	Plant-Associated Microbe Gene Ontology (Torto-Alalibo <i>et al.</i> , 2009)
PGDB :	Pathway-Genome DataBase
SIG :	Système d'Information Géographique
SparCC :	SPARse Correlation for Compositional data (Friedman et Alm, 2012)
SPIEC-EASI :	SParse InversE Covariance Estimation for Ecological Association Inference (Kurtz <i>et al.</i> , 2015)
WCNA :	Weighted Correlation Network Analysis

Liste des figures :

- Figure 1: Représentation simplifiée de la construction d'un méta-réseau métabolique à partir de deux réseaux métaboliques A et B d'organismes différents. p. 9
- Figure 2: a) Richesse en annotations KeGG des communautés en fonction de leur richesse en OTUs. b) Projection de la densité du *bootstrap*. p. 10
- Figure 3: Réseau d'OTUs basé sur les données d'abondance obtenu avec SPIEC-EASI. p. 11
- Figure 4: Nombre d'OTUs potentiellement dépendants et d'OTUs potentiellement en compétition exclusive pour chaque OTU en fonction de sa présence totale, c'est-à-dire le nombre d'échantillons dans lesquels il est détecté après *downsampling*. p. 12
- Figure 5: Représentation de notre sous-réseau jouet choisi pour effectuer une première analyse de complétion métabolique avec les arêtes SparCC et SPIEC-EASI correspondantes. p. 13
- Figure 6: Schémas de complétion de la voie métabolique de biosynthèse de l'acide jasmonique (PWY-735) pour chacun des trois méta-réseaux métaboliques étudié. p. 16
- Figure 7: Schéma bilan du *pipeline* de modélisation mis en place. p. 16

Liste des tableaux :

- Tableau 1: Liste des mesures de (dis)similarité explorées en premier lieu. p. 5
- Tableau 2: Complétions métaboliques du sous-réseau jouet. p. 14

Liste des annexes:

- Annexe 1: Voie de biosynthèse des jasmonates (PWY-735) (source : MetaCyc)

I) Introduction:

Les nombreuses avancées technologiques et l'accumulation récente de données de séquençage haut-débit dans les domaines de la génomique ont permis la réunion de deux domaines scientifiques à l'origine distincts: la génétique et l'écologie. Jizhong Zhou (2004) redéfinit le terme de 'génomique écologique' et y inclut l'ensemble des études qui visent à la compréhension génomique des bases moléculaires et des mécanismes pour déterminer la composition, la structure et les fonctions d'un système écologique dynamique dans le contexte de l'écologie. On peut alors différencier différents niveaux d'intérêt, la génomique des populations, la génomique des communautés et la génomique écosystémique.

Dans le cadre de la génomique des communautés, le microbiome* végétal constitue un support de recherche intéressant d'un point de vue agronomique, écologique et évolutif en ce sens qu'il est un déterminant clef de la santé et de la productivité d'une plante, responsable au moins en partie de la valeur sélective (*id. fitness*) de son hôte (Vandenkoornhuyse *et al.*, 2015). Il est ainsi qualifié de « deuxième génome » de la plante (Berendsen *et al.*, 2012). En effet, il a récemment été découvert une diversité abondante de bactéries, archaea et champignons qui colonisent les plantes (Bai *et al.*, Nature 2015) et forme le microbiote* végétal.

Toutes les bactéries, et plus largement tous les micro-organismes, en association symbiotique avec un hôte, élaborent des stratégies similaires. Que ce soit adhérer physiquement en premier lieu à la surface ou ensuite pénétrer dans les tissus de l'hôte, elles doivent passer le filtre des défenses de la plante et acquérir des nutriments de son hôte tout en réalisant des fonctions qui leur sont propres (*e.g.* multiplication). En pratique, tous les tissus d'une plante hôte (à l'exception des apex) hébergent des micro-organismes. Un découpage classique est de distinguer celles de la rhizosphère*, de la phyllosphère* et de l'endosphère*. Il a été estimé que plus de dix mille espèces de micro-organismes interviennent seulement au niveau de la rhizosphère de la plante. C'est cette région qui en héberge le plus, ce qui correspond parfois jusqu'à 10^{11} bactéries par gramme de sol rhizosphérique. Ce nombre est fortement réduit dans l'endosphère, la plante exerçant un rôle de filtre environnemental. Le microbiome végétal permet à la plante de tamponner les contraintes environnementales auxquelles elle fait face, le recrutement actif d'un micro-organisme particulier par la plante permettant le recrutement des fonctions qui lui sont associées (Vandenkoornhuyse *et al.*, 2015).

Les bactéries endophytes, moins diverses et moins nombreuses sont considérées comme un sous-groupe particulier de bactéries rhizosphériques (Compant *et al.*, 2010). Ces bactéries capables d'adhérer et de coloniser les tissus de la plante hôte, possèdent un arsenal moléculaire particulier reconnu par le système immunitaire de la plante (MAMPs: *Microbe-Associated Molecular Patterns*) (Monteiro *et al.*, 2012). La production d'analogues d'hormones végétales par des micro-organismes du microbiome qui permet de manipuler le phénotype de la plante hôte a été rapportée de manière récurrente (Ghosh *et al.*, 2015; Gutierrez-Manero *et al.*, 2001; Glick, 2005). La plante, en retour, influe sur les communautés bactériennes qui lui sont liées notamment

par la production de composés antibiotiques tel que des terpènes, des phénols ou des alcaloïdes. Certaines familles végétales comme les brassicacées dont fait partie *Arabidopsis thaliana*, produisent des glucosinolates qui ‘conditionnent’ les communautés microbiennes associées aux racines (Bressan *et al.*, 2009).

On appelle holobionte l’ensemble de l’hôte et de ses micro-organismes, et hologénome l’ensemble des génomes de l’holobionte*. Émise par Zilber-Rosenberg et Rosenberg (2008), la théorie hologénomique de l’évolution considère l’hologénome* comme unité de sélection. De fait, le microbiome est plus réactif (évolutivement parlant) que le génome eucaryote de l’hôte. La capacité d’amplification rapide et un potentiel réactionnel important pour faire face à une contrainte environnementale, facilite l’acclimatation et l’adaptation de l’hôte. En ce sens, il influence significativement des traits de la plante importants vis-à-vis de sa valeur sélective (*i.e.* fitness).

L’observation d’une richesse spécifique élevée dans le microbiome endophyte des plantes ne s’accorde pas en théorie au principe d’exclusion compétitive (Hardin, 1960 ; Levin, 1970). En effet, on peut supposer l’habitat endophyte racinaire comme un environnement peu hétérogène qui contiendrait un nombre limité de niches. Cette complexité théoriquement inattendu du microbiote peut s’expliquer soit i) par un contrôle de la plante, par le système immunitaire ou par allocation dirigée des ressources, ii) par des mécanismes encore inconnus qui limiteraient ces phénomènes d’exclusion, iii) de manière plus vraisemblable, par des phénomènes “d’auto-organisation” pouvant s’expliquer par exemple par un phénomène de complémentarité métabolique (complémentarité fonctionnelle) ou encore par une trajectoire d’évolution réductive de type “*Black Queen Hypothesis*”. Cette trajectoire s’interprète comme un échappement à la compétition par une évolution vers une relation de dépendance par réduction de génome (Mas *et al.*, 2016). Ces hypothèses offrent de nouvelles perspectives de compréhension des règles d’assemblage qui détermine la composition du microbiome endophyte. En conséquence de l’évolution de dépendance, un possible phénomène d’auto-organisation est attendu. Plus généralement, l’hypothèse que les communautés microbiennes endophytes s’assemblent par “auto-organisation” de sorte à optimiser les fonctions biologiques associées semble parcimonieuse. Pouvons-nous alors expliquer les règles d’assemblage des communautés bactériennes du microbiote d’*A. thaliana*, à partir de données omiques*?

Pour tester ces hypothèses, deux approches de modélisation sont développées. La première consiste à construire réseaux de communautés à partir d’une table de contingence pour détecter des interactions clés. La seconde est fonctionnelle et est basée sur l’inférence d’interactions à partir de données omiques. Cela implique la mobilisation de stratégies innovantes par delà les approches actuelles de la biologie des systèmes. En effet, les niveaux d’organisation écologiques imposent le développement du concept émergent de ‘*systems ecology*’ correspondant à l’extension des théories appliquées à la biologie des systèmes (*systems biology*).

Afin de tester ces hypothèses et le travail exploratoire associé, l'analyse de données produites dans une approche expérimentale gnotobiotique* a été envisagée. En s'appuyant sur des inoculations de bactéries connues (séquencées) isolées du microbiote naturel d'*A. thaliana*, on observe les communautés endophytes après un temps de latence. Nous cherchons à comprendre les processus sous-jacents à l'assemblage des communautés du microbiome végétal. Puis, nous initierons une stratégie d'analyse mobilisable pour l'étude d'autres microbiotes et potentiellement pour des travaux de métagénomique. En vue de rendre compte de ce projet, nous présenterons dans un premier temps les données sur lesquelles nous avons basé ce travail, ainsi que nos choix techniques et le développement de méthodologies. Dans un second temps, nous étudierons les résultats obtenus, que ce soit en termes d'analyse ou de modélisation. Enfin, nous tâcherons d'apporter un oeil critique à ces résultats et pourrons ouvrir la discussion vers des perspectives de travail.

II) Matériel, méthodes et stratégie d'étude

Le travail est réalisé dans le cadre d'une collaboration internationale engagée avec le Max Planck Institute de Cologne. Toutes les données ont été produites par ce partenaire. Le jeu de données utilisé, décrit ci-dessous, a fait l'objet d'une première publication (Bai *et al.*, 2015) et est partiellement disponible en ligne sur le site www.at-sphere.com qui recense les données publiques du projet et sur celui du NCBI.

a) Les données d'étude

A l'aide de trois techniques de mise en culture sur agar, 7943 colonies bactériennes ont été isolées de plants sains d'*A. thaliana* et conservées en banque (SIBL – Sequence Indexed Bacterial Library) par plusieurs laboratoires européens dont le Max Planck Institute de Cologne en Allemagne qui s'est chargé des échantillons racinaires. Les 432 isolats les plus abondants (206 issus de feuilles, 194 de racines et 32 de la rhizosphère) ont été soumis à un séquençage intégral de haute-qualité de leur génome afin d'être en mesure de déterminer leurs capacités fonctionnelles. Ces séquençages effectués grâce au système Illumina HiSeq 2500, ont été complétés par des séquençages de leur protéome par spectrométrie de masse avec la technologie MALDI-TOF. La collaboration avec le Max Planck Institute confère un accès NCBI aux 194 génomes (fichiers GenBank .gbff) et protéomes (fichiers FASTA .faa) de microbiotes racinaires entre autres données.

La caractérisation phylogénétique de ces génomes a confirmé que ces isolats (CFUs) couvrent une vaste gamme taxonomique, appartenant à trente-cinq familles bactériennes différentes et réparties sur cinq phyla. Un séquençage indépendant de l'ARNr 16S a permis d'obtenir un tableau de références croisées entre ces isolats et les OTUs correspondants. On possède ainsi l'information taxonomique jusqu'au genre pour tous les OTUs.

L'ensemble des 432 isolats les plus abondants (*n.b.* une communauté synthétique SynCom) a été inoculé en mélange par pulvérisation sur des plants d'*A. thaliana* aseptiques jeunes de trois semaines après germination. C'est le système "flowpots". Quatre semaines plus tard, les racines et les feuilles subissent un séquençage après amplification des ARNr 16S afin d'obtenir une table d'abondance en OTUs. Nous avons accès aux données relatives à 72 échantillons "RSclay". Ils correspondent aux racines et aux feuilles lavées trois fois avec une solution à 10 mM de chlorure de magnésium (MgCl₂) afin de ne garder que les bactéries endophytes.

L'accès à l'information fonctionnelle se fait ici par homologie en alignant (*i.e.* avec BLAST) nos gènes potentiels, ou phase de lectures ouvertes (*id.* ORFs*) à des gènes dont les fonctions biologiques sont déjà connues. De nombreuses bases de données spécifiques au métabolisme cellulaire ont été déployées, comme BiGG (Biochemical Genetic and Genomic, King *et al.*, 2016), KeGG (Kyoto Encyclopedia of Genes and Genomes, Kanehisa & Goto, 2000) ou MetaCyc (Caspi *et al.*, 2014). Elles permettent une description fine de grandes familles d'organismes, entre autres à l'échelle des réactions métaboliques. Ces bases de données sont néanmoins plus ou moins précises et adaptées à certains types d'organismes ou questions biologiques particulières. MetaCyc est, par exemple, plus vaste et riche en réactions et voies métaboliques que KeGG, qui quant à elle comporte davantage de métabolites et davantage validée par les experts (Altman *et al.*, 2013). Une première annotation KeGG des génomes séquencés, sur laquelle nous basons notre analyse, a été effectuée et était disponible en janvier 2017 (Bai *et al.*, 2015).

b) Les techniques mobilisées et évaluation des mesures

Des scripts Python ont été développés, préalablement aux étapes suivantes, afin de reformater les données d'abondances et d'annotations fournies en matrices utilisables par R. Pour cela, l'extension *numpy* (Walt *et al.*, 2011) a été employée.

i) Normalisation de la table d'abondance par *downsampling*

L'effort de séquençage sur les échantillons de table d'abondance en OTUs varie de 2169 à 166522 *reads*. Les échantillons ayant une plus grande profondeur de séquençage ont de fait, une meilleure richesse en OTUs puisque les espèces les moins abondantes sont mieux détectées. En conséquence, on normalise l'effort de séquençage pour tous les échantillons par ré-échantillonnage aléatoire sans remise à 2169 reads.

ii) Mise en évidence d'interactions potentielles entre OTUs par modélisation en réseaux

1) Construction d'un réseau de co-occurrences

La construction de matrices de co-occurrences permet de modéliser la coexistence entre organismes. Pour construire les matrices de co-occurrences à partir des tables de contingence,

nous avons développé un script R et avons adapté nos outils en fonction de la question biologique. A un élément ligne (*e.g.* OTU), elle associe le nombre de fois où il est présent simultanément avec un élément colonne (*e.g.* autre OTU). Les éléments colonnes peuvent être les mêmes que les éléments lignes ce qui crée des matrices symétriques. Les matrices de co-occurrences permettent parfois de retrouver les règles d'assemblage des communautés (Diamond, 1975). Les processus stochastiques qui relèvent de la théorie neutre de l'évolution (Hubbell, 2001) sont quant à eux, des sources d'erreurs aux différentes analyses en réseau explicitées ici. Un réseau non orienté est alors obtenu: les noeuds du réseaux étant les OTUs, et les arêtes correspondant aux liens de co-occurrence (*i.e.* deux noeuds sont reliés par une arête s'ils co-occurrent).

2) Analyses basées sur les similarités entre OTUs

Différentes mesures de similarités existent sur des données d'abondance ou des données binaires. Celles étudiées en premier lieu sont indexées dans le Tableau 1 ci-dessous.

Tableau 1: Liste des mesures de (dis)similarité explorées en premier lieu, ainsi que leur formule mathématique, le package R employé, leur intervalle de définition et la matrice sur laquelle ont été opérées ces mesures. x_1 et x_2 sont les abondances respectives des OTUs 1 et 2 dans l'échantillon g parmi un total de P échantillons.

	Formule	Package R	Intervalle de définition	Matrice d'origine
Coefficient de corrélation de Pearson	$r(x_1, x_2) = \frac{Cov(x_1, x_2)}{\sigma_{x_1} * \sigma_{x_2}}$	stats	[-1 ; 1]	Abondance
Coefficient de corrélation de Spearman	$\rho(x_1, x_2) = \frac{6 * \sum_{g=1}^P (rang(x_{1g}) - rang(x_{2g}))^2}{P^3 - P}$	stats	[-1 ; 1]	Abondance
Distance euclidienne	$de(x_1, x_2) = \sqrt{\sum_{g=1}^P (x_{1g} - x_{2g})^2}$	vegan ¹	[0 ; +∞[Abondance
Distance de Canberra	$dc(x_1, x_2) = \sum_{g=1}^P \frac{ x_{1g} - x_{2g} }{ x_{1g} + x_{2g} }$	vegan ¹	[0 ; +∞[Abondance
Distance de Manhattan	$dm(x_1, x_2) = \sum_{g=1}^P x_{1g} - x_{2g} $	vegan ¹	[0 ; +∞[Abondance
Indice de Dice-Sorensen	$s(x_1, x_2) = \frac{2 * n_{x_1>0 \& x_2>0}}{n_{x_1>0} + n_{x_2>0}}$	ade4 ²	[0 ; 1]	Présence

Indice de dissimilarité de Bray-Curtis	$BC(x_1, x_2) = 1 - \frac{2 * \sum_{g=1}^P \min(x_{1g}, x_{2g})}{\sum_{g=1}^P x_{1g} + x_{2g}}$	vegan ¹	[0 ; 1]	Abondance
Indice de Jaccard	$j(x_1, x_2) = \frac{n_{x_1>0 \& x_2>0}}{n_{x_1>0 x_2>0}}$	vegan ¹	[0 ; 1]	Présence

¹ Orksanen *et al.*, 2007

² Dray *et al.*, 2007

En premier lieu, les indices de corrélation de Pearson et de Spearman sont biaisés par nature, étant donné qu'ils concernent ici des données de composition. En effet, le nombre de reads est le même par échantillon, ce qui peut créer des corrélations artéfacts. Pearson avertissait en 1896 des risques liés à une mauvaise interprétation des corrélations entre ratios dont les numérateurs et les dénominateurs contiennent une ou des partie(s) commune(s). De plus, ces mesures sont aussi biaisées en ce sens qu'elles considèrent deux absences simultanées comme source de corrélation. Ceci n'a pas de sens biologique à l'échelle de l'OTU (contrairement éventuellement à celle de l'isolat), d'autant plus qu'un OTU est considéré comme absent à partir du moment où il n'est pas suffisamment abondant pour être détecté avec la profondeur de séquençage.

Conscients de ces limites lorsqu'il s'agit de données d'abondances obtenues par séquençage, Friedman et Alm (2012) ont développé la méthode SparCC (SPARSe Correlation for Compositional data) qui utilise la corrélation de Pearson sur les données préalablement transformées logarithmiquement. Cette méthode est une mise en pratique de la suggestion mathématique de Aitchison (1981). Elle peut être employée sous l'hypothèse que le nombre d'OTUs est important et que le réseau est peu dense (sparse). Elle repose sur l'étude de la variable t_{ij} définie telle que $t_{ij} = Var \left[\log \frac{x_i}{x_j} \right] = \omega_i^2 + \omega_j^2 - 2\rho_{ij}\omega_i\omega_j$, avec ω_i^2 et ω_j^2 les variances réciproques des abondances en base log des OTUs i et j . On peut donc interpréter cette variable de telle sorte que $t_{ij} < \omega_i^2 + \omega_j^2$ correspond à une corrélation positive et $t_{ij} > \omega_i^2 + \omega_j^2$ correspond à une corrélation négative. L'estimation de ces variances et des corrélations est permise sous l'hypothèse que la majorité des abondances des OTUs sont non corrélées donc SparCC n'est censé révéler que les interactions potentielles les plus fortes. De plus, SparCC inclut une méthode bayésienne (*i.e.* basée sur les probabilités conditionnelles) pour éviter de recourir à du *downsampling* et ainsi éliminer les OTUs les plus rares (*cf.* Partie II-b-i). Cependant, comme l'étude est déjà basée sur les OTUs les plus abondants, il a été préféré de maintenir le *downsampling* par souci de comparaison et d'interprétation des données. La sortie de SparCC est une matrice de corrélation qualifiée de "robuste".

L'intégralité des mesures exprimées précédemment sont définies sur un intervalle de définition, sur lequel elles sont continues. De façon à transformer cette information en réseau de corrélation, qui la simplifie, il est nécessaire de définir un seuil à partir duquel les deux OTUs

sont considérés significativement similaires ou significativement dissimilaires. Cela revient à faire de l'analyse de corrélation pondérée en réseau (WCNA).

Pour chacune des mesures obtenues, aucune n'avait de distribution singulière. Pour pallier à cela, deux méthodes ont été envisagées:

- ne considérer pour chaque OTU que l'OTU avec lequel il est le plus similaire et celui avec lequel il est le moins similaire;
- ne retenir que les 5% ayant les plus forts et les 5% ayant les plus faibles indices de similarité.

La première méthode suppose que tous les OTUs soient connectés, ce qui maximise la perte d'information liée à la mesure. Pour éviter cela, la seconde méthode a été choisie et mise en place sur R.

3) Interactions potentielles mises en évidence par la matrice de précision

Kurtz et al. (2015) ont fait le même constat que Friedman et Alm (2012) sur les biais liés au caractère compositionnel des données d'abondance. Dans le but d'obtenir un réseau d'interactions potentielles (information simplifiée), ils ont développé une méthode statistique nommée SPIEC-EASI (SParse Inverse Covariance Estimation for Ecological Association Inference) implémentée sur R. SPIEC-EASI s'adapte tout particulièrement aux données d'abondances pour lesquelles le nombre d'OTUs est supérieur au nombre d'échantillons. Cette méthode s'appuie sur le principe de parcimonie. En effet, selon les développeurs, la corrélation n'est pas une bonne mesure de l'association en ce sens qu'elle peut résulter d'interactions indirectes dans les communautés. Deux noeuds (*e.g.* OTUs) sont indépendants conditionnellement, si l'état (*e.g.* abondance) de l'un ne donne pas d'information supplémentaire sur l'état de l'autre, dans un réseau donné. Le but est d'obtenir un réseau non orienté, de type "champs aléatoire de Markov". De fait, un lien établi en deux OTUs signifie que leurs abondances sont liées par une relation linéaire qui ne peut pas être expliquée par un réseau alternatif ayant une connectique différente. D'une façon similaire à celle de SparCC, les données sont en premier lieu transformées en log-ratio et centrées. Ensuite, chaque noeud (*e.g.* OTU) fait l'objet d'une régression linéaire qui s'apparente à du *neighbour joining**. Ceci est complété par une sélection par maximum de vraisemblance. Le procédé revient à résoudre un problème d'optimisation, appelé "problème de sélection de la covariance" par Dempster (1972). De cette façon, les éléments non-nuls de la matrice de précision (inverse de la matrice de covariance) déterminent les arêtes du réseau. Pour résumer, SPIEC-EASI cherche à résoudre par un problème d'optimisation convexe la matrice de précision, qui correspond à la matrice d'adjacence du réseau. Deux solveurs sont intégrés à l'outil: LASSO (Tibshirani, 1996) et la méthode du cadre de sélection par voisinage aussi appelée méthode MB (Meinshausen and Bühlmann, 2006). Dans leur benchmark*, Kurtz *et al.* (2015) ont montré que la méthode MB était la plus performante. Particulièrement sensible, elle a le moins de faux positifs dans tous les cas testés, c'est pourquoi

la méthode MB a été choisie dans notre étude. Les paramètres par défaut ont été employés ($n\lambda=20$; $\lambda.\text{min.ratio}=10^{-2}$), car il semblerait, après avoir testé une centaine de combinaisons, que ce soient les plus efficaces vis-à-vis du temps de calcul, pour des résultats sans variation visible.

Enfin, une dernière méthode a été développée pour détecter les éventuelles dépendances entre OTUs : pour chaque OTU, on vérifie s'il n'est présent que si et seulement si un autre OTU est présent à travers les différents échantillons. Cela peut éventuellement permettre de détecter des dépendances entre OTUs, de type BQH si l'on peut descendre à l'échelle de l'isolat. Cette méthode n'est pas exhaustive compte-tenu de la différence d'échelle isolat/OTU. En effet une dépendance entre isolats peut être masquée par une présence dans un autre échantillon d'un autre isolat du même OTU. On obtient ainsi une matrice de "dépendances potentielles". Afin de simplifier la lecture du réseau orienté, on considère que (i) si l'OTU A n'est présent qu'en présence de B (ii) si l'OTU B n'est présent qu'en présence de C (iii) alors on ne prends pas en compte l'arête qui relie A à C, puisque A peut dépendre indirectement de C par B. De la même manière on considère une matrice d'exclusion réciproque, qui détecte les OTUs qui ne sont jamais présents simultanément. On s'attend notamment à ce que les OTUs liés par une relation de dépendance se complètent partiellement métaboliquement parlant à l'échelle des fonctions (et réciproquement que ceux qui s'excluent ne se complètent pas de façon optimale).

iii) Analyse de complétion métabolique entre OTUs: modélisation d'un méta-réseau métabolique

Une fois les connections entre OTUs identifiées, on cherche à expliquer à une échelle plus fine (métabolique) la nature de leur interaction. L'hypothèse qui suppose que les OTUs très connectés se conjuguent de manière optimale, nécessite une approche complexe pour être vérifiée. Afin de savoir si les communautés stabilisées optimisent en terme de nombre, tout d'abord et plus simplement, les "fonctions", en se basant sur la table d'annotations KeGG fournie, un grand nombre (10^4 par richesse) de communautés synthétiques en OTUs ont été simulées. Pour chaque communauté synthétique d'une richesse donnée en OTUs, les OTUs ont été tirés aléatoirement parmi les 432 isolats de l'inoculum. Au sein d'un OTU, il y a en moyenne moins de 19,6% de variabilité sur les annotations KeGG entre les isolats. En outre, nous ne pouvons obtenir l'information d'abondance à l'échelle de l'isolat, c'est pourquoi on considère qu'à partir du moment où un OTU est présent, l'ensemble des annotations KeGG possédées par au moins un de ses isolats sont représentés dans la communauté. On normalise (centrage-réduction) alors chaque valeur obtenue pour les communautés observées par rapport à la moyenne et à l'écart-type du *bootstrap** obtenu à une richesse donnée.

Comme expliqué précédemment, bien que plus contrôlée, la base KeGG est moins riche que la base MetaCyc (Altman *et al.*, 2013). Afin d'obtenir une annotation enrichie MetaCyc, l'outil Pathway Tools est ici employé. Cet outil prend en entrée nos fichiers GenBank et FASTA

(génomique et protéomique), ainsi que l'identifiant NCBI du taxon le plus proche. La sortie obtenue est un ensemble de fichiers de divers formats PGDB (*Pathway-Genome DataBase*) contenant l'ensemble des informations métaboliques prédites pour chaque isolat.

Une des modélisations métaboliques classiques se fait grâce à des graphes bipartites orientés et pondérés, dont les noeuds sont des réactions et des composés. De très nombreux outils permettent, à partir d'annotations fonctionnelles, de reconstituer des réseaux métaboliques. La solution AuReMe (AUtomatic REconstruction of MEtabolic models) (Aite *et al.*, *in prep.*) propose un espace de travail modulaire, unifié et adaptable pour la reconstruction reproductible de réseaux métaboliques à l'échelle génomique. Les reconstructions peuvent être générées en utilisant différents formats et outils de données à travers des *pipelines* personnalisables. Dans chaque étape, les informations pertinentes sont stockées, ce qui garantit la reproductibilité et traçabilité du processus. De plus, l'espace de travail établit une nouvelle façon d'explorer, de visualiser et de distribuer les modèles à l'échelle du génome et leurs métadonnées grâce à la génération de *wikis* locaux *ad hoc*. Parmi les applications pertinentes, cet espace de travail ouvre la possibilité de gérer, de manière intégrée, la reconstruction métabolique d'organismes inexplorés ou de communautés microbiennes dans des échantillons environnementaux par la fusion des réseaux métaboliques afin d'obtenir des méta-réseaux métaboliques (Figure 1). Ainsi, cela permet d'analyser la complétion du réseau par les différents isolats candidats d'origine.

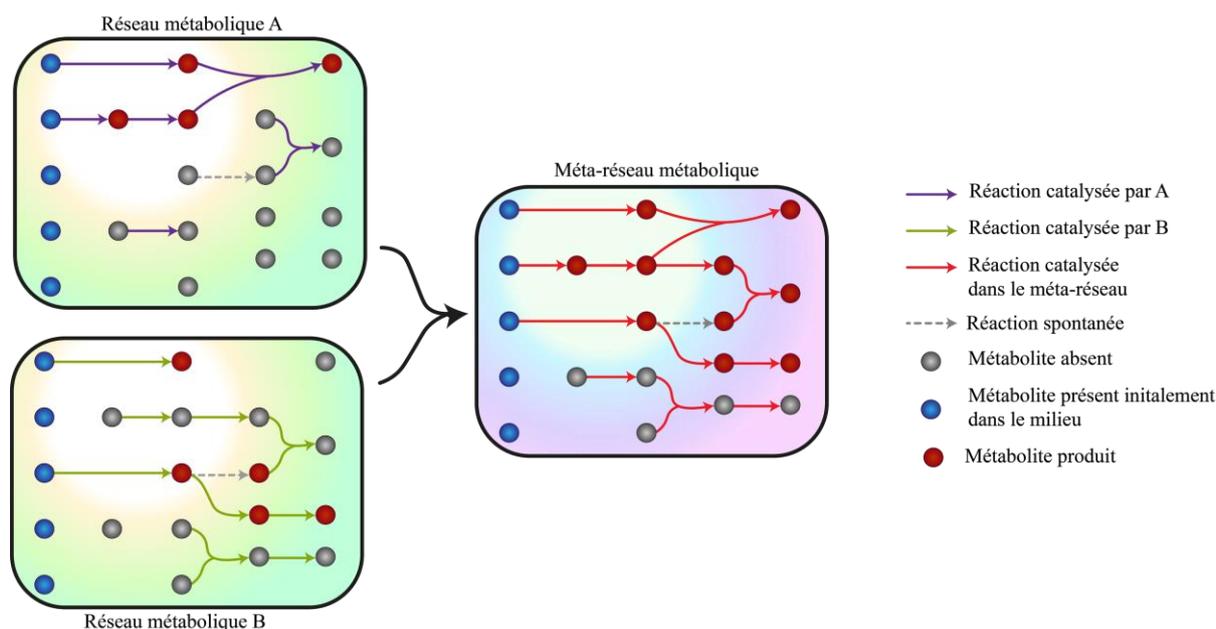


Figure 1: Représentation simplifiée de la construction d'un méta-réseau métabolique (i. e. on suppose les réactions de transport non limitantes) à partir de deux réseaux métaboliques A et B d'organismes différents.

Trois méta-réseaux métaboliques ont été reconstruits pour l'analyse de chaque couple d'OTUs pour lesquels une association de dépendance a été mise en évidence à l'étape précédente

(cf. Partie II-B-ii-3). Les deux premiers représentant chacun un OTU tandis qu'un troisième représente la fusion de ces deux premiers niveaux de méta-réseaux métabolique. Ceci permet de savoir quelles voies métaboliques sont partiellement ou complètement représentées, et quelle est la part de chaque OTUs dans la complétion de ces voies métaboliques. Les méta-réseaux métaboliques peuvent donc être calculés à partir des couples d'OTUs pour un ensemble de micro-organismes partageant un habitat.

III) Résultats

a) La richesse fonctionnelle des communautés établies n'est pas maximale

La richesse en annotations KeGG des communautés observées comparativement à celles de données simulées aléatoirement est illustrée Figure 2. Un test non paramétrique de Student indique que les communautés observées comptent significativement moins d'annotations KeGG que les données simulées ($P=2,7 \cdot 10^{-4}$) avec 62,26% des valeurs observées en deçà de la médiane du *bootstrap*, bien que leur densités ne soient pas radicalement différentes. Ceci rejette l'hypothèse selon laquelle les communautés s'organisent de manière à maximiser leur richesse fonctionnelle à l'échelle métabolique. Néanmoins, cela n'exclut pas l'existence de synergies particulières à cette échelle favorisant l'assemblage de ces communautés.

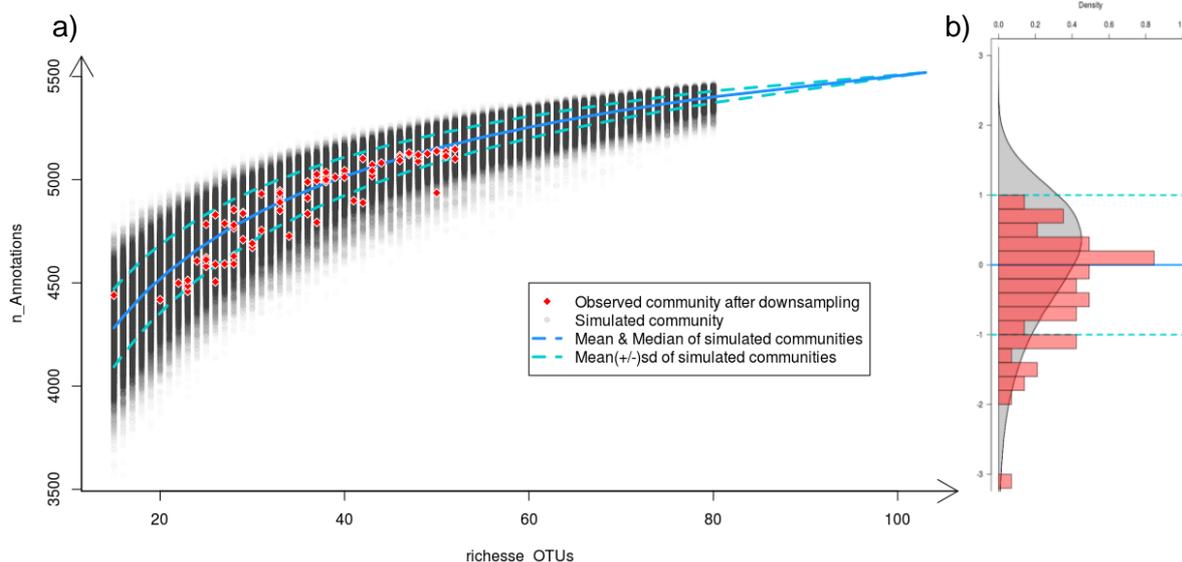


Figure 2: a) Richesse en annotations KeGG des communautés en fonction de leur richesse en OTUs. En rouge le nombre d'annotations KeGG pour les communautés observées après downsampling, et en gris celles pour les communautés simulées par bootstrap, en fonction de la richesse en OTUs. b) Projection de la densité du bootstrap en gris et des communautés observées en rouge (dans chacun des deux cas, les données sont centrées-réduites par rapport à la moyenne spécifique à leur richesse en OTUs).

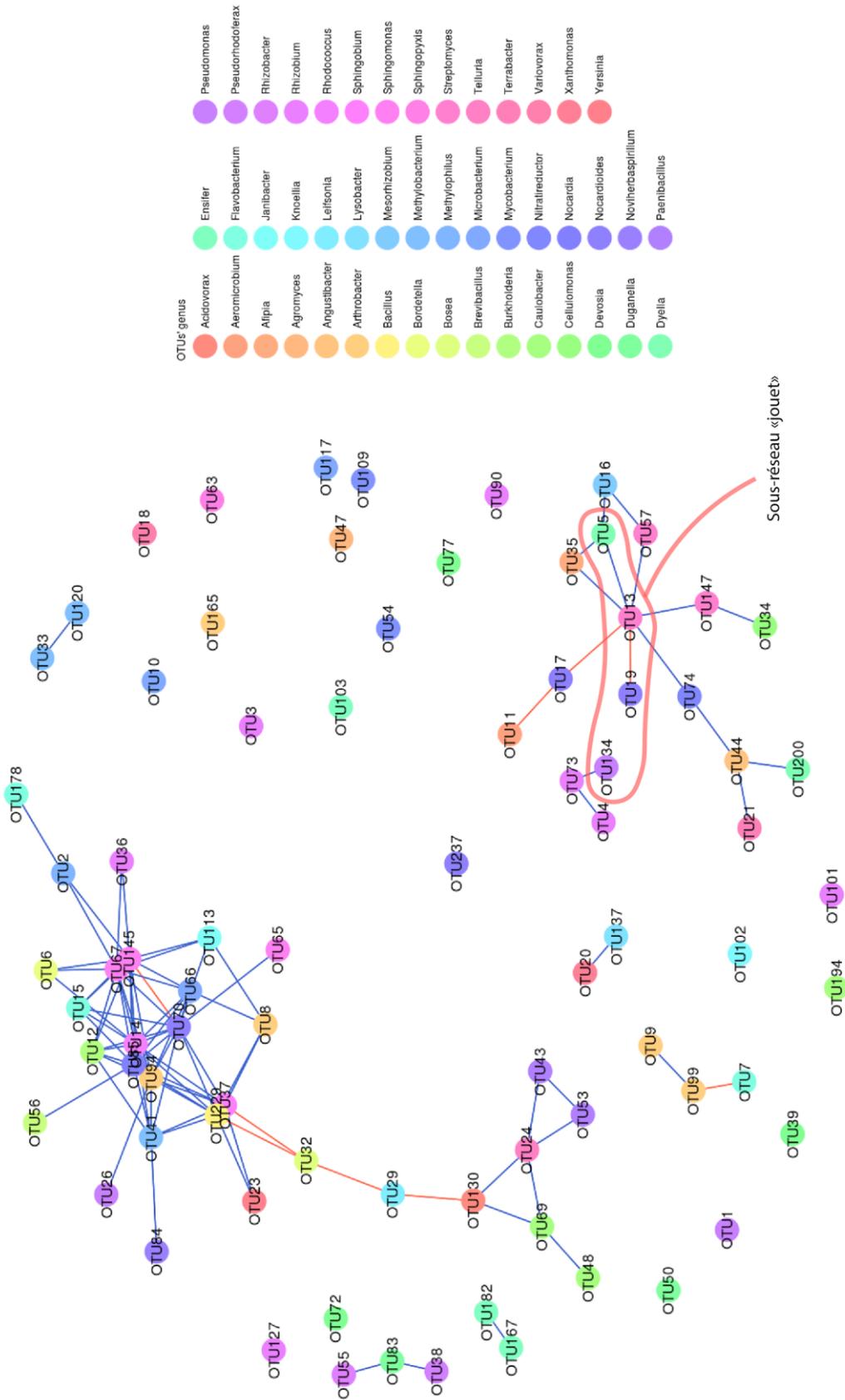


Figure 3: Réseau d'OTUs basé sur les données d'abondance obtenu avec SPIEC-EASI. Afin de déterminer la nature de l'interaction, en rouge sont représentées les arêtes dont le score SparCC est négatif et réciproquement, en bleu celles pour lesquelles il est positif.

b) Modélisation en réseaux des interactions potentielles entre OTUs

Afin de mettre en évidence les potentielles synergies entre OTUs (deux à deux), on identifie ici les interactions les plus probables et les plus fortes afin de cibler l'analyse de complétion métabolique, encore trop complexe si elle est abordée de manière globale.

Les mesures de similarités entre OTUs détaillées Tableau 1 donnent des réseaux pondérés qui comportent le même nombre d'arêtes étant donné que seuls les 10% couples les plus similaires sont considérés. Si l'on compare ces réseaux, on constate qu'il sont très différents. En effet, ils n'ont en moyenne que 11,5% d'arêtes communes pondérés dans le même sens, contre 5,7% d'arête communes identifiées dans des pondération opposées. Ces réseaux extrêmement dissimilaires montrent l'importance d'une mesure plus appropriée. SparCC est radicalement différent de la majorité des autres mesures, mais partage tout de même 29,8% d'arêtes communes pondérées dans le même sens pour les corrélations de Spearman et 42,3% pour les corrélations de Pearson. Plus de 57% des interactions identifiées par SparCC ne le sont pas avec ces corrélations. 28,8% des arêtes de SPIEC-EASI (cf. Figure 3) se retrouvent dans le réseau SparCC (réciproquement 10,4%).

Si l'on étudie le réseau de dépendances et d'exclusions réciproques potentielles obtenu, on ne constate aucune 'anomalie', en ce sens que les résultats sont mathématiquement attendus (Figure 4). Plus un OTU est présent dans de nombreux échantillons, plus il a de potentiels 'dépendants' (OTUs présents uniquement en sa présence) et moins il a de potentiels compétiteurs exclusifs (OTUs qui ne se retrouvent jamais dans les mêmes échantillons). Pour renforcer ce résultat, une étude qui viserait à ré-organiser aléatoirement les OTUs en conservant le même nombre de présence dans les échantillons est envisagée. Ceci serait complété par une comparaison des coefficients obtenus après régression logistique.

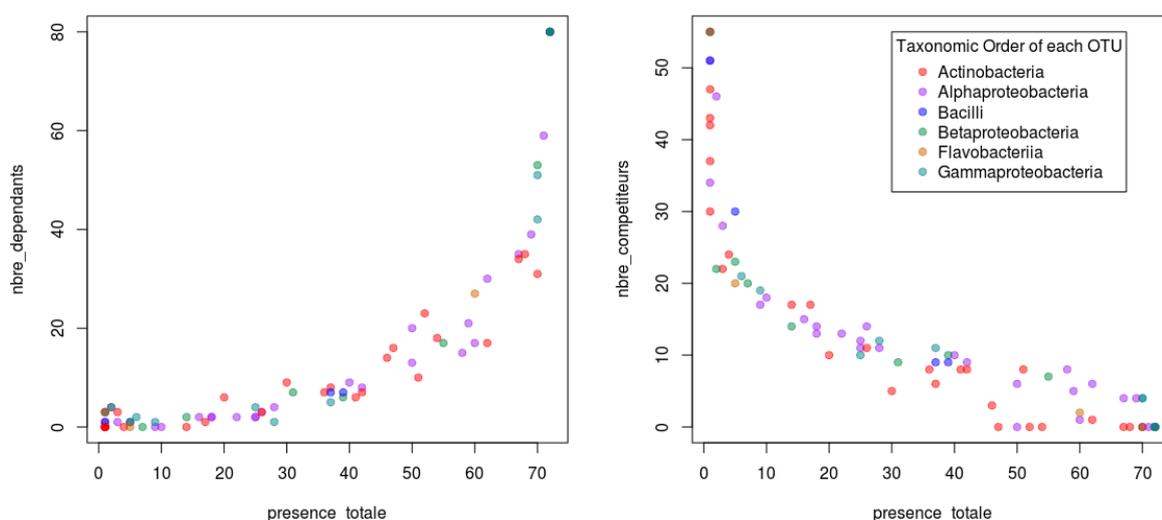


Figure 4: Nombre d'OTUs potentiellement dépendants et d'OTUs potentiellement en compétition exclusive pour chaque OTU en fonction de sa présence totale, c'est-à-dire le nombre d'échantillons dans lesquels il est détecté après downsampling.

c) Analyse de la complétion métabolique entre OTUs

i) Choix d'un sous-réseau jouet

Avant de réaliser une analyse globale des complétions métaboliques, il a été choisi d'étudier un sous-réseau jouet d'OTUs comme preuve de concept. C'est cette partie qui est détaillée ici. Parmi les hubs révélés par l'analyse du réseau SPIEC-EASI (Figure 3), la stratégie a été de travailler sur un hub particulier. Ce hub devait comporter un nombre limité d'arêtes et contenir peu d'isolats. Ce hub devait permettre de tester l'hypothèse d'auto-organisation entre les micro-organismes du hub pour tester l'hypothèse $H_{0(1)}$ de complémentarité métabolique (Figure 1). De plus, un OTU devait comporter plusieurs isolats de la même espèce de manière à pouvoir comparer les compositions génomiques (comparaison d'annotations KeGG sous *R* pour l'hypothèse $H_{0(2)}$, l'existence de phénomènes d'évolution réductive pour échapper à l'exclusion compétitive de type BQH. De plus, il faut posséder les données omiques des isolats de ce hub d'OTUs. Seul l'OTU19 (*Nocardioides spp.*) satisfait cette dernière condition et est un bon candidat pour l'hypothèse $H_{0(2)}$. En effet, d'après la table d'annotations KeGG, l'isolat *Root79* ne posséderait pas deux protéines: la phosphatase-NUDJ (K12152) et l'inositol-phosphate transport system ATP-binding (K17240) que posséderait *Root240*, un autre isolat le l'OTU19. Le hub porté principalement par l'OTU13 (*Streptomyces spp.*) a donc été choisi car il satisfait ces conditions, étant donné que l'OTU19 est lié de façon unique à ce dernier dans le réseau SPIEC-EASI. Nous nous sommes focalisés en premier lieu sur un sous-réseau de ce hub: l'OTU13, l'OTU19 et l'OTU5 (*Dyella spp.*).

L'OTU19 et l'OTU13 sont corrélés négativement à -0,55 avec SparCC et le couple OTU13-OTU5 est corrélé positivement avec un coefficient SparCC de 0,76. Afin de compléter notre analyse et de comparer avec le réseau SparCC, en se basant sur notre OTU19 d'intérêt, on intègre l'OTU134 (*Pseudomonas spp.*) avec lequel il est fortement corrélé (à 0,51) à notre sous-réseau (Figure 5), bien que cette interaction ne soit pas détectée par SPIEC-EASI. En outre, l'OTU134 est potentiellement dépendant de l'OTU19, en ce sens qu'il n'est présent que dans les échantillons où l'OTU19 est présent.

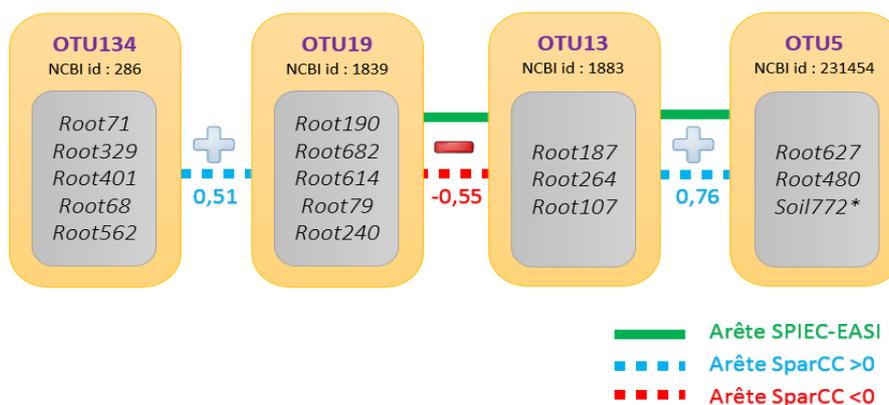
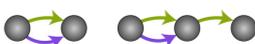
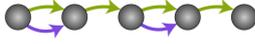


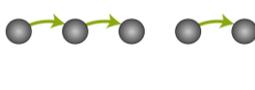
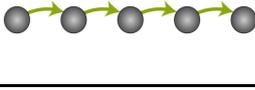
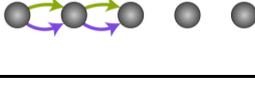
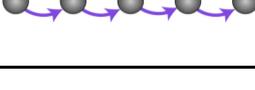
Figure 5: Représentation de notre sous-réseau jouet choisi pour effectuer une première analyse de complétion métabolique avec les arêtes SparCC et SPIEC-EASI correspondantes. Dans les rectangles gris sont inscrits les identifiants des isolats qui composent l'OTU.

ii) Approche globale

Si l'on considère qu'il y a complétion à proprement parler entre deux OTUs, cela signifie qu'ils contribuent chacun individuellement à au moins une réaction de la voie métabolique considérée. D'après la table de sortie obtenue (Tableau 2), il y a des cas de complétions pour les trois couples étudiés. 18 voies métaboliques sont concernées entre l'OTU13 et l'OTU19 qui sont anti-corrélé très fortement, contre 30 et 48 pour les couples corrélés positivement, respectivement OTU13 et OTU5 puis OTU134 et OTU19. Une seule voie métabolique est alors fonctionnelle (complète) dans le premier cas contre cinq et six pour les deux autres (Tableau 2). Cette voie métabolique courte (deux réactions) est la dégradation III de la tréhalose en β -D-glucose-6-phosphate, que la plante produit déjà lors de la photosynthèse. Toutefois, les voies métaboliques complètes pour les couples corrélés positivement n'ont pas tous été décrits chez les plantes ^{et/ou} les bactéries, (certaines uniquement chez les mammifères comme la dégradation II de la glycine bétaine). Par contre, deux voies métaboliques sont complètes dans l'association OTU19 et OTU134 qui correspondent au cycle de l'acide tricarboxylique, proches du cycle de Krebs qui produisent de l'énergie et du pouvoir réducteur. Le cycle de l'assimilation du carbone (cycle de la photosynthèse des plantes en C4) est aussi en théorie complet dans les associations chez les deux couples d'OTUs corrélés positivement. Il ne faut pas oublier qu'ici on a étudié deux à deux les OTUs, mais la complétion sera probablement plus importante à l'échelle d'un *hub* complet.

Tableau 2: Complétions métaboliques du sous-réseau jouet. Chaque type de complétion représenté à gauche est défini par un nombre de sources uniques (nombre d'OTU(s) apportant au moins une réaction qui lui est spécifique), le nombre de sources totales, et si la voie métabolique est complète. Le ratio dans les trois dernières colonnes indique le niveau de complétion et le nombre au dessus du ratio indique le nombre d'étapes réalisées pour la voie métabolique par la paire d'OTUs.

Schéma de complétion	Nombre de sources uniques	Nombre de sources	Voie complète			
	2	2	non	25 0,58	17 0,53	42 0,56
	2	2	oui	5 1,00	1 1,00	6 1,00
	1	2	non	165 0,52	113 0,54	171 0,54
	1	2	oui	125 1,00	71 1,00	118 1,00

	1	1	non	245 0,32	204 0,31	204 0,31
	1	1	oui	109 1,00	67 1,00	108 1,00
	0	2	non	296 0,43	416 0,42	296 0,40
	0	2	oui	157 1,00	219 1,00	200 1,00

iii) Approche ciblée: analyse de la production de phytohormones

Nous avons donc cherché à étudier particulièrement les voies métaboliques associées à des composés d'intérêt pour la plante, en particulier les hormones végétales, afin d'évaluer si la nature des voies métaboliques est un élément déterminant de l'assemblage de ses communautés microbiotiques. Sur les sept voies métaboliques de la biosynthèse de l'auxine indexées par MetaCyc, six sont partiellement constituées dans les quatre OTUs d'intérêt. La complétion totale de la biosynthèse de l'auxine V est obtenue pour les différentes paires d'OTUs étudiés alors que isolément seuls deux d'entre eux possède l'unique enzyme de cette voie. Un exemple marquant (bien qu'il ne corresponde pas à de la complétion au sens propre du terme) est la dégradation VII du L-tryptophane. Ce dernier est précurseur de l'auxine. Cette voie est constituée de trois réactions successives, dont seuls les OTU5 et OTU134 possèdent l'enzyme tryptophane-aminotransférase. Donc les deux couples corrélés positivement catalysent cette réaction mais pas le couple OTU19/OTU13 corrélé négativement. De plus, on a une meilleure complétion pour les biosynthèses de l'auxine II dans le couple OTU134 et OTU19 que dans les autres couples. Aucun enzyme associé aux voies de biosynthèse de l'acide abscissique ou des gibbérellines n'a été retrouvé. L'ensemble des voies métaboliques associées à la biosynthèse de l'éthylène sont représentées partiellement pour les quatre OTUs, mais ne se complètent pas mutuellement. Pour les jasmonates, comme pour les cytokinines et les brassinostéroïdes, une seule des voies métaboliques est représentée. La biosynthèse de l'acide jasmonique (voie métabolique détaillée en Annexe 1) présente un taux de complétion par couple d'OTUs variant de 0,58 à 0,79 (Figure 6).

c) Élaboration d'une stratégie pour la modélisation de type "systems ecology"

Les différentes analyses décrites précédemment sont le résultat d'une démarche exploratoire ayant pour objectif le test de notre hypothèse de départ. La stratégie globale du plan expérimental, avec ses différentes étapes, est explicitée Figure 7.

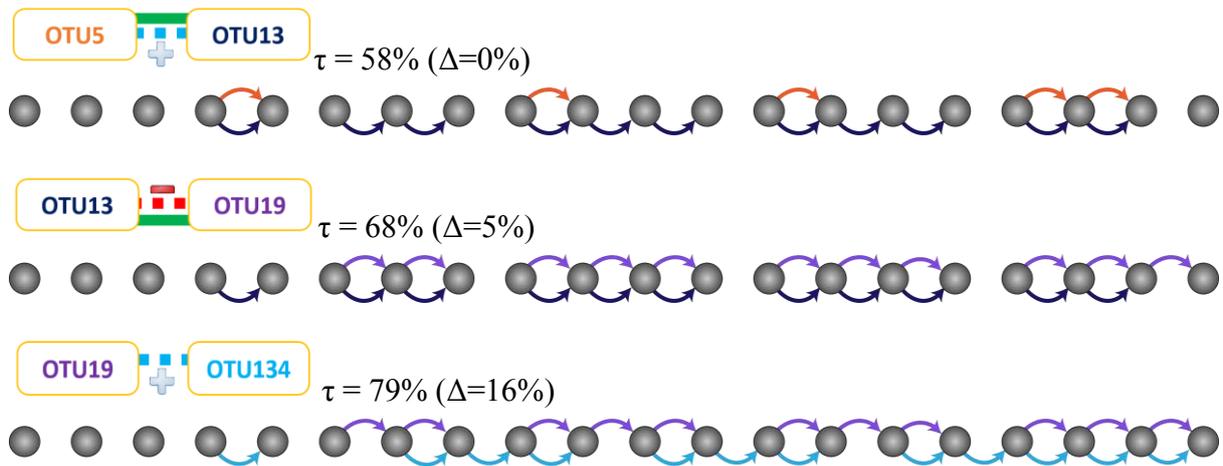


Figure 6: Schémas de complétion de la voie métabolique de biosynthèse de l'acide jasmonique (PWY-735) pour chacun des trois méta-réseaux métaboliques étudiés. Leur taux de complétion τ est indiqué, ainsi que leur 'gain' de complétion $\Delta = \tau_{\text{méta-réseau}} - \max(\tau_{\text{réseau1}}, \tau_{\text{réseau2}})$.

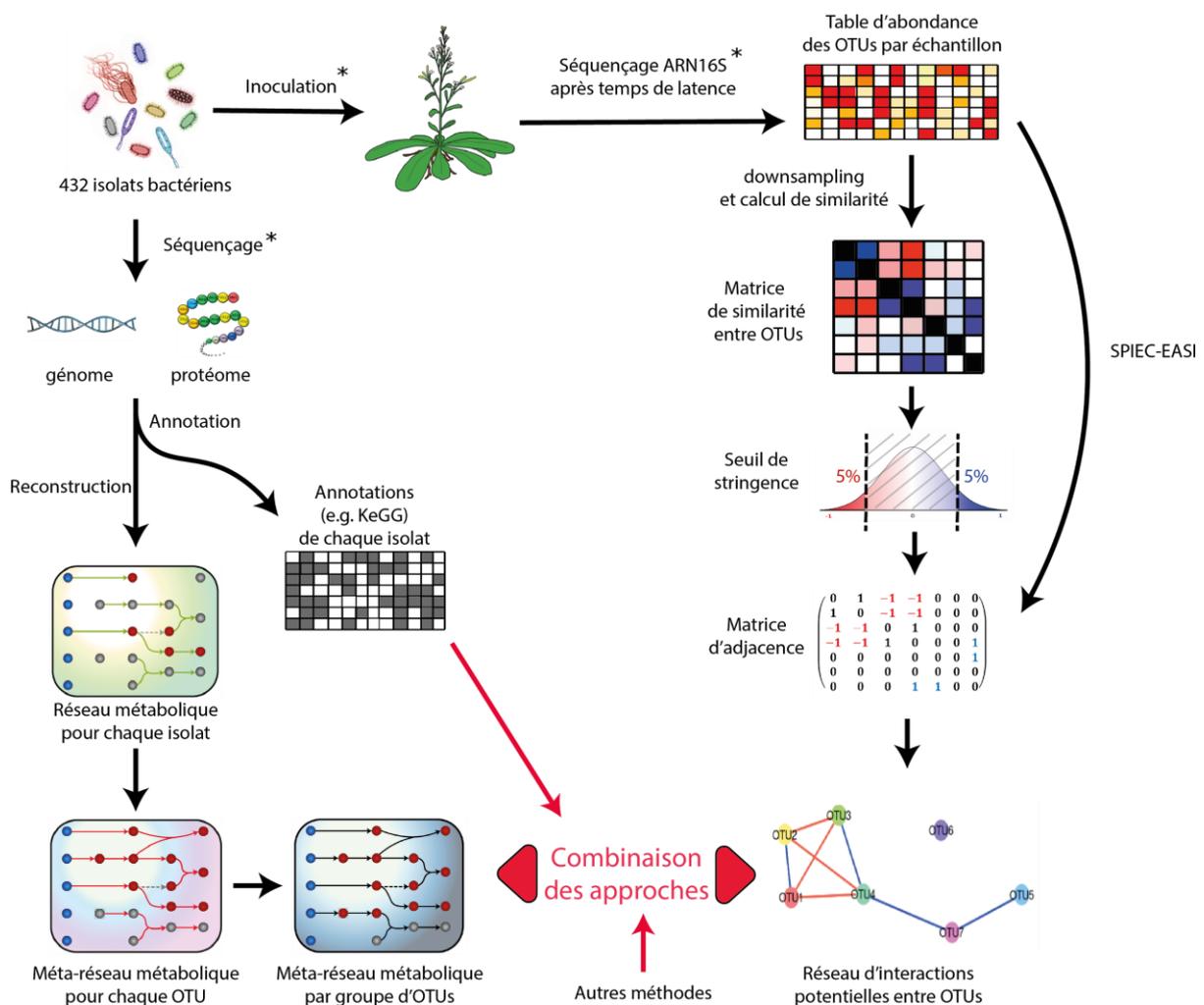


Figure 7: Schéma bilan du pipeline de modélisation mis en place. Les champs marqué d'une (*) correspondent aux étapes réalisées en amont de notre étude.

IV) Discussion

a) Avantage et originalité du jeu de données

Le choix du jeu de données pour effectuer ce travail exploratoire a été fondé sur la possibilité d'avoir accès à différents types d'informations qui sont combinées dans notre étude pour permettre une modélisation multi-réseaux. Par ailleurs le jeu de données 'communautés' qui contient des informations produites à partir de plans d'expériences pour l'analyse du microbiote de plantes soumises à différentes contraintes environnementales intègre également des réplicats expérimentaux, une information précieuse sur la reproductibilité des comportements des phénomènes biologiques. Ainsi, ce jeu de données permet d'extraire les interactions significatives en enlevant les bruits liés à l'expérimentation individuelle, mais ne permet donc pas d'évaluer l'impact des conditions expérimentales (Bonneau, 2008).

Un autre atout majeur de ce jeu de données est que les isolats bactériens cultivés représentent 66% des bactéries qui colonisent spontanément en milieu naturel les racines d'*A. thaliana* (Hacquart, communication personnelle). Comme la majorité des bactéries constituant le microbiote végétal d'*A. thaliana* sont cultivés, des approches gnotobiotiques peuvent être envisagées. L'idée sous-jacente est donc aussi la possibilité offerte par ce modèle d'étude pour tester expérimentalement les hypothèses générées par modélisation (innoculations contrôlée d'*Arabidopsis thaliana*).

La détermination de la composition des communautés microbiennes par l'analyse d'amplicons d'ARNr 16S est la méthode de référence. Cette méthode ne permet pas de distinguer la variabilité à l'intérieur des OTUs (e.g. Vandenkoornhuyse et al., 2010). Pour réduire le nombre d'isolats regroupés sous un OTU, il est possible d'utiliser l'*oligotyping* sur les séquences amplifiées (Eren et al., 2013). L'*oligotyping* est une méthode qui évalue le signal phylogénétique par minimum d'entropie afin de minimiser les bruits (Eren et al., 2013). Ceci pourrait permettre d'obtenir une table d'abondance en isolats, donc d'une meilleure précision lors de l'analyse. On considère l'OTU comme *proxy*. Nous ne pouvons considérer les isolats dans les OTUs puisque dans ce cas le calcul de toutes les combinaisons devient impossible. A titre d'exemple, pour l'échantillon "Root.RSclayVR1b", dont la richesse en OTUs est de quarante, il faudrait étudier les $3,391262 \cdot 10^{41}$ possibles considérant les isolats. Notons qu'à l'heure actuelle, les réseaux d'abondance au sujet du microbiome sont analysés majoritairement sur des données issues de séquençage d'un fragment du gène de l'ARNr 16S. C'est le cas de l'outil Faprotax (Comeau et al., 2017). La finesse de l'enrichissement fonctionnel sur les *hubs* excède donc encore rarement l'échelle de l'OTU.

b) Rétrospective sur la démarche et les résultats

i) Les possibles raisons d'une richesse fonctionnelle plus faible (Figure 2)

1) La nature de l'annotation, un élément crucial.

Les annotations KeGG sont restreintes à un cadre très vérifié du métabolisme cellulaire (ne prennent pas en compte les éléments relatifs à l'immunité de l'hôte) et ne sont pas exhaustives. La base de données COG (*Cluster of Orthologous Groups of proteins*) est basée sur une classification phylogénétique de l'ensemble des protéines décrites de 21 génomes intégralement séquencés. Cette base de données plus ancienne, a l'avantage de distinguer multiples catégories fonctionnelles comme les protéines associées au métabolisme certes, mais aussi à la mobilité, l'immunité ou encore le cytosquelette (Tatusov *et al.*, 2000).

L'importance des gènes codant pour d'autres fonctions biologiques (*e.g.* chimiotaxie), dans le contexte du microbiome végétal, a déjà été montrée dans de nombreuses études (Hardoim *et al.*, 2008). Initié en 2006, le consortium PAMGO (*Plant-Associated Microbe Gene Ontology*) met en place des termes descriptifs (au nombre de 700 en 2009) pour décrire les fonctions associées à un gène bactérien donné, dans le contexte de sa relation avec la plante. Initialement orienté sur la pathogénie, il s'est rapidement étendu aux attributs non-pathogènes (Torto-Alalibo *et al.*, 2009). En effet, il existe une grande complexité dans la relation entre ces bactéries et la plante. La frontière entre parasitisme et mutualisme est souvent ténue, et la pathogénie peut ne s'exprimer que dans certaines conditions, à un certain stade de développement, *etc* (Monteiro *et al.*, 2012). Les termes PAMGO restent toutefois assez généraux et basés sur uniquement sur trois génomes bactériens et un génome fongique (Torto-Alalibo *et al.*, 2009). Un complément à notre analyse pourrait donc consister à réitérer l'expérience avec ce type d'annotations non restreintes au métabolisme cellulaire. Néanmoins, cela exclut toujours les gènes qui n'ont pas encore été annotés fonctionnellement, car nos organismes ne sont pas des organismes modèles largement étudiés (Ungerer *et al.*, 2008). Une grande fraction des gènes des organismes non-modèles se retrouve ainsi non-annotés (Bonneau, 2008). Les réseaux basés sur les similarités de séquence permettent d'identifier les super-familles auxquelles les protéines appartiennent le plus probablement (Atkinson *et al.*, 2009). Bien que difficilement interprétable biologiquement, cela serait plus approprié pour étudier la maximisation en terme de nombre de fonctions au sein d'une communauté dont on possède les séquences en acides aminés (protéomique). Dans tous les cas, l'inférence de fonctions basée sur de la similarité de séquence néglige de nombreux facteurs (*e.g.* structure tertiaire et quaternaire des protéines) qui peuvent influencer considérablement sur le rôle des protéines impliquées. Compléter l'étude avec des données métabolomiques n'est sans doute pas pertinent. En effet de nombreux métabolites 'carrefour' sont issues de réactions différentes, ce qui complique considérablement la liaison de ces données à l'inférence de fonctions sur un gène ou un groupe de gènes donnés (Zhou, 2004).

2) La plante, un censeur potentiel.

Il a été montré dans plusieurs cas que certains micro-organismes endophytes peuvent se substituer à d'autres (Crisp *et al.*, 2015). Pour Doolittle et Booth (2016), ce n'est pas tant l'identité du micro-organismes qui importe mais la nature des interactions qu'il entretient avec la plante. Ils illustrent métaphoriquement leur hypothèse par “what matter is the song, not the singer” (de l'anglais, “c'est la chanson qui importe, ce n'est pas le chanteur”). Néanmoins, on peut s'interroger sur le filtrage par le système immunitaire de l'hôte qui autorise un micro-organisme pas une fonction. En effet, dans notre cas les communautés observées ne maximisent pas en nombre les annotations KeGG (Figure 2). Plusieurs hypothèses permettent d'expliquer ce résultat. Tout d'abord, les bactéries étudiées sont à l'origine isolées de plants d'*A. thaliana*, et en étaient les plus abondantes. Elles sont peut-être déjà suffisamment “optimisées” en ce sens. De plus, supposer qu'il y ait une maximisation métabolique suppose que la plante n'agit pas en facteur limitant de ce point de vue. Or il est tout-à-fait envisageable que la plante, qui agit comme un facteur déterminant des communautés endosphériques, conditionne les communautés de telles sortes qu'elles ne maximisent pas en nombre leurs fonctions métaboliques. Par ailleurs, la maximisation du nombre de fonctions dans le microbiome suppose que toutes ces fonctions sont utiles^{et/ou} nécessaires à la plante. L'interprétation de nos résultats ne soutient pas cette idée. Ceci révèle peut-être aussi l'incapacité à inférer des communautés microbiennes sans prendre en compte l'hôte. Nos résultats supposent que le “chanteur” est aussi important que la “chanson” qu'il interprète, en opposition à l'assertion de Doolittle et Booth (2016).

3) La spécialisation de niche reste à explorer

Il est à noter que si l'on considère les fonctions comme éléments de définition d'une niche pour les bactéries, sur le principe de l'hypothèse de Doolittle et Booth (2016), un biais supplémentaire peut être la différence entre la niche réalisée (fonctions effectivement présentes) et la niche potentielle (fonctions non représentée dans notre communauté observée). Plusieurs niches peuvent ainsi exister. La nomenclature EC (*Enzyme Commission number*), qui a de nombreuses correspondances avec KeGG, possède différents niveaux de description des enzymes. La richesse en enzymes appartenant à la même catégorie supérieure peut donner un aperçu de la spécialisation de niche (Rubino *et al.*, 2017). La spécialisation de niche peut être la cause de cette relativement faible quantité de fonctions enzymatiques.

ii) Combinaison des approches en réseaux dans le contexte de l'émergence d'une 'systems ecology'

1) Un choix impératif parmi les multiples réseaux d'interactions potentielles entre OTUs.

Les mesures de (dis)similarité généralement utilisées dans la modélisation de réseaux (Tableau 1) sont sensibles au nombre et au type d'échantillons. En outre, il y a une perte d'information substantielle sur les mesures de similarité (et de dépendance) qui ne prennent en compte que la matrice binaire de présence/absence des OTUs dans un échantillon. La non-connaissance des conditions expérimentales pour les échantillons, ne permet pas de prendre en compte les différents facteurs associés qui pourraient éventuellement améliorer ces différentes mesures. La faible congruence des réseaux obtenus par ces mesures montre la nécessité d'utiliser d'autres métriques plus complexes dont l'efficacité a été démontrée telles que SPIEC-EASI ou SparCC (Kurtz *et al.*, 2015 ; Friedman et Alm, 2012).

2) Retours sur les résultats de la combinaison et appréciation de l'efficacité et de l'aspect innovant de la méthode

Les résultats obtenus sur le sous-réseau "jouet", montrent une assez bonne performance conjuguée de SPIEC-EASI et de SparCC. En effet, les OTUs liés positivement dans le réseau d'abondance obtenu se complètent à l'échelle des voies métaboliques. Ceci va dans le sens de notre hypothèse de départ qui suppose que les communautés du microbiote végétal s'assemblent de sorte à optimiser la complémentarité fonctionnelle métabolique. Nous validons donc par ce travail l'hypothèse d'une auto-organisation du microbiome médiée par la complémentarité métabolique entre individus. Cela est vérifié pour la production de phytohormones par nos trois couples d'OTUs. La production d'auxine est favorisée dans les couples corrélés. Le même constat est fait concernant l'acide jasmonique qui est une molécule régulatrice importante dans de nombreux processus développementaux qui incluent la fertilité des plantes, l'élongation des racines et le mûrissement des fruits (Fonseca *et al.*, 2009). L'absence de complétion des voies métaboliques de l'éthylène ainsi que l'absence d'enzyme impliqué dans la production d'acide abscissique ne peut présager l'absence de complétion en analysant un autre hub.. Cette hypothèse sera vérifiée par une étude élargie à l'ensemble du réseau d'OTUs. Il est utile de constater que la synthèse de certains analogues de phytohormones apparaît dans la complétion des méta-réseaux métaboliques de la partie du *hub* analysé ici. Cette observation suggère la possibilité d'une manipulation de l'hôte par ces bactéries. Cette hypothèse devra être testée.

Sous jacent à ces méta-réseaux métaboliques, on considère ici une diffusion ou transmission des intermédiaires métaboliques, qu'il faudra contrôler ultérieurement pour perfectionner le processus de modélisation. Un autre outil permettant la construction de méta-réseaux métabolique et prenant en compte cet aspect, est *MetaPathways*. Il serait intéressant d'en comparer les résultats (Konwar *et al.*, 2013 ; Hanson *et al.*, 2014).

Un phénomène d'évolution réductive est fortement suspecté au sein des isolats de l'OTU19. Des travaux de comparaison de génomes en cours visent à mettre en évidence les traces d'évolution réductive. L'évolution réductive peut également intervenir entre individus d'espèces différentes en compétition pour la même niche écologique. D'une certaine manière les phénomènes de complétion pourraient être une conséquence d'une trajectoire évolutive d'évolution réductive. Cette hypothèse nécessite d'être approfondie. L'hypothèse de la Reine Noire comme élément expliquant la complexité du microbiome végétal ne peut pas être rejetée et cette question semble fondamentale dans la compréhension de la complexité du microbiote végétal.

La combinaison de la modélisation en méta-réseaux métaboliques et de la modélisation en réseau d'interactions potentielles entre OTUs basé sur les données d'abondance est une démarche novatrice. Elle s'appuie sur les mêmes principes que la biologie de systèmes mais appliqué à un système plus complexe qui intègre la dimension relative aux interactions bactériennes au sein du microbiome. Ce travail initie ce que pourrait devenir la '*systems ecology*' et nourrir de nouveaux concepts associés à cette '*systems ecology*' qui est promis à un développement important par la possibilité d'approfondissements dans l'interprétation des données, pour mieux comprendre les microbiotes, émettre des hypothèses de fonctionnement et d'interactions sur la base de ces modélisations de réseaux (avec l'hôte y compris), par une nouvelle voie d'analyse des metagenomes et des assemblages *de novo* qui en sont produits.

iii) Un pipeline modulable, orientable vers une démarche d'optimisation théorique du microbiome

Une autre façon d'envisager les réseaux est d'utiliser des règles de raisonnement logiques (exemple : « A est présent si et seulement si B et C présents et D absent »), grâce à des langages adaptés à la modélisation et la résolution de problèmes de recherche combinatoire, tels que la programmation par ensembles réponses (*Answer Set Programming, ASP*). Différents algorithmes ont été proposés afin de construire ces réseaux depuis de larges jeux de données (Agrawal *et al.*, 1993). De plus, cette approche qui résout les problèmes d'ordre combinatoire permet d'aborder les interactions d'ordre supérieur à deux (*e.g.* l'étude de la présence de trois OTUs simultanément et de leurs interactions avec d'autres groupes potentiels d'OTUs pour donner un métagraphe). On pourrait imaginer la combinaison de notre pipeline topologique à de la FVA (*Flux Value Analysis*) et de la FBA (*Flux Balance Analysis*) qui prennent en compte la stoechiométrie des réactions et à un programme ASP cherchant à maximiser différentes fonctions telles que la production de biomasse. Ceci permettrait éventuellement d'identifier les communautés idéales potentielles, de les inoculer ainsi que des communautés non-optimales afin de comparer leurs performances. Une approche de modélisation mathématique développée par Taillefumier *et al.* (2017) basée sur les flux métaboliques impliqués dans la fabrication de biomasse permettrait d'expliquer la coexistence de différents métabolismes. Ceci correspond à

l'hypothèse d'un assemblage des communautés microbiennes gouverné par l'optimisation du rendement de biomasse. Les processus de compétition restreindraient les ressources de sorte qu'aucune autre stratégie puisse envahir, et que l'état optimal serait stable. Ceci est une alternative au principe d'exclusion compétitive pour lequel on considère qu'à un état stable, la diversité ne peut excéder le nombre de ressources disponibles (Hardin, 1960 ; Levin, 1970). L'utilisation de notre pipeline alors implémenté pourrait permettre de vérifier cette hypothèse dans le cas du microbiome végétal.

c) Trois perspectives envisagées

i) Intégrer la dimension spatiale

L'endosphère végétale est un milieu très structuré ne serait-ce que par l'organisation tissulaire végétale. Cette structure est importante dans le processus de colonisation des bactéries dans l'endosphère (Hardoim *et al.*, 2008). Des études (Lundberg *et al.*, 2012 & Bulgarelli *et al.*, 2012) sur *Arabidopsis thaliana* qui ont utilisé les ultrasons (sonication) pour ôter les couches superficielles des tissus végétaux, ont montré que les bactéries endophytes résident principalement dans l'apoplaste intercellulaire, les cellules mortes et mourantes, mais aussi dans les vaisseaux de xylèmes par lesquels elles peuvent migrer (Lundberg *et al.*, 2012). Ceci suggère une possible spatialisation au sein du végétal dans nos échantillons. L'interprétation des réseaux pourrait permettre de prédire quels micro-organismes partagent le même habitat si des plans d'expérience adaptés sont mis en oeuvre. Dans le cas contraire d'une analyse globale des communautés tel que généralement effectué, on ne peut exclure des liens de co-occurrences ayant peu de signification biologique.

De façon plus générale, il serait intéressant de comparer le microbiome de la phyllosphère au microbiome endophyte racinaire dont les annotations KeGG sont très différentes et qui sont souvent anti-corrélées selon SparCC dans les données d'abondance. Cela est probablement dû à une spécialisation du microbiome phyllosphérique.

La structure spatiale est possiblement importante à l'échelle du tissu, mais aussi si l'on généralise à l'échelle du micro-paysage racinaire. La combinaison de notre *pipeline* de modélisation (Figure 7) avec des systèmes d'information géographique (SIG) permettrait éventuellement de mieux comprendre les interactions microbiennes.

ii) Intégrer la dimension temporelle

Un second point limitant de cette étude est l'aspect statique de l'analyse. En effet, la nature du séquençage pour obtenir des matrices d'abondance est destructif. La dimension de la dynamique temporelle des interactions bactériennes (microbiote/hôte) est pourtant très important pour comprendre la stabilité du système (Bonneau, 2008). Notre étude menée sur des données produites au bout de quatre semaines qui ne sont probablement pas dans un état d'équilibre

stable, si l'on considère les dynamiques des microbiotes. Néanmoins dans notre cas de figure, nous mélangeons des informations de plusieurs expérimentations menées de manière indépendante et observons un signal de cooccurrence qui peut donc être qualifié de robuste.

L'emploi des équations généralisées de Lotka-Volterra s'est révélé pertinent dans la prédiction de la prépondérance de *Clostridium difficile* dans le microbiome intestinal humain sous l'effet d'antibiotiques, en détruisant les fèces naturellement renouvelées (Stein *et al.*, 2013). Le couplage avec des études de réseaux métaboliques et de flux a aussi donné des résultats encourageants (Bucci & Xavier, 2014).

iii) Moduler les conditions environnementales:

Une dernière perspective envisagée serait l'application d'un stress (biotique ou abiotique) sur la plante et d'en mesurer la perturbation à l'échelle de ses communautés microbiennes mais il serait alors difficile d'évaluer l'impact du stress sur le microbiote lui-même. On pourrait aussi mesurer l'influence bactérienne (seule ou en mélange) sur les performances de la plante comparativement à une plante aseptique, et dans différentes conditions. Une mesure de similarité supplémentaire sur les échantillons pourrait alors être envisagée: UniFrac (Lozupone et Knight, 2005). qui prend en compte la proximité phylogénétique entre les OTUs. Bien que cela n'ait pas été envisagé lors du développement de nos réseaux d'interactions potentielles entre OTUs (basés sur les données d'abondance), on peut, dans ce contexte, visualiser l'influence des conditions expérimentales via des réseaux de classification double, ou "biclusters" (Bonneau, 2008).

Le développement rapide des technologies "single-cell" déjà qualifiées de "révélateur de la matière noire du microbiome" par Xu *et al.* (2017), permettra probablement de mieux prendre en compte cette dimension."

V) Conclusion

L'objectif des études en génomique écologique est d'identifier les gènes et les voies génétiques sous-jacents aux interactions et aux réponses écologiques importantes. C'est aussi de déterminer dans quelle mesure ces gènes et ces voies sont variables, de caractériser ces variations du point de vue écologique mais aussi les conséquences évolutives associées (Ungerer *et al.*, 2008).

L'hypothèse selon laquelle les communautés microbiennes endophytes s'assemblent préférentiellement par "auto-organisation" de sorte à optimiser la complémentarité fonctionnelle, peut expliquer la forte richesse spécifique observée du microbiome végétal. L'approche de modélisation en réseaux d'OTUs a permis de révéler les interactions-clés potentielles au sein d'un système complexe tel que le microbiome végétal. La modélisation en réseau métabolique, permise par une démarche gnotobiotique qui donne accès aux données omiques du microbiote,

peut dans une certaine limite expliquer les relations de corrélations détectées dans les réseaux d'OTUs basés sur les données d'abondances. Pour le *hub* choisi dans ce travail, on peut observer une complémentarité métabolique à l'échelle de certaines voies métaboliques. D'après cette preuve de concept, il semble possible d'envisager la complexité des interactions au sein d'un écosystème complexe et d'en déduire des propriétés émergentes.

Il est possible grâce à notre approche de définir la partie conservée du microbiote, « *core microbiota* », au sein d'un groupe d'individus hôtes défini en fonction de l'échelle écologique étudiée. Cette fraction du microbiome serait la plus essentielle pour l'hôte (que l'on oppose à l'«éco-microbiome» considéré comme accessoire). Elle est bien entendu dépendante de l'échelle hiérarchique écologique et de la précision taxonomique choisies concernant les micro-organismes constituant le microbiote (Vandenkoornhuyse *et al.*, 2015). Il serait intéressant de voir si l'on peut expliquer l'appartenance des bactéries à cette fraction conservée du microbiote grâce aux réseaux métaboliques, afin de mieux comprendre les fonctions de chaque organisme. De plus, notre preuve de concept s'est limitée pour l'instant à l'étude des bactéries, pour des raisons pratiques d'accès aux données. Il ne faut cependant pas négliger le grand nombre de champignons et d'archeae (plus complexe biologiquement parlant) qui constituent, eux aussi, le microbiome végétal. Notre étude pourra être facilement adaptée à l'étude de ces micro-organismes.

La compréhension de l'organisation et du rôle du microbiote végétal est un point crucial en agronomie et est un levier d'action potentiel pour relever les grands défis du XXI^{ème} siècle (changement climatique, maintien des écosystèmes, *etc*) (Jousset *et al.*, 2017). Une réorientation des pratiques culturales vers une intégration du microbiome dans les logiques de sélection et d'itinéraires techniques au champs dans le contexte de la "Révolution Verte" semble prometteur (Duhamel & Vandenkoornhuyse, 2013). Le projet *Phytobiomes Alliance* qui est mis en place en 2017 à l'échelle internationale, envisage d'accentuer l'éclairage du phytobiome (micro-organismes associés aux plantes) dans une trajectoire d'amélioration de la santé des plantes et des agro-écosystèmes et leur productivité. On peut imaginer que la transition des connaissances du microbiome végétal puisse se traduire dans des nouvelles formes d'agriculture ou de sélection variétale pour une production agricole plus douce et plus durable (Duhamel & Vandenkoornhuyse, 2013).

Bibliographie:

AGRAWAL, Rakesh, IMIELIŃSKI, Tomasz et SWAMI, Arun, 1993. Mining association rules between sets of items in large databases. S.l. : ACM Press. 1993. p. 207- 216.

AITCHISON, J., 1981. A new approach to null correlations of proportions. In : *Journal of the International Association for Mathematical Geology*. avril 1981. Vol. 13, n° 2, p. 175- 189. DOI 10.1007/BF01031393.

AITE, Meziane, CHEVALLIER, Marie, FRIOUX, Clémence, TROTTIER, Camille, GOT, Jeanne, PAZ CORTES, Maria, LOIRA, Nicolas, CARRIER, Grégory, DAMERON, Olivier, GUILLAUDEUX, Nicolas, LATORRE, Mauricio, MENDOZA, Sebastián, MARKOV, Gabriel V., MAAS, Alejandro et SIEGEL, Anne, in prep. Traceability, reproducibility and wiki-exploration of genome-scale metabolic model reconstructions. In : . in prep.

ALTMAN, Tomer, TRAVERS, Michael, KOTHARI, Anamika, CASPI, Ron et KARP, Peter D, 2013. A systematic comparison of the MetaCyc and KEGG pathway databases. In : *BMC Bioinformatics*. 2013. Vol. 14, n° 1, p. 112. DOI 10.1186/1471-2105-14-112.

ATKINSON, Holly J., MORRIS, John H., FERRIN, Thomas E. et BABBITT, Patricia C., 2009. Using Sequence Similarity Networks for Visualization of Relationships Across Diverse Protein Superfamilies. In : *PLoS ONE*. 3 février 2009. Vol. 4, n° 2. DOI 10.1371/journal.pone.0004345.

BAI, Yang, MÜLLER, Daniel B., SRINIVAS, Girish, GARRIDO-OTER, Ruben, POTTHOFF, Eva, ROTT, Matthias, DOMBROWSKI, Nina, MÜNCH, Philipp C., SPAEPEN, Stijn, REMUS-EMSERMANN, Mitja, HÜTTEL, Bruno, MCHARDY, Alice C., VORHOLT, Julia A. et SCHULZE-LEFERT, Paul, 2015. Functional overlap of the Arabidopsis leaf and root microbiota. In : *Nature*. 17 décembre 2015. Vol. 528, n° 7582, p. 364- 369. DOI 10.1038/nature16192.

BERENDSEN, Roeland L., PIETERSE, Corné M.J. et BAKKER, Peter A.H.M., 2012. The rhizosphere microbiome and plant health. In : *Trends in Plant Science*. août 2012. Vol. 17, n° 8, p. 478- 486. DOI 10.1016/j.tplants.2012.04.001.

BONNEAU, Richard, 2008. Learning biological networks: from modules to dynamics. In : *Nature Chemical Biology*. novembre 2008. Vol. 4, n° 11, p. 658- 664. DOI 10.1038/nchembio.122.

BRESSAN, Mélanie, RONCATO, Marie-Anne, BELLVERT, Floriant, COMTE, Gilles, HAICHAR, Feth el Zahar, ACHOUAK, Wafa et BERGE, Odile, 2009. Exogenous glucosinolate produced by Arabidopsis thaliana has an impact on microbes in the rhizosphere and plant roots. In : *The ISME Journal*. novembre 2009. Vol. 3, n° 11, p. 1243- 1257. DOI 10.1038/ismej.2009.68.

BUCCI, Vanni et XAVIER, Joao B., 2014. Towards Predictive Models of the Human Gut Microbiome. In : *Journal of Molecular Biology*. novembre 2014. Vol. 426, n° 23, p. 3907- 3916. DOI 10.1016/j.jmb.2014.03.017.

BULGARELLI, Davide, ROTT, Matthias, SCHLAEPPI, Klaus, VER LOREN VAN THEMAAT, Emiel, AHMADINEJAD, Nahal, ASSENZA, Federica, RAUF, Philipp, HUETTEL, Bruno, REINHARDT, Richard, SCHMELZER, Elmon, PEPLIES, Joerg, GLOECKNER, Frank Oliver, AMANN, Rudolf, EICKHORST, Thilo et SCHULZE-LEFERT,

Paul, 2012. Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. In : *Nature*. 1 août 2012. Vol. 488, n° 7409, p. 91- 95. DOI 10.1038/nature11336.

CASPI, Ron, ALTMAN, Tomer, BILLINGTON, Richard, DREHER, Kate, FOERSTER, Hartmut, FULCHER, Carol A., HOLLAND, Timothy A., KESELER, Ingrid M., KOTHARI, Anamika, KUBO, Aya, KRUMMENACKER, Markus, LATENDRESSE, Mario, MUELLER, Lukas A., ONG, Quang, PALEY, Suzanne, SUBHRAVETI, Pallavi, WEAVER, Daniel S., WEERASINGHE, Deepika, ZHANG, Peifen et KARP, Peter D., 2014. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. In : *Nucleic Acids Research*. janvier 2014. Vol. 42, n° D1, p. D459- D471. DOI 10.1093/nar/gkt1103.

COMEAU, Andr? M., DOUGLAS, Gavin M. et LANGILLE, Morgan G. I., 2017. Microbiome Helper: a Custom and Streamlined Workflow for Microbiome Research. In : EISEN, Jonathan (éd.), *mSystems*. 28 février 2017. Vol. 2, n° 1, p. e00127-16. DOI 10.1128/mSystems.00127-16.

COMPANT, Stéphane, CLÉMENT, Christophe et SESSITSCH, Angela, 2010. Plant growth-promoting bacteria in the rhizo- and endosphere of plants: Their role, colonization, mechanisms involved and prospects for utilization. In : *Soil Biology and Biochemistry*. mai 2010. Vol. 42, n° 5, p. 669- 678. DOI 10.1016/j.soilbio.2009.11.024.

CRISP, Alastair, BOSCHETTI, Chiara, PERRY, Malcolm, TUNNACLIFFE, Alan et MICKLEM, Gos, 2015. Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. In : *Genome Biology*. 2015. Vol. 16, n° 1, p. 50. DOI 10.1186/s13059-015-0607-3.

DEMPSTER, A. P., 1972. Covariance Selection. In : *Biometrics*. mars 1972. Vol. 28, n° 1, p. 157. DOI 10.2307/2528966.

DIAMOND, Jared M. (éd.), 1975. In : DIAMOND, Jared M. (éd.), *Ecology and evolution of communities*. Cambridge, Mass : Belknap Press of Harvard University Press. p. 342- 444. ISBN 978-0-674-22444-5. QH540 .E3115

DOOLITTLE, W. Ford et BOOTH, Austin, 2016. It's the song, not the singer: an exploration of holobiosis and evolutionary theory. In : *Biology & Philosophy*. octobre 2016. Vol. 32, n° 1, p. 5- 24. DOI 10.1007/s10539-016-9542-2.

DRAY, Stéphane, DUFOUR, Anne-Béatrice et OTHERS, 2007. The ade4 package: implementing the duality diagram for ecologists. In : *Journal of statistical software*. 2007. Vol. 22, n° 4, p. 1-20.

DUHAMEL, Marie et VANDENKOORNHUYSE, Philippe, 2013. Sustainable agriculture: possible trajectories from mutualistic symbiosis and plant neodomestication. In : *Trends in Plant Science*. novembre 2013. Vol. 18, n° 11, p. 597- 600. DOI 10.1016/j.tplants.2013.08.010.

EREN, A. Murat, MAIGNIEN, Loïs, SUL, Woo Jun, MURPHY, Leslie G., GRIM, Sharon L., MORRISON, Hilary G. et SOGIN, Mitchell L., 2013. Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. In : *Methods in Ecology and Evolution*. 1 décembre 2013. Vol. 4, n° 12, p. 1111- 1119. DOI 10.1111/2041-210X.12114.

FONSECA, Sandra, CHICO, Jose M. et SOLANO, Roberto, 2009. The jasmonate pathway: the ligand, the receptor and the core signalling module. In : *Current Opinion in Plant Biology*. octobre 2009. Vol. 12, n° 5, p. 539- 547. DOI 10.1016/j.pbi.2009.07.013.

FRIEDMAN, Jonathan et ALM, Eric J., 2012. Inferring Correlation Networks from Genomic Survey Data. In : VON MERING, Christian (éd.), *PLoS Computational Biology*. 20 septembre 2012. Vol. 8, n° 9, p. e1002687. DOI 10.1371/journal.pcbi.1002687.

GHOSH, Pallab Kumar, KUMAR DE, Tarun et MAITI, Tushar Kanti, 2015. Production and Metabolism of Indole Acetic Acid in Root Nodules and Symbiont (*Rhizobium undicola*) Isolated from Root Nodule of Aquatic Medicinal Legume *Neptunia oleracea* Lour. In : *Journal of Botany*. 2015. Vol. 2015, p. 1- 11. DOI 10.1155/2015/575067.

GLICK, Bernard R., 2005. Modulation of plant ethylene levels by the bacterial enzyme ACC deaminase. In : *FEMS Microbiology Letters*. octobre 2005. Vol. 251, n° 1, p. 1- 7. DOI 10.1016/j.femsle.2005.07.030.

GUTIERREZ-MANERO, Francisco Javier, RAMOS-SOLANO, Beatriz, PROBANZA, Agustín, MEHOUACHI, Jalel, R. TADEO, Francisco et TALON, Manuel, 2001. The plant-growth-promoting rhizobacteria *Bacillus pumilus* and *Bacillus licheniformis* produce high amounts of physiologically active gibberellins. In : *Physiologia Plantarum*. février 2001. Vol. 111, n° 2, p. 206- 211. DOI 10.1034/j.1399-3054.2001.1110211.x.

HANSON, Niels W., KONWAR, Kishori M., WU, Shang-Ju et HALLAM, Steven J., 2014. MetaPathways v2.0: A master-worker model for environmental Pathway/Genome Database construction on grids and clouds. S.l. : IEEE. mai 2014. p. 1- 7.

HARDIN, Garrett, 1960. The Competitive Exclusion Principle. In : *Science*. 29 avril 1960. Vol. 131, n° 3409, p. 1292- 1297. DOI 10.1126/science.131.3409.1292.

HARDOIM, Pablo R., VAN OVERBEEK, Leo S. et ELSAS, Jan Dirk van, 2008. Properties of bacterial endophytes and their proposed role in plant growth. In : *Trends in Microbiology*. octobre 2008. Vol. 16, n° 10, p. 463- 471. DOI 10.1016/j.tim.2008.07.008.

HUBBELL, Stephen P, 2001. *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)*. Princeton : Princeton University Press. ISBN 978-1-4008-3752-6.

JOUSSET, Alexandre, BIENHOLD, Christina, CHATZINOTAS, Antonis, GALLIEN, Laure, GOBET, Angélique, KURM, Viola, KÜSEL, Kirsten, RILLIG, Matthias C., RIVETT, Damian W., SALLES, Joana F., VAN DER HEIJDEN, Marcel G. A., YOUSSEF, Noha H., ZHANG, Xiaowei, WEI, Zhong et HOL, W. H. Gera, 2017. Where less may be more: how the rare biosphere pulls ecosystems strings. In : *The ISME Journal*. avril 2017. Vol. 11, n° 4, p. 853- 862. DOI 10.1038/ismej.2016.174.

KANEHISA, M. et GOTO, Susumu, 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. In : *Nucleic Acids Research*. 1 janvier 2000. Vol. 28, n° 1, p. 27- 30. DOI 10.1093/nar/28.1.27.

KING, Zachary A., LU, Justin, DRÄGER, Andreas, MILLER, Philip, FEDEROWICZ, Stephen, LERMAN, Joshua A., EBRAHIM, Ali, PALSSON, Bernhard O. et LEWIS, Nathan E., 2016. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. In : *Nucleic Acids Research*. 4 janvier 2016. Vol. 44, n° D1, p. D515- D522. DOI 10.1093/nar/gkv1049.

KONWAR, Kishori M, HANSON, Niels W, PAGÉ, Antoine P et HALLAM, Steven J, 2013. MetaPathways: a modular pipeline for constructing pathway/genome databases from environmental sequence information. In : *BMC Bioinformatics*. 2013. Vol. 14, n° 1, p. 202. DOI 10.1186/1471-2105-14-202.

KURTZ, Zachary D., MÜLLER, Christian L., MIRALDI, Emily R., LITTMAN, Dan R., BLASER, Martin J. et BONNEAU, Richard A., 2015. Sparse and Compositionally Robust Inference of Microbial Ecological Networks. In : *PLOS Computational Biology*. 7 mai 2015. Vol. 11, n° 5, p. e1004226. DOI 10.1371/journal.pcbi.1004226.

LEVIN, Simon A., 1970. Community Equilibria and Stability, and an Extension of the Competitive Exclusion Principle. In : *The American Naturalist*. 1970. Vol. 104, n° 939, p. 413- 423.

LOZUPONE, Catherine et KNIGHT, Rob, 2005. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. In : *Applied and Environmental Microbiology*. 12 janvier 2005. Vol. 71, n° 12, p. 8228- 8235. DOI 10.1128/AEM.71.12.8228-8235.2005.

LUNDBERG, Derek S., LEBEIS, Sarah L., PAREDES, Sur Herrera, YOURSTONE, Scott, GEHRING, Jase, MALFATTI, Stephanie, TREMBLAY, Julien, ENGELBREKTSON, Anna, KUNIN, Victor, RIO, Tijana Glavina del, EDGAR, Robert C., EICKHORST, Thilo, LEY, Ruth E., HUGENHOLTZ, Philip, TRINGE, Susannah Green et DANGL, Jeffery L., 2012. Defining the core Arabidopsis thaliana root microbiome. In : *Nature*. 1 août 2012. Vol. 488, n° 7409, p. 86- 90. DOI 10.1038/nature11237.

MACKEY, David et MCFALL, Aidan J., 2006. MAMPs and MIMPs: proposed classifications for inducers of innate immunity. In : *Molecular Microbiology*. septembre 2006. Vol. 61, n° 6, p. 1365- 1371. DOI 10.1111/j.1365-2958.2006.05311.x.

MAS, Alix, JAMSHIDI, Shahrad, LAGADEC, Yvan, EVEILLARD, Damien et VANDENKOOORNHUYSE, Philippe, 2016. Beyond the Black Queen Hypothesis. In : *The ISME Journal*. septembre 2016. Vol. 10, n° 9, p. 2085- 2091. DOI 10.1038/ismej.2016.22.

MEINSHAUSEN, Nicolai et BÜHLMANN, Peter, 2006. High-dimensional graphs and variable selection with the Lasso. In : *The Annals of Statistics*. juin 2006. Vol. 34, n° 3, p. 1436- 1462. DOI 10.1214/009053606000000281.

MONTEIRO, Rose Adele, BALSANELLI, Eduardo, WASSEM, Roseli, MARIN, Anelis M., BRUSAMARELLO-SANTOS, Liziane C. C., SCHMIDT, Maria Augusta, TADRA-SFEIR, Michelle Z., PANKIEVICZ, Vânia C. S., CRUZ, Leonardo M., CHUBATSU, Leda S., PEDROSA, Fabio O. et SOUZA, Emanuel M., 2012. Herbaspirillum-plant interactions: microscopical, histological and molecular aspects. In : *Plant and Soil*. juillet 2012. Vol. 356, n° 1- 2, p. 175- 196. DOI 10.1007/s11104-012-1125-7.

MORRIS, J. J., LENSKI, R. E. et ZINSER, E. R., 2012. The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss. In : *mBio*. 2 mai 2012. Vol. 3, n° 2, p. e00036-12-e00036-12. DOI 10.1128/mBio.00036-12.

OKSANEN, Jari, KINDT, Roeland, LEGENDRE, Pierre, O'HARA, Bob, STEVENS, M. Henry H., OKSANEN, Maintainer Jari et SUGGESTS, MASS, 2007. The vegan package. In : *Community ecology package*. 2007. Vol. 10, p. 631-637.

PEARSON, Karl, 1896. Mathematical Contributions to the Theory of Evolution.--On a Form of Spurious Correlation Which May Arise When Indices Are Used in the Measurement of Organs. In : *Proceedings of the Royal Society of London (1854-1905)*. 1 janvier 1896. Vol. 60, n° 1, p. 489- 498. DOI 10.1098/rspl.1896.0076.

RUBINO, Francesco, CARBERRY, Ciara, M WATERS, Sinéad, KENNY, David, MCCABE, Matthew S et CREEVEY, Christopher J, 2017. Divergent functional isoforms drive niche specialisation for nutrient acquisition and use in rumen microbiome. In : *The ISME Journal*. avril 2017. Vol. 11, n° 4, p. 932- 944. DOI 10.1038/ismej.2016.172.

SOKAL, Robert R. et SNEATH, Peter Henry Andrews, 1963. *Principles of Numerical Taxonomy*. S.l. : W. H. Freeman.

STEIN, Richard R., BUCCI, Vanni, TOUSSAINT, Nora C., BUFFIE, Charlie G., RÄTSCH, Gunnar, PAMER, Eric G., SANDER, Chris et XAVIER, João B., 2013. Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intestinal Microbiota. In : VON MERING, Christian (éd.), *PLoS Computational Biology*. 12 décembre 2013. Vol. 9, n° 12, p. e1003388. DOI 10.1371/journal.pcbi.1003388.

TAILLEFUMIER, Thibaud, POSFAI, Anna, MEIR, Yigal et WINGREEN, Ned S, 2017. Microbial consortia at steady supply. 5 mai 2017. Vol. 6. DOI 10.7554/eLife.22644.

TATUSOV, Roman L., GALPERIN, Michael Y., NATALE, Darren A. et KOONIN, Eugene V., 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. In : *Nucleic Acids Research*. 1 janvier 2000. Vol. 28, n° 1, p. 33- 36.

TIBSHIRANI, Robert, 1996. Regression shrinkage and selection via the lasso. In : *Journal of the Royal Statistical Society. Series B (Methodological)*. 1996. p. 267–288.

TORTO-ALALIBO, Trudy, COLLMER, Candace W. et GWINN-GIGLIO, Michelle, 2009. The Plant-Associated Microbe Gene Ontology (PAMGO) Consortium: community development of new Gene Ontology terms describing biological processes involved in microbe-host interactions. In : *BMC Microbiology*. 2009. Vol. 9, n° 1, p. S1. DOI 10.1186/1471-2180-9-S1-S1.

UNGERER, M. C., JOHNSON, L. C. et HERMAN, M. A., 2008. Ecological genomics: understanding gene and genome function in the natural environment. In : *Heredity*. février 2008. Vol. 100, n° 2, p. 178- 183. DOI 10.1038/sj.hdy.6800992.

VANDENKOORNHUYSE, Philippe, DUFRESNE, Alexis, QUAISER, Achim, GOUESBET, Gwenola, BINET, Françoise, FRANCEZ, André-Jean, MAHÉ, Stéphane, BORMANS, Myriam, LAGADEUC, Yvan et COUÉE, Ivan, 2010. Integration of molecular functions at the ecosystemic level: breakthroughs and future goals of environmental genomics and post-genomics: Environmental genomics. In : *Ecology Letters*. 21 avril 2010. Vol. 13, n° 6, p. 776- 791. DOI 10.1111/j.1461-0248.2010.01464.x.

VANDENKOORNHUYSE, Philippe, QUAISER, Achim, DUHAMEL, Marie, LE VAN, Amandine et DUFRESNE, Alexis, 2015. The importance of the microbiome of the plant holobiont. In : *New Phytologist*. juin 2015. Vol. 206, n° 4, p. 1196- 1206. DOI 10.1111/nph.13312.

WALT, Stéfan van der, COLBERT, S. Chris et VAROQUAUX, Gaël, 2011. The NumPy Array: A Structure for Efficient Numerical Computation. In : *Computing in Science & Engineering*. 1 mars 2011. Vol. 13, n° 2, p. 22- 30. DOI 10.1109/MCSE.2011.37.

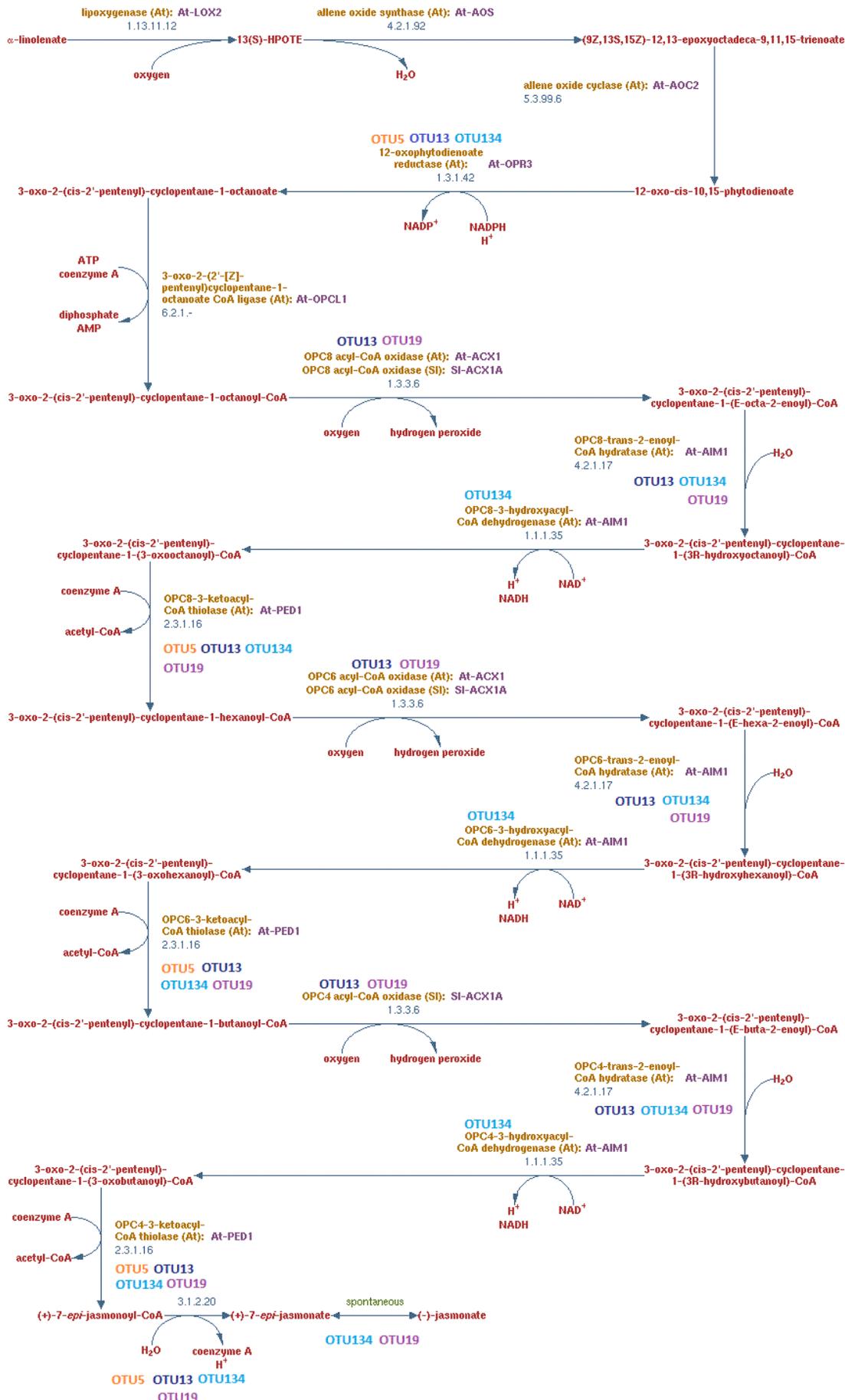
XU, Jian, MA, Bo, SU, Xiaoquan, HUANG, Shi, XU, Xin, ZHOU, Xuedong, HUANG, Wei E. et KNIGHT, Rob, 2017. Emerging Trends for Microbiome Analysis: From Single-Cell Functional Imaging to Microbiome Big Data. In : *Engineering*. février 2017. Vol. 3, n° 1, p. 66- 70. DOI 10.1016/J.ENG.2017.01.020.

ZHOU, Jizhong, 2004. Chapter 1: Genomics: Toward A Genome-Level understanding of the Structure, Functions, and Evolution of Biological Systems. In : *Microbial functional genomics*. Hoboken, N.J. : Wiley-Liss. p. 5- 10. ISBN 978-0-471-07190-7.

ZILBER-ROSENBERG, Ilana et ROSENBERG, Eugene, 2008. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. In : *FEMS Microbiology Reviews*. août 2008. Vol. 32, n° 5, p. 723 - 735. DOI 10.1111/j.1574-6976.2008.00123.x.

Annexe 1:

Voie de biosynthèse des jasmonates (PWY-735) (source : MetaCyc)



	Diplôme : Ingénieur Agronome et Master Spécialité : MODE (Modélisation en Ecologie) Spécialisation / option : NA Enseignant référent : Frédéric Hamelin
Auteur(s) : Victoria Potdevin Date de naissance* : 10 novembre 1994	Organisme d'accueil : ECOBIO UMR6553 Adresse : Avenue du Général Leclerc Campus de Beaulieu F-35042 RENNES Cedex
Nb pages : 30 Annexe(s) : 1	
Année de soutenance : 2017	Maître de stage : Pr. Philippe VANDENKOORNHUYSE
Titre français : Combinaison de modélisations en réseaux pour la réinterprétation des règles d'assemblages du microbiome végétal. Titre anglais : Combination of network modelling methods for the reinterpretation of the assembly rules of the plant microbiome.	
Résumé (1600 caractères maximum) : <p>Le microbiote végétal est un élément déterminant de la productivité et de la robustesse des plantes. Ce microbiote est complexe à comprendre, à étudier et à envisager dans son intégralité. Pour appréhender les règles d'assemblage de ce microbiote, l'hypothèse qu'il s'auto-assemble de façon optimale, complémentaire sur l'échelle métabolique est analysée. A partir de données accumulées sur la plante modèle <i>A. thaliana</i> par une approche gnotobiotique, des travaux de modélisation en réseaux sont réalisés. Les données exploitées correspondent à une table d'abondance en OTUs et des données omiques (génomique et protéomique) de bactéries du microbiote de la plante hôte. On combine à partir de ces données deux types de réseaux : des réseaux d'interactions potentielles entre OTUs et des méta-réseaux métaboliques afin d'étudier l'influence de la complémentarité métabolique sur l'assemblage des communautés microbiotiques. Il n'y a pas de maximisation en terme de nombre de fonctions (annotations fonctionnelles) dans notre jeu de données. Cependant il y a bien, d'après notre sous-réseau jouet d'OTUs, une meilleure complétion métabolique entre les OTUs les plus souvent associés que pour ceux qui ne le sont pas. Ces résultats sont encourageants et valident la preuve de concept de notre hypothèse de départ. Ces nouvelles méthodes en modélisation s'inscrivent dans l'émergence du concept récent de "systems ecology" qui, sur le même principe que la biologie des systèmes, s'intéresse à des systèmes écologiques complexes, notamment de par les interactions entre organismes.</p>	
Abstract (1600 caractères maximum) : <p>Plant microbiota is a key determinant of plant fitness and productivity. This microbiota is complex to understand, to study and to consider in its entirety. In order to apprehend the rules of the plant-microbiota assembly, we analyse the hypothesis that the microbiota is optimally self-assembling (complementary on the metabolic scale). Based on data from a gnotobiotic approach applied on the model plant <i>A. thaliana</i>, modelling studies were undertaken. The studied data here are an OTU abundance table and omics data (genome and proteome) about the plant microbiotic bacteria. Two types of networks are built and combined in order to study the influence of metabolic complementarity on microbiotic community assemblies: networks of potential interactions between OTUs and meta-metabolic networks. It appears that there is no maximization in numbers of functional annotations in our dataset. But according to a toy subnetwork of OTUs, there is a better metabolic completion between OTUs that are often associated than for those that are not. These results are encouraging and validate the proof of concept of our starting hypothesis. These new modeling methods are part of the emergence of the "Systems Ecology": a very recent concept which is interested in complex ecological systems, on the same principle as the Systems Biology, especially through the interactions between organisms.</p>	
Mots-clés : Modélisation Réseaux Assemblage Microbiome Végétal Key Words: Network Modelling Plant Microbiome Assembly	

* Élément qui permet d'enregistrer les notices auteurs dans le catalogue des bibliothèques universitaires