



Genome-scale modeling, prediction and experimental testing of cross-feeding interactions within the root microbiota of *Arabidopsis thaliana*: a microbial system ecology framework

Victor Mataigne

► To cite this version:

Victor Mataigne. Genome-scale modeling, prediction and experimental testing of cross-feeding interactions within the root microbiota of *Arabidopsis thaliana*: a microbial system ecology framework. Ecology, environment. Université de Rennes, 2021. English. NNT : 2021REN1B050 . tel-03640143

HAL Id: tel-03640143

<https://theses.hal.science/tel-03640143v1>

Submitted on 13 Apr 2022

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

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

THÈSE DE DOCTORAT DE

L'UNIVERSITE DE RENNES 1

ECOLE DOCTORALE N° 600

Ecole doctorale Ecologie, Géosciences, Agronomie et Alimentation

Spécialité : *Ecologie et Evolution*

Par

Victor Mataigne

Genome-scale modelling, prediction, and experimental testing of cross-feeding interactions within the root microbiota of *Arabidopsis thaliana* : a microbial system ecology framework

Thèse présentée et soutenue à Rennes, le 08/12/2021

Unité de recherche : UMR 6553 ECOBIO

Rapporteurs avant soutenance :

François Buscot
Jérémie Bourdon

Professeur, Helmholtz Centre for Environmental Research - UFZ
Professeur, UMR 6004 - Laboratoire des Sciences du Numérique de Nantes

Composition du Jury :

Anne Siegel (Présidente)
Christophe Mougel
François Buscot
Jérémie Bourdon

Directeur de recherche CNRS, IRISA - INRIA
Directeur de recherche INRA, UMR IGEPP
Professeur, Helmholtz Centre for Environmental Research - UFZ
Professeur, UMR 6004 - Laboratoire des Sciences du Numérique de Nantes

Dir. de thèse : Philippe Vandenkoornhuyse
Co-dir. de thèse : Stéphane Hacquard
Research

Professeur, UMR 6553 ECOBIO - Université de Rennes 1
Research Group leader, Max Planck Institute for Plant Breeding

Genome-scale modelling, prediction, and experimental testing of cross-feeding interactions within the root microbiota of *Arabidopsis thaliana* : a microbial system ecology framework

Victor Mataigne

PhD thesis

Doctoral school EGAAL, University of Rennes 1, UMR CNRS 6553 ECOBIO

Co-supervised and co-financed by the Max Planck Institute for Plant Breeding Research

2021

Publications

Mataigne V., Vannier N., Vandenkoornhuyse P and Hacquard S (2021) Microbial Systems Ecology to Understand Cross-Feeding in Microbiomes. *Front. Microbiol.* 12:780469. doi: 10.3389/fmicb.2021.780469

Mataigne V., Vannier N., Vandenkoornhuyse P., Hacquard S. Multi-genomes metabolic modelling predicts functional inter-dependencies in the *Arabidopsis* root microbiota. (*submitted to Microbiome*).

Mataigne V., Vannier N., Vandenkoornhuyse P., Hacquard S. Cross-feeding and metabolic diversity promote coexistence in *Arabidopsis thaliana* root microbiota. *In prep.*

Mataigne V., Vannier N., Vandenkoornhuyse P., Hacquard S. Phylogenetic and functional clustering in a representative bacterial consortia of the *Arabidopsis thaliana*'s root microbiota. *In prep.*

Oral presentations

Mataigne V., Vannier N., Vandenkoornhuyse P., Hacquard S. Omics metrics for ecological niche potential and metabolic interactions. *World Microbe Forum 2021*, June 2021 (Rapid Fire Presentation)

Mataigne V., Vannier N., Vandenkoornhuyse P., Hacquard S. Hypothetical cooperations and growth media strongly affect the metabolic potential of a bacterial community. *Journées Scientifiques (EGAAL 2021)*, Rennes, France, July 2021 (Poster)

Remerciements

Tout ce travail ne s'est pas fait tout seul. Vous étiez si nombreux à m'accompagner que si je prenais le temps de vous remercier un par un comme il se doit, cette section du manuscrit serait plus longue que n'importe quelle autre.

À Philippe et Stéphane, un immense merci pour votre encadrement, idées, corrections, confiance et soutien. Comme le disais un célèbre anonyme, « la liberté commence quand on a deux chefs ». En dépit du caractère humoristique de cette (fausse ?) citation, cela n'a jamais été aussi vrai au regard de la grande autonomie et indépendance que vous m'avez laissé dans mon emploi du temps et mon travail. C'est bien connu, l'aspect le plus important d'une thèse c'est de bien accrocher avec ses encadrants, et on ne peut certainement pas dire que j'ai été déçu de ce côté là.

Nathan, on peut te considérer comme le 3ème co-directeur de cette thèse. J'aurais été bien incapable de faire ces fichues manips sans toi (et ce sans compter ma maladresse avérée dès que je suis ailleurs que devant du code). Tu t'es même farci tout seul la seconde (heureusement plus petite) manip, pour cause de confinements à répétition en raison d'une certaine pandémie bien malvenue. Je te suis redevable en bières à l'œil pour un certain temps.

Je remercie également chaleureusement les membres de mon comité de thèse, pour le temps et l'intérêt qu'ils m'ont consacré et leurs conseils avisés : Anne Siegel, Damien Eveillard, Cédric Wolf, Ning Ling (just call him Ning, as he said once ; Ning if you ever read this page : thank you !), et Philippe Bordron, qui en fait me coach depuis le master de bioinfo à Roscoff. Merci également aux membres de mon jury de thèse, qui ont accepté d'évaluer mon travail sous sa forme finale : Anne Siegel (de nouveau !), François Buscot, Christophe Mougel, et Jérémie Bourdon.

Au labo, il y a un bureau à citer, un bureau particulier, rempli de gens bizarres, un peu dingues, un peu dissipés, avec parfois les éclairs de lucidité et de discipline requis pour avancer dans son planning. J'ai atterri dans ce bureau là, et de mon plein gré en plus. Bien caché dans mon coin de la pièce, j'ai pu échapper à la folie qui a gagné petit à petit les autres occupants : des gens qui pensent que la tomate est un fruit qui se récolte à coups de pieds, rendez-vous compte. Cela dit, il faut reconnaître que c'était sacrément marrant. Merci Marine, Lucie, Claire, Léa, Paola, Solène, pour des moments aussi absurdes que nécessaires : pour garder une bonne santé mentale, autant faire semblant d'être déjà fou.

Dans le bâtiment adjacent, d'autres ne sont pas en reste : Lorène, Laura, Emma, Kaïna, Sumen, vous êtes les meilleurs, autant au labo qu'au bar, sur un mur d'escalade, en musique, sur un tapis de yoga, ou quand il s'agit de me filer à manger. Cela aurait été trois ans beaucoup moins fun sans

vous. Et un jour viendra où 1) on passera du 9A à la grimpe, 2) on jouera et chantera en rythme, 3) je réussirai à manger très pimenté, et surtout 4) on aura le dernier mot : des bières belges ou bretonnes, lesquelles sont les meilleures ? (divulgâchage : c'est les belges).

Merci à tous les autres membres d'ECOBIO, ceux de mon couloir comme les autres pour leur sympathie, leur présence, ainsi que tous leurs restes de repas (mention spéciale à Cendrine pour les chocolats en abondance, à Bertrand pour tous ces jolis dessins très inspirés, un style très personnel, à Yann pour la musique, et à Isabelle, « parce que Isabelle c'est la meilleure »).

Sortons un peu du labo. Il y a une foule de gens là dehors : Émilie, Gaëlle, Prisca, Aliénor (la sale gosse la plus cool), Anna, Ana (notez la subtile différence dans le nombre de « n ») Cassandre, Eladio, Gilles, Lou, Babak , Roberta, Maëlle, Meg, -je continue hein, je ne me permettrai pas d'oublier quelqu'un, mais vous êtes sacrément nombreux-, Adrien (ou les aventures de deux Francs-comtois en Bretagne), Hugo, Hugo 2 (les mecs faites un effort quoi, prenez exemple sur Anna), Nicolas, Aby, Farzan, Mehdi, Adam, Quentin, Jean-Gab, Tifenn, Laura 2, Laura 3 (même remarque que pour les Hugo), et Zoé. Émilie et Gaëlle, une reconnaissance toute particulière pour ne m'avoir jamais entraîné dans un traquenard à base de houblon en pleine semaine (comment ça personne ne me croit ?). Merci à mes copains de master Geoffray, Annthomy, et Kévin, sans qui mes parties de jeux vidéos n'auraient pas la même saveur (mais seraient cependant plus reposantes... « *Calmez vous !* »). On ne se voit pas souvent, mais grâce à Discord, on s'entend beaucoup. A propos des copains, il y en a tant d'autres depuis le début de ces longues études, qu'ils soient de Franche-Comté, Belgique, ou Madagascar ... Merci à vous tous de m'avoir rappelé qu'il n'y a pas que la recherche dans la vie.

Retournons au labo, mais un autre : un grand merci à tous les membres de mon équipe au Max Planck Institute pour leur accueil et leur aide (pour les curieux, ça s'écrit *Max-Planck-Institut für Pflanzenzüchtungsforschung*, pour la prononciation, adressez-vous à Aliénor svp). Merci également aux musicos de l'institut pour toutes ces jam !

Les meilleurs pour la fin : merci à mes adorables parents toujours là pour me soutenir, même s'il serait temps d'en terminer définitivement avec les études, même payées. N'oublions pas mes chères petites pestes de frangines (je vous adore), et bien évidemment, ma famille dans son ensemble.

A l'instar d'un célèbre scribe, « travailler dans l'urgence, ça peut être très intéressant ». L'urgence s'est ressentie sur la fin, donc merci à vous tous de m'avoir maintenu sur pied sur la fin de ces trois années, quand il a fallu tout finir en même temps.

Table of contents

Chapter 1. General Introduction.....	11
Part I – Microbiomes and microbial ecology.....	13
1. The world of microorganisms.....	13
1.1 What and where are microorganisms ?.....	13
1.1.1 The biogeography of microorganisms.....	14
1.1.2 Temporal patterns in microbial communities.....	16
1.2 Microbiology and microbial ecology.....	18
1.3 Microbes are fundamental in ecosystems.....	18
1.3.1 Implications in biogeochemical cycles and ecological processes.....	18
1.3.2 Microbiotas as a host symbionts.....	20
1.3.2.1 The gut microbiome.....	21
1.3.2.2 The plant microbiome.....	22
1.3.3 Microbiome functional diversity and redundancy, core microbiome.....	25
2. Microbiomes' assembly rules.....	27
2.1 Processes behind communities assembly.....	27
2.1.1 The relative importance of deterministic and stochastic events.....	30
2.2 The niche theory, habitat filtering, and the neutral theory.....	30
2.3 Factors and events affecting microbial communities.....	34
2.3.1 External or abiotic factors.....	34
2.3.2 Selection by the host.....	34
2.3.2.1 Vertical transmission.....	34
2.3.2.2 Horizontal transmission and pseudo-vertical transmission.....	35
2.3.3 Microbe-microbe interactions.....	35
2.3.3.1 An overview of ecological interactions.....	35
2.3.3.2 Time and spatial patterns affect existence, magnitude, and outcomes of ecological interactions.....	38
2.3.3.3 Interactions existence, magnitude and outcome are environment-dependent.....	39
3. Data and methods in Microbial ecology.....	41
3.1 The -omics revolution.....	41
3.2 Metabolic networks and system biology are the basis to study metabolic interactions.....	43
Part II. Metabolic cross-feeding with Microbial System Ecology to disentangle coexistence in microbiomes: A mini-review.....	49
1 Introduction.....	51
2 Metabolic cross-feeding as a major driver of microbiota assemblages.....	52
2.1 Evolution and stability of cross-feeding.....	54
2.2 The growing importance of metabolic cross-feeding compared to competition.....	55
3 Microbial Systems Ecology: a crossroads between system biology and community ecology.....	57
3.1 Microbial Systems Ecology approaches and framework.....	5
3.1.1 Metabolic network reconstruction.....	58
3.1.2 The microbial systems ecology framework calls for shifts between top-down and bottom-up approaches.....	60

Table of contents

4 Conclusion.....	62
Chapter 2 – Multi-genomes metabolic modelling predicts functional inter-dependencies in the <i>Arabidopsis</i> root microbiome.....	63
1 Introduction.....	66
2 Materials and Methods.....	68
2.1 Genomes data.....	68
2.2 Reference database.....	69
2.3 Metabolic networks (GEMs) reconstruction.....	69
2.4 Metrics from genomes and GEMs.....	69
2.5 Targeted Predicted Producible Metabolites (TPPM).....	70
2.6 Nutritional constraint (growth media) modeling	70
2.7 Putative GEMs combinations for metabolic interactions.....	71
2.8 Quasi-Poisson GLMs	72
2.9 Other statistical analyses.....	73
2.10 Scripting.....	73
3 Results.....	74
3.1 A link between genome-predicted unconstrained metabolism and phylogeny.....	74
3.2 SynCom unconstrained PPM are greater than that of single strains and rapidly reaches saturation	74
3.3 PPM and TPPM number and composition depends on nutritional constraints.....	78
3.4 Simplest SynComs are predicted to produce TPPM through metabolic exchanges.....	80
4 Discussion.....	85
4.1 Fundamental ecological niche signature in GEMs.....	85
4.2 Phylogenetic distance, similarity and complementarity, antagonism and cooperation in SynComs.....	85
4.3 Metabolism is nutritional-constraint dependent.....	86
4.4 Metabolic dependencies are predicted to be major drivers of microbial communities structure. 87	87
4.4.1 Metabolic exchanges are nutritional-constraint dependent and compensate severe growth constraints.....	87
4.4.2 Minimal combinations of GEMs reflect functional redundancy for targeted compounds.....	87
4.4.3 Genomes size effects remains unclear.....	89
4.4.4 Few strains are enough to reach the community's potential.....	89
5 Conclusions and prospects.....	90
Supplements.....	91
Chapter 3. Cross-feeding and predicted metabolic diversity promote coexistence in <i>A. thaliana</i> root microbiota.....	99
1 Introduction.....	102
2 Material and methods.....	104
2.1 Culture collection and genome-scale metabolic models.....	104
2.2 Construction of the predicted produced metabolites gradient.....	104
2.3 Bacteria cultures.....	106
2.4 Cultures in split-system.....	107
2.5 DNA extraction, purification, and sequencing.....	108

Table of contents

2.6 Formatting sequencing data.....	109
2.7 Data analysis.....	109
3 Results.....	110
3.1 Reduced competition in SynComs with high PPM.....	110
3.2 Cases of cross-feeding identified among many competitive situations.....	110
3.3 An <i>Achromobacter</i> sp with an apparent benefit from cross-feeding in three SynComs.....	111
3.4 The constant strain is a slow grower and a weak competitor that could benefit of non-contact secretions.....	113
4 Discussion.....	114
4.1 The prevalence of antagonistic interactions.....	115
4.2 Some strong cross-feeding patterns for <i>Achromobacter</i> sp.....	115
4.3 Cross-feeding could exist in many cases but might be masked by antagonistic interactions.....	116
4.4 The importance of spatial configuration and metabolites flow.....	116
5 Perspectives.....	117
Supplements.....	119
Chapter 4. General discussion.....	134
1 Metabolic cross-feeding, coexistence, and community assembly : a summary.....	134
1.1 The prevalence and importance of cross-feeding.....	134
1.2 Cross-feeding selective drivers.....	135
1.3 Integration of cross-feeding with competition and environmental factors : from <i>in silico</i> predictions to experimental validation.....	136
1.3.1 The difficulty to find simple and appropriate metrics.....	136
1.3.2 Cross-feeding as a mostly underlying process compared to competition ?.....	137
1.3.3 A few competitive behaviors at the origin of many cross-feeding opportunities ?.....	138
1.3.4 Environmental constraints.....	138
2 Metabolic cross-feeding is not only about GEMs and various padlocks have to be considered.....	139
2.1 Metabolites availability, secretion, transport, and uptake.....	139
2.2 Root exudates composition and their effect on microbiota.....	140
2.3 Spatial scale.....	140
2.4 Genomes annotation.....	141
2.5 Models complexity or simplicity ?.....	141
3 Conclusion and perspectives.....	142
Supplementary chapter 5 Phylogenetic and functional clustering in a representative bacterial consortia of the <i>Arabidopsis thaliana</i> 's root microbiota.....	144
1 Introduction.....	146
2 Material and methods.....	148
2.1 Data acquisition.....	148
2.2 Data filtering.....	149
2.3 Co-occurrences network.....	149
2.4 Bacterial strains metrics, and scale to the OTU level.....	149
2.5 Modules' phylogenetic distances.....	150
2.6 Accumulative null models.....	150
3 Results.....	151

Table of contents

3.1 The co-occurrences network follows a scale-free topology.....	151
3.2 Phylogenetic clustering and functional redundancy in modules.....	152
4 Discussion.....	154
4.1 The network is a basis to hypothesize interactions, notably between keystone species and the host.....	154
4.2 Phylogenetic and functional clustering reveal habitat filtering.....	155
5 Conclusion and prospects.....	156
General Bibliography.....	158

Chapter 1. General Introduction

Since their discovery in the 16th and 19th centuries, micro-organisms are known to be dramatically diverse and distributed everywhere around the globe (Martiny et al., 2006; Opal, 2009). In common language, they are often referred to as microbes, but microorganisms include not solely bacteria but also archaea and unicellular Eukarya. Long believed to be pathogenic, microbiology was consequently mainly oriented towards hygienist policies or development of antibiotics (Opal, 2009). Such approaches were successful in terms of public health but, in parallel, researchers (then people) realized the huge, beneficial impacts they have of their environment (Whitman et al., 1998; Gentile and Weir, 2018).

Microbes are indeed implied in numerous environmental processes such as nutrients cycle and other biochemical processes, and interact in multiple ways with other organisms, particularly when they live as a host's symbionts (Larimer et al., 2010; Fierer, 2017). Such communities of microbes, named microbiotas when associated to a host (or microbiomes to consider the microbial community genomes), are indeed beneficial to their host, such as protection against pathogens or diseases, nutrition, resistance to various stress, such as pH, heat, drought, or salinity in plants microbiotas (Rodriguez et al., 2019).

Such a paradigm shift led researchers to wonder how microbes interact with animals and plants, as well as with their whole biotope, giving birth to a new field of research, microbial ecology, which aims to decipher how microbial communities are structured, evolve, and react to environmental variations. Thanks to tremendous advances in molecular biology and DNA sequencing, a major shift occurred in past decades. A set of tools and techniques permitted the characterization of microbial communities of various environments, from soil and oceans to plants and animals compartments (Quince et al., 2017). Firstly descriptive, these techniques are now putting more and more effort into obtaining functional insights of microbial systems.

A way to embrace functional studies is to dive into ecological interactions between microbes, their host or their environment. Interactions are widely diverse, from antagonist behaviors as competition, to mutualistic symbiosis. Effects of the environment include for example climate, nutrients availability, pH... (Yan et al., 2021). Effects from the host encompass for example interactions among organisms that comprise a causal system and gives rise to emergent properties selection from the host, the existence of mutual benefits and their co-evolution and stability, species or individual specificity (Andreote et al., 2014).

A major part of microbial ecology deals with microbes-microbes interactions, from species pairs to multi-species relationships. Indeed, such a quantity of organisms of various species are permanently interacting each other, forming an intricate and complex network (R. Poudel, 2016). In the past few years, metabolic dependencies (often referred as cross-feeding and relying on the secretion of public goods) took more and more attention in research as a possible key for a generic understanding of microbial communities and microbiota assembly process.

Such complexity raises issues with experimental approaches which, despite being the preferred approach of demonstration, are often complemented or preceded with modelling and computational approaches. Thus, recent developments in microbial ecology uses a wide panel of system siology methods which has recently being conceptualize as microbial system ecology (Muller et al., 2018).

Hereafter, we present a summary of microbial ecology, followed by a mini-review of microbial system ecology approaches, mainly focused on bacteria. We then used these methods combined with experimental work in order to explore putative metabolic dependencies within a subset of the bacterial microbiota of *Arabidopsis thaliana*'s roots.

Part I – Microbiomes and microbial ecology

1. The world of microorganisms

Outlines:

In this first section, we provide a general summary of what microbes and microbial communities are, their distribution, and their importance in their hosts and ecosystems balance.

1.1 What and where are microorganisms ?

Micro-organisms (or microbes) are an ubiquitous form of life, in the way they can be found in every ecosystem on Earth. In soils, the first 10 cm are known as the main reservoir of microorganisms on earth, consisting of communities of free-living organisms. Animals and plants contain dense communities of microbes bound to their host, their microbiota, that are involved into many host-microbe and microbe-microbe interactions (Marchesi and Ravel, 2015; Pacheco and Segrè, 2019). Microorganisms exist in tremendous quantities : prokaryotes (bacteria and archaea) alone are estimated to be around $4\text{-}6 \times 10^{30}$ individuals, accumulating 350-550 Pg of carbon, which is around 60-100% of the estimated total carbon in plants (Whitman et al., 1998). In oceans, their biomass is estimated to 10^4 to 10^6 cells per milliliter (Sunagawa et al., 2015). Microbes are far more diverse than any other taxonomic or phylogenetic group : if we stick to their basic definition, i.e. organisms that are so tiny that they cannot be seen by eyesight, they include bacteria, archaea, various eukaryotes like protists, some fungi, some plankton, or even algae (figure 1 and box 1). Some microorganisms can however be visible in some ways, since some are capable of aggregating together, like fungi and algae in combinations forming lichens, or in communities forming biofilms. Their physiological (i.e. functional) diversity is also considerably greater than any other group (Prosser et al., 2007). For example, a community of 800 bacteria recovered from arctic ice displayed several categories of temperature tolerance, and various enzymatic activities (proteolytic, glycosidic, amylolytic, lipolytic, and DNA-hydrolyzing). A wide variability of antibiotics sensibility was also found (Miteva et al., 2004). In another study, a community of 198 bacterial morphotypes from two distinct deep subsurface geological formations near South Carolina were distributed among 21 biotypes (but in an uneven manner, some types being more prevalent than others) (Frederickson et al., 1991).

Box 1: An overview of micro-organisms

Prokaryotes	Bacteria	Bacteria (gathered into the eponymous taxonomic domain) are the first organisms we think of when speaking about microorganisms. Being the first to have been studied, microbiology methods were developed towards how to handle them. They are single-cell prokaryotes (meaning they do not have a nucleus : the DNA exists as a long, folded thread with no specific location in the cell). They exist in various shapes and often appear grouped together. They form a domain of life.
	Archaea	Archaea look similar to bacteria, but this domain of life diverged early in evolutionary history. Archaea widely differ and have specific chemical composition, biochemical activities, and are known to be extremophile, colonizing environments such as hot springs, deep-sea vents, salt flats and of course also more conventional habitats. Archaea are closer to Eukarya than to Bacteria.
Eukaryotes	Fungi	Fungi is a sister group of animals in Opisthokonta and are both multicellular and unicellular.
	Algae	Algae, like plants, are photosynthetic and have cell walls. They occur in moist and aquatic environments. Algae can be unicellular or multicellular (in the late case, they exhibit a wide variety of shapes). Some can reach hundreds of meters long and are thus not considered as microorganisms, but single-celled ones aggregate in mono or multi-specific colonies.
Others	Viruses	Viruses are barely living beings, since they are just a DNA sequence encapsulated into a protective protein capsid. They lack the basic metabolism of other organisms and are obligate parasites which need to infect their host in order to usurp their metabolic functions to reproduce.

Source: Biologydictionary.net Editors. “Microorganism.” *Biology Dictionary*, Biologydictionary.net, 27 Apr. 2017, <https://biologydictionary.net/microorganism/>. & <https://www.britannica.com/science/microbiology/>

1.1.1 The biogeography of microorganisms

Despite the absolute prevalence of microbes all around the globe, microbial biogeography exhibits clear spatial patterns, for instance driven by abiotic factors like ocean depth, which, as the outcome of several environmental factors, stratifies the diversity of oceans microbiomes (Sunagawa et al., 2015). The clearest evidence is the uneven distributions of species along locations (even for taxa found everywhere), for example latitudinal gradients, and the existence of endemic (i.e. restricted to a location) species microorganisms (Oakley et al., 2010). Different environments exhibit different communities (example in figure 2, taken from Schulz et al., (2017), particularly extreme environments such as arctic zones, hot springs or abysses which exhibit highly specific microorganisms communities (Hanson et al., 2012). Most bacteria are then restricted to specific environments (animal, plants, seawater, soil ...) which overlap only a few (Nemergut et al., 2011).

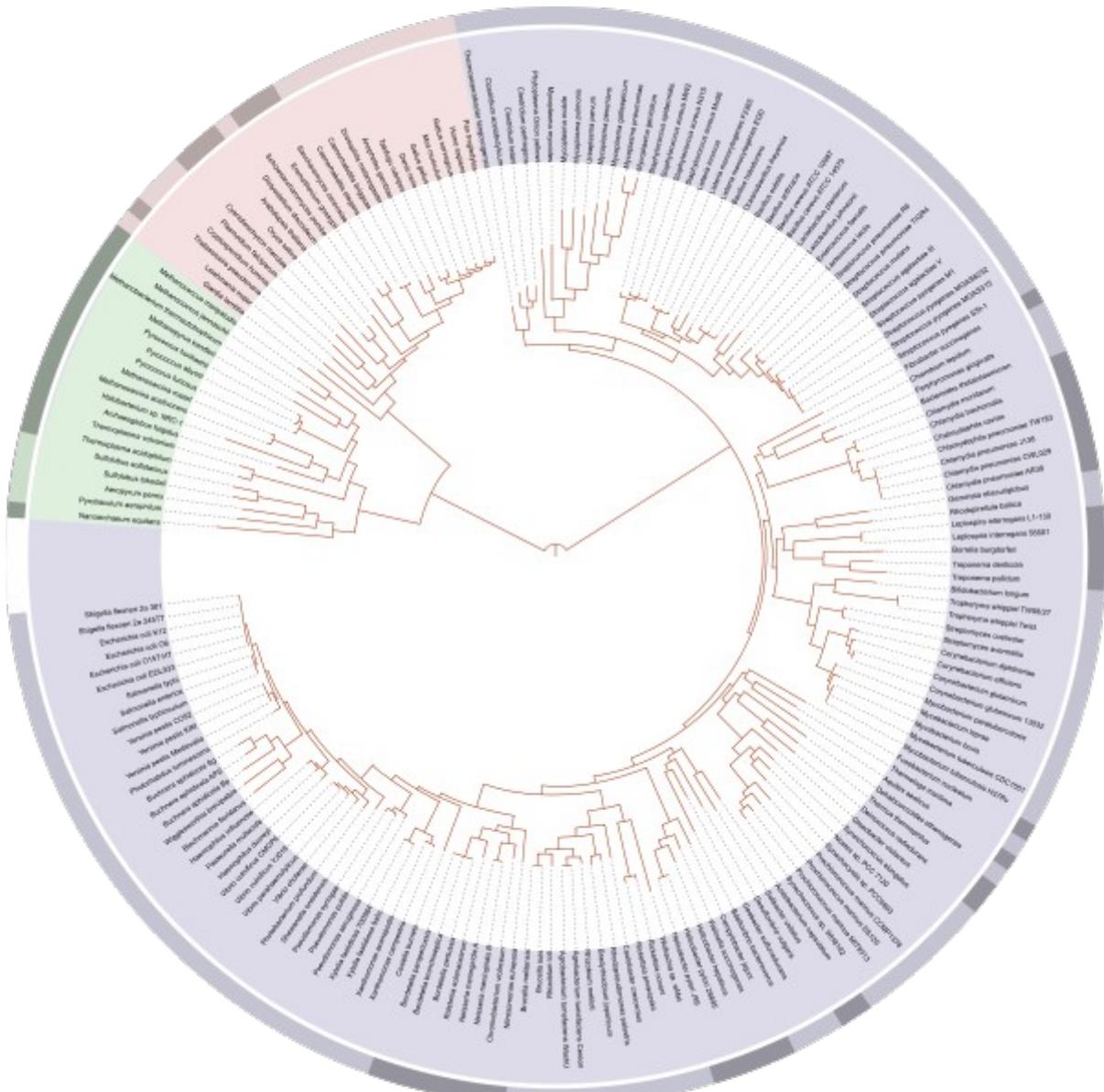


Figure 1: The phylogenetic tree displaying the three kingdoms of life. Bacteria (gray) are the most diverse group. Archaea are in green, and eukaryotic microorganisms are in red. Note that leaves names are species cited as examples representative of their whole taxa: many internal branches are not displayed, as well as taxonomic levels such as phyla, order or families. (Creative Commons figure, from Wikimedia Commons).

A second evidence is that genetic similarity patterns among microbial taxa in different locations are known. Population genetics used these patterns to study divergence among locations and find a history of their spatial and temporal evolution. It showed for example that genetic divergence is correlated with spatial distance (as for macro-organisms), formalized as the distance-decay relationship (Soininen et al., 2007), illustrated on figure 3 (after Nemergut et al., (2013), modified). For example, distance-decay is observed in microbial eukaryotic communities of Sturt National Park, Australia (Green et al., 2004).

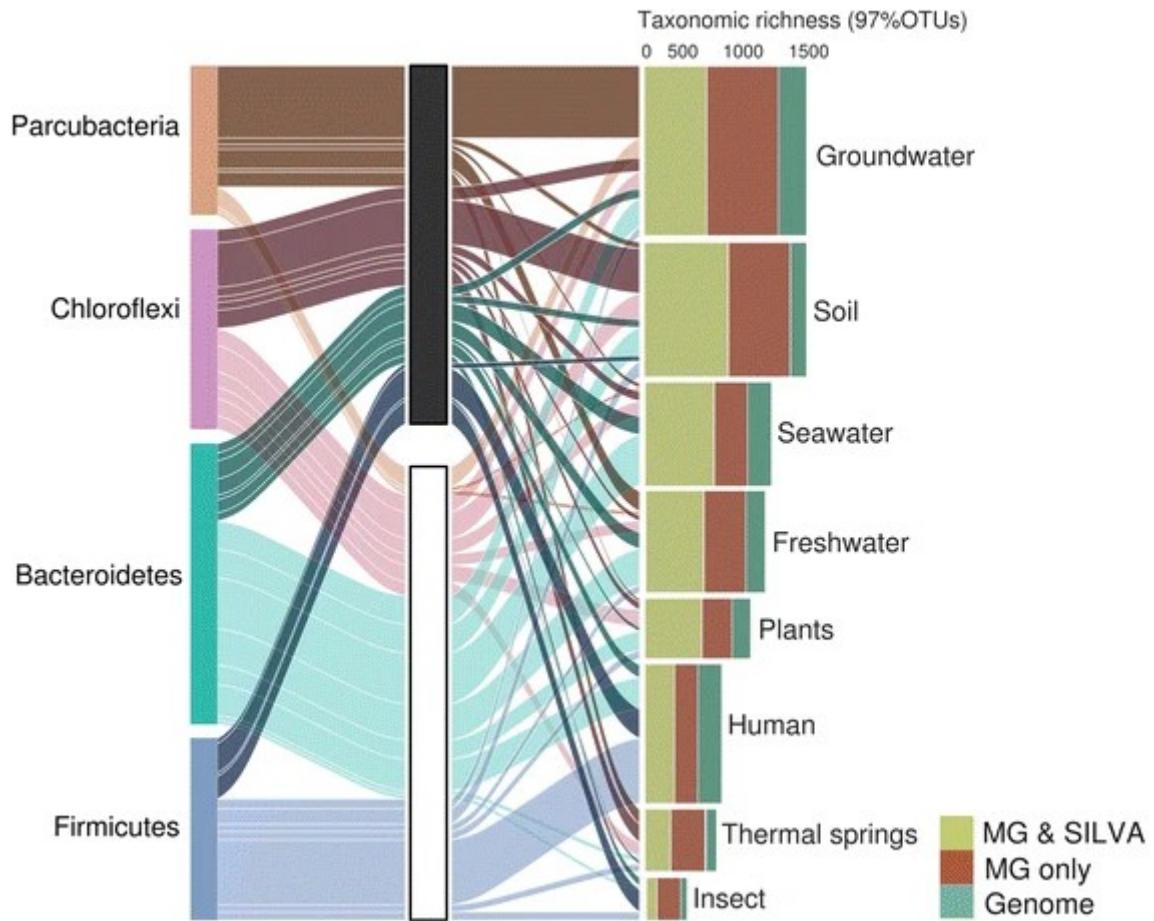


Figure 2: Bacterial phyla prevalence varies among environments. For example, Parcubacteria are mostly associated with groundwater, and Firmicutes to humans. The central column and colored stack bars in the second and third column detail if genomic sequences were obtained from metagenomes exclusively (black+brown) or were previously observed either in the SILVA database or by genome sequencing (white+green and blue). Sequences are clustered in OTUs (Operational Taxonomic Units) (Creative Commons figure, from Schulz et al., 2017).

1.1.2 Temporal patterns in microbial communities

Microbial composition also displays temporal patterns, within years but also within seasons or days. A recent example is the existence of a persistent microbial seed bank in the English Channel, microbial relative abundance being variations of this reservoir (Caporaso et al., 2012). Depending on the community and environment, timescales and spatial scales can vary dramatically. For example, in surface seawater communities, variations are typically in the order of days to weeks, kilometers in the horizontal direction, and in the vertical direction, millimeters for the immediate sea-surface micro-layer (Fuhrman, 2009). Another example is air microbiome : its composition was found to be affected by multiple temporal factor : seasonality, night/day cycle (which might be confounded with temperature variation) affected both absolute and relative abundance of microbial taxa (Gusareva et al., 2020). (figure 4).

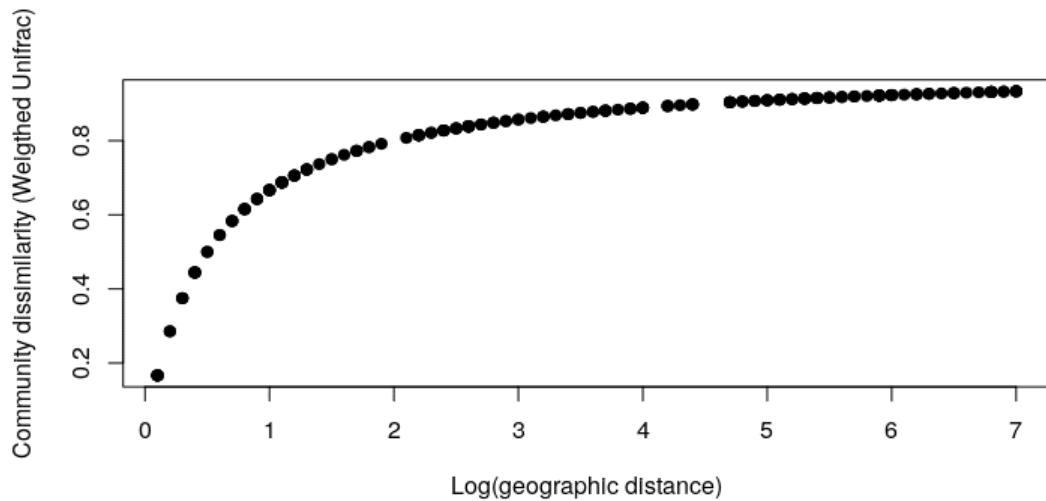


Figure 3: The relation between geographic distance and dissimilarity (represented as UniFrac distance, which is a measure of phylogenetic dissimilarity) between soil rotifers communities (after Nemergut et al. 2013, modified).

Finally, differential functional diversity is at stake regarding the environment or the community : microbes living in plants or guts will not display the same metabolism as free-living microbes in the ocean. Each community is representative of its environment. However these differences can be more subtle. For example, Gut and ocean microbiomes display large physiological differences, but were found to have an overlap of functions which gathers much more genes abundances than specific functions. In addition, significant difference of genes abundances between functional categories were found between the two ecosystems (Sunagawa et al., 2015).

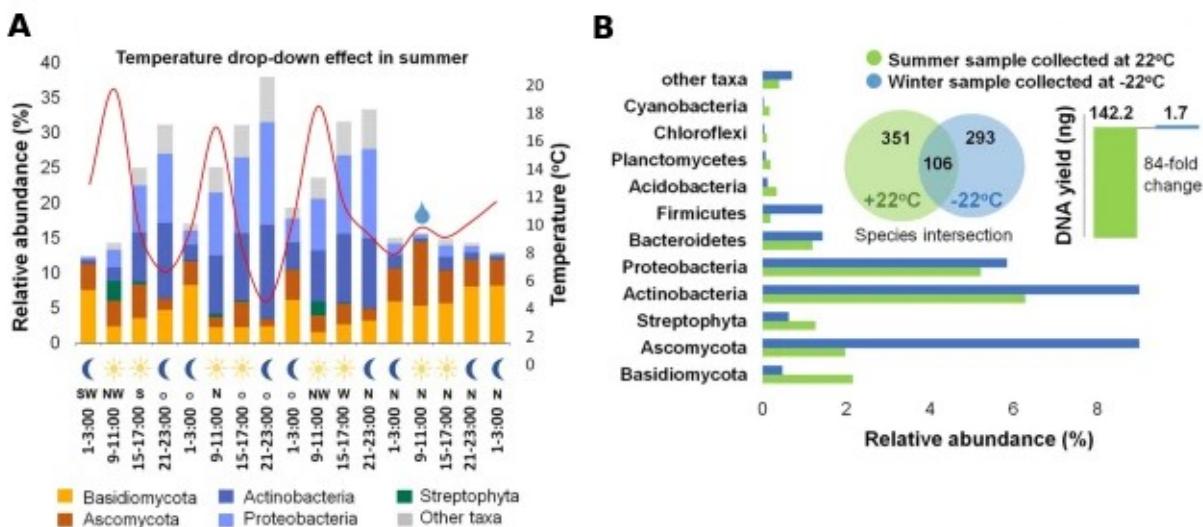


Figure 4: temporal patterns of the Air microbiota. A) variation in the relative abundance of bacterial taxa according to a day/night cycle (temperature effect). B) abundance variation of bacterial taxa according to season. The sampling location was in West Siberia (Creative Commons figure from Gusareva 2020).

1.2 Microbiology and microbial ecology

In 1859, Darwin formalized the theory of evolution by random variation and natural selection. Coupled to the re-discovery of Mendel's laws of heredity, the elucidation of DNA's structure by Franklin, Watson and Crick, and the development of molecular biology, this was the foundation of all biology in the 20th century, including microbiology (Opal, 2009). Various applied branches appeared, such as soil microbiology or food microbiology. Finally, microbiology expanded dramatically with the development of microbial ecology, which studies microbes in their ecosystem and roughly aims to find “who's there” and “what are they doing”. Community ecology concepts started to be applied on microbial communities in an attempt to decipher interactions among microbes that comprise a causal system and give rise to emergent properties. However it encountered issues rapidly due to the lack of theoretical framework, preventing any mechanistic analysis and predictive power, for two main reasons. First, an initial lack of data (mainly identification and *in vitro* culture issues) prevents the acquisition of theoretical tools required to detect underlying processes of microbial ecology. Second, ecological theories were not part of the microbiology mindset, which had a reductionist, mainly physiological approach (Prosser et al., 2007). Nonetheless, microbial ecology flourished rapidly and underwent a revolution in the 2000'. Thanks to progress in DNA sequencing and the development of other -omics data technologies, molecular data accumulated quicker and quicker, revealing the huge taxonomic and functional diversity of microbes. The main issue of microbial ecology was then to define rigorously what is a microbial community, notably for delimiting contiguous environments at various time and spaces scales, in order to differentiate communities, microbes being everywhere (Konopka, 2009).

1.3 Microbes are fundamental in ecosystems

1.3.1 Implications in biogeochemical cycles and ecological processes

Microorganisms play a fundamental role because of their implication into biogeochemical processes and organic matter recycling. Oceans' microbial life is particularly essential, as they are responsible for nearly all the primary production and respiration of this ecosystem (Moran, 2015). Soil microbial communities are key components of nutrient cycling, and their activity drives plant community composition and productivity (Bérard et al., 2015). The soil microbial decomposers are thus one of the components of the soil fertility ecosystem service (Guo et al., 2020a).

Various biogeochemical processes (methane oxidation, nitrogen/carbon fixation and cycles linkage, phosphorus uptake, sulfite oxidation ...) are insured by diverse microbial communities

(ocean water and sediment, freshwater soil and sediment, contaminated groundwater, bioreactors ...) (Madsen, 2011). For instance, nutrient cycling by microorganisms involves numerous reactions, such as nitrogen fixation, nitrification/denitrification, ammonification (...). Consequently, very little nitrogen escapes to the atmosphere, organic and inorganic nitrogen quantity in the biosphere being at an apparent homeostasis. Both soils and oceans participate to nutrients cycles, but differ in which nitrogen-transforming reactions prevalence, and are assured by different microbial networks of interacting micro-organisms (Kuypers et al., 2018). Such capacities allow microorganisms are also able to treat nitrogen pollution from agriculture, fossil fuels or waste waters (Ogbonna et al., 2000; Kuypers et al., 2018) and there are microbial consortia able to degrade pollutants as among many examples, hydrocarbons from oil facilities and reservoirs (Liu et al., 2018).

Environmental remediations by microbial consortia are developing, as well as agricultural soil enhancement to improved crop yield or biocontrol, for instance with *Trichoderma* strains (Powlson et al., 2001; Woo and Pepe, 2018). More broadly, microbial communities engineering is a growing business (see Box 2). Understanding microbial communities structure is essential to understand ecosystems functions and predict Earth's response to global changes such as warming and ocean acidification, and calls for a deeper knowledge about microbial functions, processes and interactions (Fuhrman, 2009). Indeed, some processes might not be performed by single species, but induced or enhanced by interactions, such as in Dombrowski et al., (2017), where biogeochemical interdependencies between members of hydrothermal vents' microbial communities for carbon-cycling are suggested.

Box 2 : Microbial factories

Microbes have long been exploited for food and beverages : microbial fermentation is indeed at the origin of many alcohols, such as beer. Currently, microbes-mediated chemical production is widely used for the production of various organic chemicals, with applications in biofuels, cosmetics, pharmaceuticals, flavors (Liu and Nielsen, 2019), or plant biocontrol and growth-promotion (Chubukov et al., 2018)... Microbial factories developed with metabolic engineering, a discipline which emerged in the early 1990s, which aims to improve the yield and productivity of microbial processes toward specific biochemical production (Jiang et al., 2020). The framework to build a microbial factory starts from the discovery of implied genes, then pathway design and optimization (i.e. genetic engineering), host optimization if needed, scaling-up from lab to industrial level. Several tools and techniques are employed, from molecular biology to computational biology (Chubukov et al., 2018). However, issues are numerous and improving the

stability and robustness is a major challenge, such as the management of the accumulation of toxic byproducts, accumulation of cofactors, and competition between production and cell growth (Orozco-Mosqueda et al., 2018).

An extend of microbial factories and metabolic engineering is microbiome engineering, where the factory is roughly a part of a host's microbiota, dedicated towards a function, such as pathogen protection, nutrients uptake, or growth promotion in plants, or treatment for obesity in humans (Orozco-Mosqueda et al., 2018; Lim et al., 2020). Methods to inoculate an engineered microbiota into a host are diverse. As an example, a plant can be inoculated from its soil and rhizosphere, its seeds or seedlings, or directly into tissues (Orozco-Mosqueda et al., 2018). Microbiome engineering does not necessarily imply sophisticated molecular techniques, such as genetically modified organisms, but can make great use of ecological processes. For example, soil microbiome can be influenced by agricultural management practice such as organic farming, or by a change in land utilization (Foo et al., 2017). The use of synthetic microbial communities as factories also imply deciphering microbe-microbe interactions. Thus, microbial consortia incorporating interacting strains, for example co-dependent strains by syntrophic exchanges or protection against antibiotics can improve the productivity (for example via distribution of tasks and pathways separation) and robustness of the system.

The design of a synthetic microbial consortia goes with four elements : (1) viability (the capacity to survive the gastrointestinal tract), (2) colonization (the ability to become part of the native gut microbiota), (3) localization of the disease, (4) genetic tractability (whether an organism can feasibly be engineered) (Inda et al., 2019).

1.3.2 Microbiotas as a host symbionts

Microbial communities do not exist as free-living organisms only. Microorganisms are organised within a microbiota when associated with a host. Symbiotic microorganisms within the microbiota are essential to their host development and homeostasis. These microbiota are becoming important objects in microbiome engineering research (box 2).

1.3.2.1 The gut microbiome

Animal-associated microbes mainly colonize the skin surface or the gut epithelium, 99% of them being in the large intestine. In humans, there are an average of 1012 bacterial cells per gram of intestinal content, reaching 1000 different species (probably more due to the difficulty to detect rare species), with a dominance of a few phyla (up to 99% of total abundance) (Zilber-Rosenberg and Rosenberg, 2008) including mainly Firmicutes, Bacteroidetes, Actinobacteria and Verrucomicrobia (Hacquard et al., 2015). The gut microbiota is influenced by multiple factors of the host : phylogeny, age, disease status, geography, diet (Yatsunenko et al., 2012; Gerber, 2014; Li et al., 2017; Gentile and Weir, 2018) (figure 5). For example, regarding diet, the rise of herbivory could have driven the evolution of many mammals' gut microbiome (Hacquard et al., 2015). Diet is however less likely to drive the gut microbiome of birds and flying mammals (Song et al., 2020). Microbe-microbe interactions are highly relevant in the gut microbiome, and a well-known example is about a case of metabolic cross-feeding, where microbes of the human colon form a complex metabolic network of producers and consumers in order to produce fatty acids and other beneficial metabolites (Vázquez-Castellanos et al., 2019).

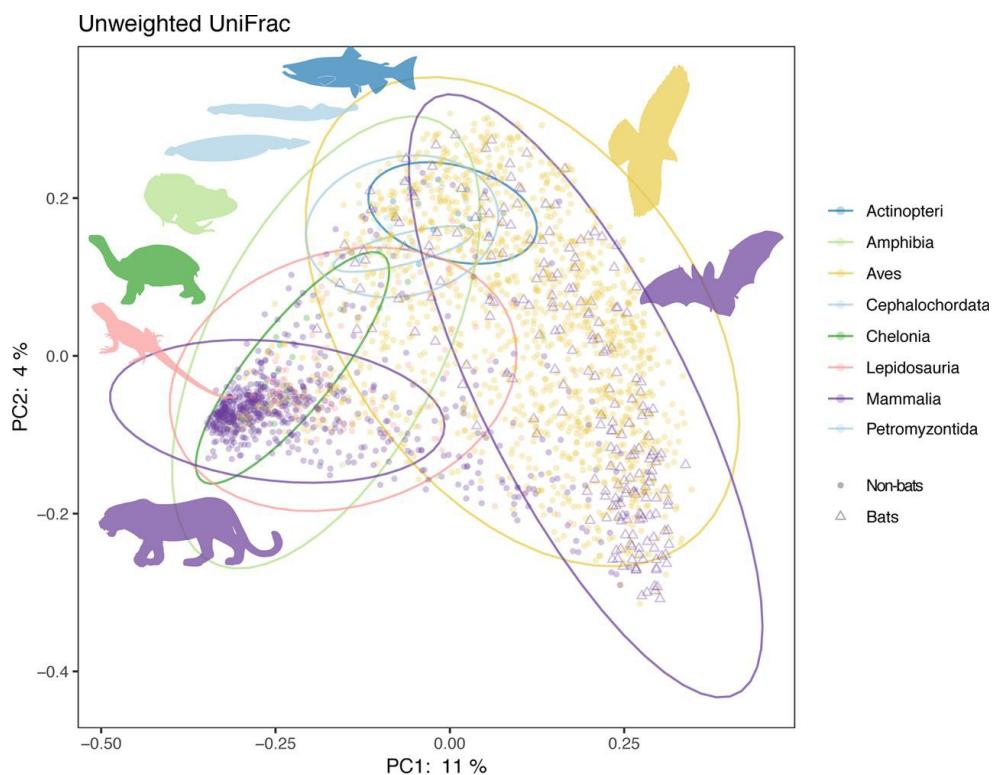


Figure 5: Similarity and dissimilarity of the gut microbiota of various animal taxa, represented by a principal-coordinate analysis of unweighted UniFrac distances between fecal samples. Samples are gathered by class, with a distinction for the order Chiroptera. Species from the same taxa have similar microbiotas, while species belonging to different taxa are dissimilar (from Song et al. 2020, used with authorization).

The gut microbiota has thus an immense potential in medicine, since its has been linked with pathogen and disease resistance (for example by competing with pathogens for resources and space, inhibitory metabolites, killing) (Pickard et al., 2017; McLaren and Callahan, 2020), but its composition can be cause or consequence of host troubles as obesity, malnutrition, diabetes, inflammatory bowel diseases, and even asthma (Kumar et al., 2018; Inda et al., 2019). Imbalanced microbiota can for example be the result of spatial redistribution of species (Tropini et al., 2017) or relative-abundances variations, taxonomic and functional shifts (Lloyd-Price et al., 2019). A direct application, in reference with the previously mentioned microbial factories, would be the design of synthetic microbial consortia aiming to modulate the gut microbiome and restore homeostasis (Vázquez-Castellanos et al., 2019). Some probiotics or lactic bacteria have already been engineered to diagnose conditions in the gut and synthesize therapeutics or anti-inflammatory cytokines *in vivo* (Inda et al., 2019).

1.3.2.2 The plant microbiome

In plants, the endosphere root microbial communities are overall structured similarly, belonging to only four bacterial taxa : Proteobacteria, Actinobacteria, Bacteroidetes and Firmicutes (Proteobacteria being always enriched compared to the surrounding soil). Such a taxonomic signature is ubiquitous, indicating the robustness of the forces behind root microbiota composition (Hacquard, 2016). In contrast, leaf-associated microorganisms display more compositional and seasonal variations, being more subject to stochastic processes (Maignien et al., 2014; Vandenkoornhuyse et al., 2015). Also, seeds ecology has its importance into the acquisition of a plant's microbiota (Nelson, 2018). Root and leaves compartments harbor shared ‘systemic’ microorganisms including bacteria and fungi (Vandenkoornhuyse et al., 2015). The composition of the plant microbiome depends on various factors, whose effect size vary according to the compartments (rhizosphere, endosphere, phyllosphere) and the microorganisms taxa : soil properties, land use, host genotype and functional profile (Schöps et al., 2018; Wu et al., 2018).

As for the animal microbiota, the plant microbiota displays important functions to the host, from an enhanced nutrition to resistance to various stresses as drought, heat, or salinity (figure 6) (Hacquard et al., 2015; Orozco-Mosqueda et al., 2018). Regarding plant nutrition, mycorrhizas, a symbiosis that ‘help to feed the world’ (Marx, 2004) represent arguably the most frequent symbiosis on earth (Smith and Read, 2008), i.e. a symbiosis formed with ~90% of the land plants whatever the ecosystem considered. Conditioned mutual exchanges of mineral nutrients and water from a fungi to the plant, in exchange of photosynthesis-derived organic compounds have allowed the stability of this key symbiosis since more than 400 millions of years (Kiers et al., 2011). Bacteria also

participate in plant nutrition. The most representative examples are symbiotic nitrogen-fixing bacteria in leguminous plants. For instance, various *Pseudomonas*, *Bacillus*, *Azotobacter*, *Serratia*, *Azospirillum* are capable of improving nutrient availability in soil, plant nutrient uptake, as well as being involved in nitrogen cycling. Regarding plants defense, *Trichoderma* may activate a state of alert in the plant, inducing a response to pathogen attack, which eventually anticipates the establishment of an acquired resistance (Woo and Pepe, 2018). Another example is the fungus *Claviceps purpurea* which, despite having a pathogenic behavior, produces alkaloids that are toxic to mammalian species, reducing herbivory of the host plant (Hardoim et al., 2015).

Plants being immobile organisms, unable to migrate to avoid environmental stresses, these host-microbe interactions become even more important. Thus, given the complex microbial consortia and its impact on plants growth and productivity, and the dependency of plants to their microbiota, plants cannot be considered as standalone entities anymore (Vandenkoornhuyse et al., 2015). They are rather defined as holobionts, that is to say a host and its numerous microorganisms associates, in which all organisms are interacting to maintain the stability of the system (box 3) (Vandenkoornhuyse et al., 2015).

To make it short, the plant microbiome impacts the host phenotype and fitness (i.e. its success to survive and reproduce). The fitness measurement of the microbiota components is possible, however measuring plant fitness of the plant alone is much more difficult, because this would require axenic growth and therefore a disconnection from most of the environmental factors. Measures of plant fitness consequently include in general both the endosphere and ectosphere, and match in fact to the fitness of the plant holobiont (Vandenkoornhuyse et al., 2015).

Deciphering the mechanisms underlying the structure and dynamics of the plant microbiome is crucial to go towards a sustainable agriculture. Indeed, maximizing the functions of plant microbiotas, or designing core microbiotas, either gathering all functions or enhancing particular functions, destined to be inoculated are promising ways to counter emerging pathogens, climate change, and make pesticides and fertilizers obsolete (Duhamel and Vandenkoornhuyse, 2013; Toju et al., 2018)..

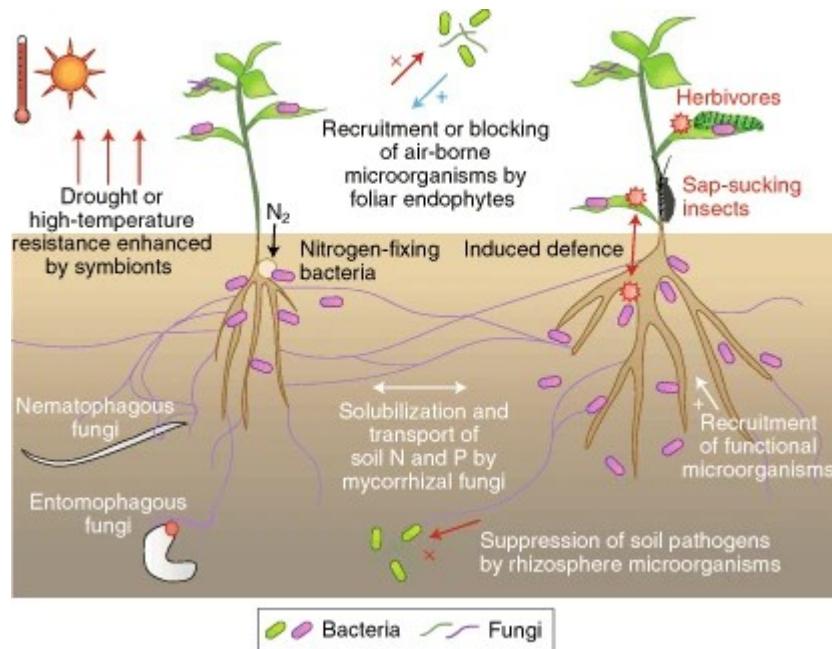


Figure 6: A summary of the benefits brought by the plant microbiota to its host : enhanced nutrition, stress and pathogen resistance (from Toju et al. 2018, used with authorization).

Box 3 : The holobiont and the hologenome

The holobiont is the animal or plant with all of its associated microorganisms, and the hologenome is the sum of the genetic information of the host and its microbiota (figure 7) (Zilberman-Rosenberg and Rosenberg, 2008). The hologenome theory of evolution considers the holobiont as a single entity, in which a majority of the genetic information and variability is brought by microorganisms. Animals and plants alone are then not considered as autonomous entities (Bordenstein and Theis, 2015). The theory was raised from empirical data. First, all animals and plants have symbiotic relationships with microorganisms. Second, part of the symbionts can be transmitted between generations, vertically or pseudo-vertically. Third, the holobiont's fitness is affected by these associations. Finally, the genetic variation encoded in the holobiont can be modulated and change according to environmental conditions faster than the host alone (Zilberman-Rosenberg and Rosenberg, 2008). Highlighting the role of microbes in the evolution of their host, leading to the hologenome theory of evolution, a recent eco-evolutionary framework. Its conceptual and evidence-based foundation are expected to serve as a road map for hypothesis-driven, experimentally validated research on holobionts and their hologenomes, by catalyzing the fusion of biology's sub-disciplines (Theis et al., 2016).

Although coevolution is at stake, it is not the sole feature of the hologenome, and the

hologenome do not arise by cooperation only. All evolution drivers and ecological interactions are involved (Bordenstein and Theis, 2015), and the rules of evolutionary biology are compatible with the holobiont, which are shaped by selection and neutrality. Thus, microbial genomes encode traits which can either be damaging, beneficial, or neutral to the holobiont. A frequent misconception is to assimilate the holobionts as an organ, a superorganism, or a metagenome (Theis et al., 2016). They are neither of that. An organ is composed of similar cells from the same genome, performing specific functions. Superorganisms are used in a context of multiple individuals from the same species, such as ants, bees, or termites colonies. Metagenome refers also to the sum of genetic information in a sample, but does not incorporate the symbiotic aspect of the holobiont (Theis et al., 2016).

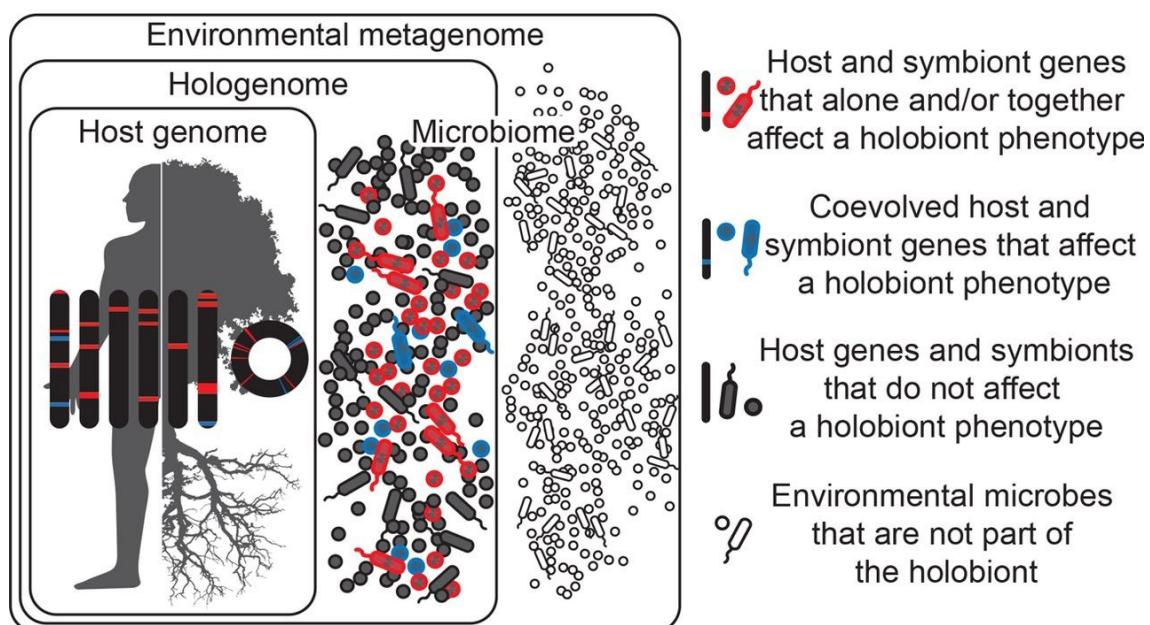


Figure 7: Holobionts are entities composed of the host and all of its symbiotic microbes, whether they have or have not co-evolved with the host. (from Theis et al. 2016, used with authorization).

1.3.3 Microbiome functional diversity and redundancy, core microbiome

Clustered organisms display “core” and “flexible” genomes, which are respectively common to all organisms and organism-specific. Core genomes tend to perform central functions, such as DNA replication or protein synthesis while flexible genomes encode generally accessory functions such as transport reactions, and are suspected to control important aspects in the definition, diversification and overlap of species’ ecological niches (Fuhrman, 2009; Curiel Yuste et al., 2014).

Consequently, microbial communities display both a functional diversity and redundancy, which has its importance in how the community is structured. Functional aspects have to be considered when designing core microbiotas or microbial factories, respectively in order to mimic the functions

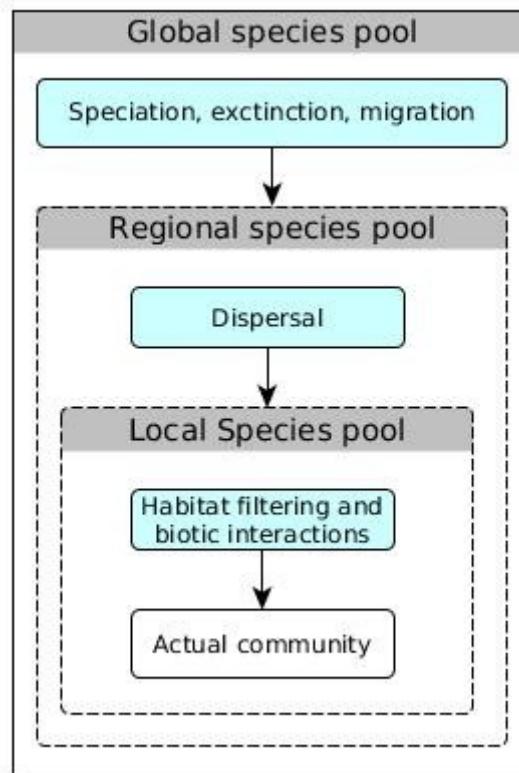


Figure 8: The processes that produce assembly rules in plant communities and the scales at which they are most influential. A global species pool defines a regional species pool through speciation, extinction and migration of species (phylogeographic assembly). The regional species pool defines a local site with species that are able to disperse there (dispersal assembly). Habitat filtering and biotic interactions define the actual community from the local pool (after Götzberger et al., 2012, modified).

of a whole community with a reduced pool of species, or to optimize a biological process (Toju et al., 2018). Redundant functions will be likely to allow broader species choices, while specific ones will likely be more restrictive. In addition, functions taken alone are not enough : since a species might be influenced by another, ecological interactions between microbes must be taken into account, which implies to dive into the study of microbiomes' assembly rules.

2. Microbiomes' assembly rules

Outlines:

This second part deals with the rules and processes governing the assembly of communities. An overview of deterministic and stochastic processes is provided, as well as an overview of abiotic factors at play, followed by biotic factors, i.e. host-microbe and microbe-microbe interactions. Metabolic dependencies between microbes are then more deeply detailed.

2.1 Processes behind communities assembly

A community is defined as a group of organisms representing multiple species living in a specified place and time (Vellend, 2010). Community ecology is the discipline seeking to analyze how biological assemblages are structured (which species and their abundances), what are their functional interactions and how community structure changes in space and time (Konopka, 2009). Understanding how communities assemble has been a central question since the early days of ecology, and the term “assembly rule” was introduced by Diamond (1975), who identified forbidden combinations of species among fruit-eating birds in New Guinea. Assembly rules were then defined as any ecological process selecting for or against a species from the regional species pool, thus defining the local community composition (Götzenberger et al., 2012). In some ways, microbial communities differ from macro-organisms communities, because these processes can act a bit differently (box 4) but in every case however, mechanisms underlying communities patterns are numerous and sorted in four distinct processes (Vellend, 2010):

- Drift, random changes in species abundances (i.e. births, deaths, offspring production).
- Dispersal, the movement of organisms across space, leading to migration.
- Speciation, the creation of new species (relying on mutation)
- Selection, a deterministic process where the most adapted species / individuals can survive, reproduce, and spread whereas the others decline.

Figure 8 displays how these processes articulate together to shape, as an example, the structure of a community of plants. Speciation and dispersal are the two forces which bring new organisms into communities, and drift and selection are the ones affecting changes in the presence, absence, and abundance of species (Nemergut et al., 2013). Correlated to spatial distance (and connectivity), selection and drift increase dissimilarity between communities while dispersal decreases it.

Speciation (mutations) increases dissimilarity regardless of spatial distance (Hanson et al., 2012).

Communities patterns are various (Vellend, 2010):

- Species-area relationships
- Abundances and relative abundances
- Composition-environment relationships
- latitudinal gradients
- Distance-decay of similarity
- Diversity-productivity relationships
- Diversity-disturbance relationships

On the first hand, communities are then driven by stochastic factors. Indeed, drift, speciation and dispersal all rely on random (n.b. : by random, we mean observable but not predictable) events, and on the other hand, by deterministic processes gathered under the term “selection”, because all these processes imply a response of an organism/species to environmental abiotic and biotic factors. Abiotic factors can be environmental, like pH, temperature, humidity, nutrients availability (Nemergut et al., 2013). Biotic factors gather a wide set of ecological interactions between species, each with different outcomes (beneficial/neutral/detrimental, box 6) for each species.

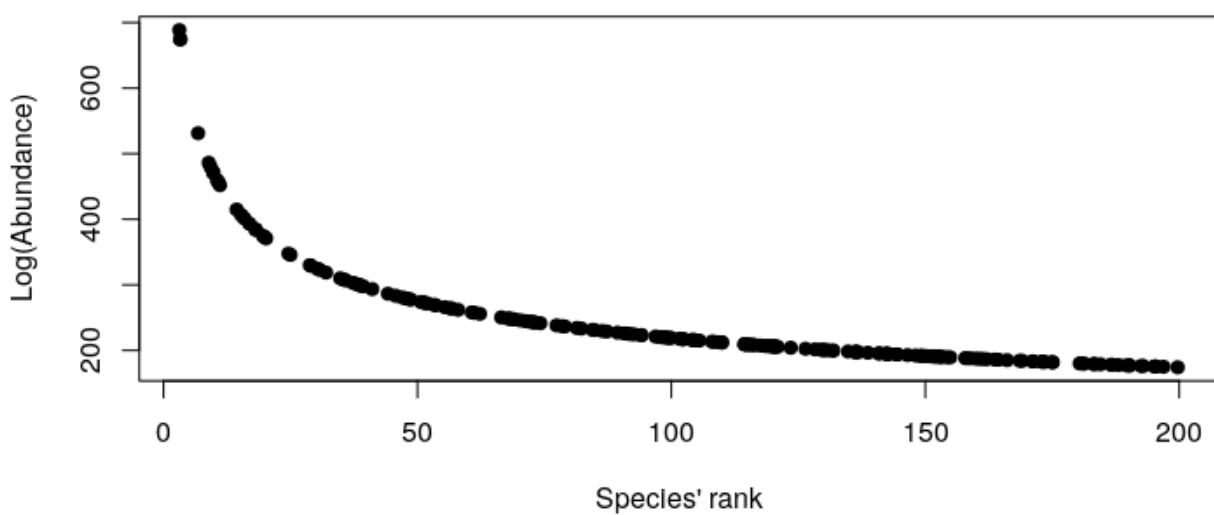


Figure 9: Typical rank-abundance plot followed by most samples of microbial communities. A few abundant species dominate the community, and most species coexist in low abundances (after Nemergut et al. 2013, modified).

Box 4 : the specificity of microbial community ecology

Despite many common principles, microbial community ecology differs in several ways from classical community ecology (even if some macro-organisms can display some of the following traits) (Nemergut et al., 2013). First, the species scale is not always available. For example, bacterial strains can be isolated thanks to the identification of different phenotypes, or species can be differentiated by fixing a threshold of similarity in their 16S RNA (Rosselló-Mora and Amann, 2001). Second, most microbes have no or weak active dispersal activity : dispersal is mainly passive, through wind, water, or “hitchhiking” on moving macro-organisms. Paradoxically, this gives microbes a much higher dispersal potential than macro-organisms, without an expense of energy (Martiny et al., 2006). Third, many microbes have dormancy capacities in response to environmental stresses (Martiny et al., 2006). Microbes also have a short generation time and have rapid evolutionary capacities able to mask interactions (Yoshida et al., 2007), and can exchange genetic material (Ochman, Howard et al., 2000). Fourth, spatial and temporal structure are more complex in microbial communities; for example, because a microbial community can occupy a very tiny spatial scale, even only one gram of soil samples will capture all variations of multiple factors, making it difficult to discover and explain assembly patterns. All together, these traits can lead to ecological responses unique to microbial communities. Last, microbial communities tend to be more phylogenetically structured than expected by chance, containing groups of closely related taxa (however a few communities exhibit the opposite pattern) (Nemergut et al., 2013).

Species abundances of microbial communities usually follow a power law distribution, with a few extremely abundant species followed by numerous species of low and very low abundance (figure 9, Lennon and Jones, 2011). In the meantime, abundant species in a given sample are also highly prevalent in others samples from a similar environment (Nemergut et al., 2011).

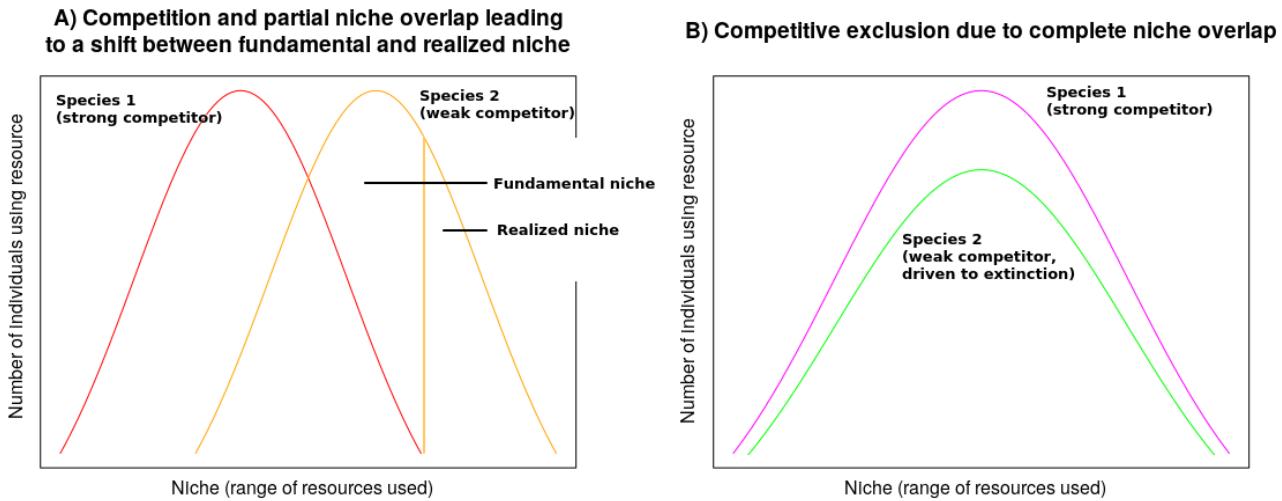


Figure 10: Illustration of how overlapping niches between species are avoided. In (A), two species have partially overlapping fundamental niches for a given resource, which leads to competition. The weakest competitor sees its fundamental niche reduced on a smaller range of the usable resource (realized niche). By natural selection, such cases can lead to niche differentiation, where the two fundamental niches would be clearly distinct. In (B), the two species have a complete niche overlap : the weakest competitor cannot avoid extinction because of competitive exclusion.

2.1.1 The relative importance of deterministic and stochastic events

The relative contribution of stochastic and deterministic events is not clearly known, particularly for microbial communities (Morrison-Whittle and Goddard, 2015). Extreme views exist on the subject, from Hubbell's theory to authors like Clark (Clark, 2009), for which stochasticity is only an attribute of models, artifacts of unknown or left-aside processes. Mostly, it is admitted that deterministic and stochastic processes combine to generate coexistence, with a primacy of selection, neutral processes being mostly underlying. The contribution of each process can also vary depending on time and spatial scales (Morrison-Whittle and Goddard, 2015; Zhou and Ning, 2017). However, it has to be mentioned that the effect of stochasticity is in general more difficult to measure because of a recurrent bias, which is sampling across environmental gradients or habitat types. This sampling emphasizes the strength of environmental selection, and therefore may artificially minimize the effect of drift (Hanson et al., 2012).

2.2 The niche theory, habitat filtering, and the neutral theory

The niche theory assumes that deterministic factors, both abiotic (pH, temperature, salinity...) and biotic (species traits, ecological interactions) shape community structure. It implies that species differ in their niche, the niche being a set of biotic and abiotic conditions defining the volume in which the species can persist (figure 10) (a more precise definition is in box 5). The idea that two species coexisting in the same place must occupy different niches, already present in Darwin's work, and Hutchinson formalized the concept (Pocheville, 2015). A distinction is made between the

fundamental niche, i.e. the maximal, theoretical volume where a species can survive indefinitely , and the realized niche, i.e. a reduced volume, limited because of interactions with present competitors, where the species actually survives. Various frameworks exist under this theory, such as niche differentiation, which implies microbial communities less phylogenetically clustered than expected by chance, and habitat filtering, which relies on the opposite statement.

Niche differentiation (or niche segregation/partitioning/separation) is the process in which competing species use the environment differently to coexist, despite having similar ecological niches. Indeed, according to the competitive exclusion principle, two species with identical niches cannot coexist because of competition, leading one of them to exclusion. Differentiating their respective niche, such as occupying different spaces or consuming different foods (Caldwell and Vitt, 1999; Brochet et al., 2021), facilitate their coexistence. Other species differentiate in their competitive abilities based on varying environmental conditions : for example, some might be more efficient in dry season while others perform better in rainy seasons, such as plants in the Sonoran Desert (Angert et al., 2009). Another kind of partition can be caused by predators by maintenance of low enough densities of competing species (Grover, 1994). Niche differentiation is then a shift from potential to realized niche, which may cause evolutionary changes afterwards. Niche differentiation has been widely used to explain community patterns both in field and experimental studies, mostly on relatively small scales. Indeed, neutral processes are believed to be of a greater importance at very large scales, but niche differentiation might still be at play (Tang and Zhou, 2011). An examples of field study is by (Kang et al., 2020), where niche differentiation caused by grazing implied various responses from species. For experimental examples, mixtures of several plants led to relative differences of height and leaf surface between species compared to monocultures (Zuppinger-Dingley et al., 2014). Closer to microorganisms, (Burson et al., 2019) experimentally highlighted the coexistence of phytoplankton species thanks to differential use of the underwater light spectrum, which can result from niche partitioning.

Habitat filtering works as the opposite by selecting the ecological traits that confer the best tolerance to a site, leading to a convergence in traits distributions among species (no differentiation). Such process might lead to competition between species, however competition might in return also lead to trait convergence, because only strongly competitive and ecologically equivalent species would remain (Zhang et al., 2017). It is in general hard to predict who between niche differentiation or habitat filtering shapes a community. However, we dispose of hints on spatial and temporal scales. For example, the study of spatial distribution of functional traits in

plants highlighted an higher effect of habitat filtering at small spatial scales, during early plant succession (Ulrich et al., 2017). Habitat filtering is likely to determine the niche occupancy, thus community structure of many plants communities worldwide, as demonstrated by (Li et al., 2018). Concerning microbial communities, (Yang et al., 2019) supposed habitat filtering in Chinese grassland in reason of numerous unique OTUs in different habitats, interpreted as habitats specialists.

In opposition to the niche theory, Hubell's neutral theory of biodiversity is a null hypothesis assuming that all species are ecologically equivalent and have equal rates of birth, death, immigration, emigration. A community's structure is then independent of species traits and only determined by stochastic processes (drift, dispersal, speciation) (Vellend et al., 2014). Both neutral and niche theories managed to explain community structure, depending on which pattern, time and spatial scales were considered.

Box 5 : An history of ecological niche theory (Pocheville, 2015)

Niche is a core concept in ecology and has been widely used to explain the assembly and coexistence of communities. It is roughly the description of a species's ecology. Grinnell (1917) first described the niche as every factor (abiotic or biotic) conditioning the existence of a species at a given location. Those can be for instance temperature, humidity, food, competitors, predators. Grinnell then got interested in ecologically equivalent organisms which are driven to occupy the same niche in different locations because of evolutionary convergence. Ecological equivalents brought Elton (1927) to define the niche as the position in trophic chains prior to habitat parameters.

In 1957, Posterior to the principle of competitive exclusion, Hutchinson formalized the niche as an attribute of species instead of the environment. Niche was then not the “niche occupied by a species” but the “niche of that species”, and was described as a space of multiple environmental variables, representing the limits of species viability, and delimiting the formed area as the fundamental niche. In contrast, the niche actually occupied, smaller because of restrictions caused by competitors, was named the realized niche. The Hutchinson shift is considered as a major step and was the start of quantification and predictive theories.

In 1960, Mac Arthur and Levins extended Hutchinson's approach with the concept of resource utilization distribution. Niche was then equivalent to the frequency of utilization of a resource by a population. Resources can either belong to food, space, or time variables. The vision of the niche furnished easy to measure items and got dramatically used, in what is known today as the niche theory, which essentially dealt with competition. It built the basis to study assembly rules and coexistence of communities, with models mostly based on Lotka-Volterra's equations.

The 1980 were a backslash, because research could converge towards a usable theory. Niche theory's validity started to be questioned by Simberloff, Strong, and others. Meanwhile, other factors such as environmental stresses, predation, or mutualism started to get a better consideration in the study of community assembly.

An ultimate revision was brought by Leibold and Chase, who defined the niche as the union of responses of an organism to various factors, as well as the organism's impact on factors. A typical example is a species in need of a resource in a given quantity, but its consumption depletes the resource in a proportional way. The rate of the resource renewal is then a major component.

2.3 Factors and events affecting microbial communities

2.3.1 External or abiotic factors

Various processes can affect a microbial community structure. Among them, wind dispersal and deposit (via aerosols and currents), as well as rain, are known to be relevant vectors affecting microbial communities, with a capacity to connect different microbial ecosystems at a local scale. Microbes are however differentially dispersed due to their life-strategies and morphological features (Womack et al., 2010; Griggs et al., 2021). As previously mentioned, physio-chemical properties of the biotope are determinant : climate, nutrients availability, pH, humidity, temperature (Zogg et al., 1997; Carrero-Colón et al., 2006; Drenovsky et al., 2010; Meron et al., 2012; Mello et al., 2016; Zhou et al., 2018; Chai et al., 2019; Cui et al., 2021). Landscape features, connectivity to other microbial sources, vectoring by animal hosts, neighboring plants and animals, are also multiple environmental factors shaping microbiota composition (Hacquard, 2016; Griggs et al., 2021).

Stochastic events such as the order of arrival (also called assembly history or priority effects) of microbial species in the system may also be at play. For example, manipulation of early immigration history in wood decomposer communities revealed differences in fungal species richness and composition (resulting in different carbon dynamics), most likely associated with different magnitudes in species interactions (Fukami et al., 2010). Echoing these results, Diamond (1975) observed that community composition varied among sites which seemed similar in several ways, thus suggesting that assembly history could lead to multiple stable equilibria. A single equilibrium is more likely to be reached in systems which have a small species pool, a high connectivity, a low productivity and frequent disturbances. Multiple equilibria are more likely to exist in the opposite scheme (Chase, 2003). The same microbe might then have facilities or trouble to establish in a system according to its timing of arrival, which echoes the barrier against pathogens offered by a microbiota to its host, a supplemental filter after host selection.

2.3.2 Selection by the host

2.3.2.1 Vertical transmission

In plants, microbes associated with the embryo and endosperm are more likely to be transmitted vertically than those associated with the seed coat. Despite the existence of a large cohort of studies, knowledge of implied mechanisms is scarce (Vandenkoornhuyse et al., 2015). Limits are similar in humans where, despite a lack of evidence in humans, there is support for vertical transfer (from mother to offspring) in animal models, but the colonization mechanisms are unclear (Walker et al., 2017).

2.3.2.2 Horizontal transmission and pseudo-vertical transmission

In plants, root colonization by soil microbes is in part deterministic and a host-controlled process. Bulk soil constitutes a reservoir of microbes in which variations in root morphology and exudates allow plants to actively recruit their rhizosphere microbiota in the surrounding soil (Bulgarelli et al., 2013; Griggs et al., 2021). The process of recruitment lead to a significant reduction in microbial diversity from soil to rhizosphere and endosphere, and is influenced by plant age, which exert a continuous selective pressure over time (Griggs et al., 2021). Furthermore, the colonization of the internal compartments of the plant could be an attractive goal because of reduced competition to reduced microbial diversity (Hartmann et al., 2009). The effect of host plant is not limited to the roots but extend to other compartments, which exert distinct selective pressure. Indeed, compartments of cultivated agaves were found to have a convergent microbiota's structure, independent to spatial distance (up to 2000km) (Coleman-Derr et al., 2016). Plants emit from their roots a large variety and quantities of organic exudates (carbohydrates, carboxylic acids, phenolic amino acids) and inorganic ions, according to plant physiological and developmental stage. These exudates have the capacity to condition the rhizospheric environment and therefore the rhizospheric microbial community composition. Furthermore, inhibitory (antimicrobial) or stimulatory (sugars) exudates modulate selective pressure (Hartmann et al., 2009). Because most of the seeds fall and germinate close to the mother plant, the plantlets come in contact with similar microorganisms than the mother. Thus associated to this short term seed dispersion has been hypothesized this pseudo-vertical transmission (Wilkinson and Sherratt, 2001).

The gut microbiota is known to be colonized during early life, although much more research has been conducted on bacteria than on archaea and eukaryotes. The initial microbiome colonization is crucial for the development of individuals. At birth, a subset of the maternal microbiota is supposed to be transferred, here again a pseudo-vertical transmission of microorganisms from the mother to the baby. Within a year (for humans), significant shifts in composition and abundances were highlighted, with strong fluctuations of the microeukaryotic community, as well as a diversification delay caused by cesarean-section and formula milk (Wampach et al., 2017).

2.3.3 Microbe-microbe interactions

2.3.3.1 An overview of ecological interactions

In comparison to abiotic factors, much less is known about how microbial interactions shape microbial communities (Nemergut et al., 2013). Determining how species interact is challenging : *in situ*, the observed behavior of a species is the integration of all its interactions with the other community members. Disentangling the nature of individual interactions is then hard. In addition,

when it comes to microbes, particularly host-associated ones, it becomes difficult to directly observe the community *in situ* (Logan et al., 2018). Ecological interactions between microbes are wide and range from mutualistic exchange of metabolic products to antagonistic secretion of antibiotics and direct predation (Coyte and Rakoff-Nahoum, 2019; Pacheco and Segrè, 2019).

The simplest form of an interaction is pairwised (two participants only) and is classified according to its outcome, which ranges from positive for both (mutualism, noted $+/+$) to negative for both (competition, $-/-$). Various outcomes lie in between, with combinations of positive, neutral, or negative outcomes such as ammensalism ($-/0$) or commensalism ($+/0$) (Figure 11, box 6, (Zélé et al., 2018). However, microbe-microbe interactions are far from being limited to pairwise-associations. Microbial communities incorporate numerous species of a high taxonomic diversity, characterized by a dense network of high-order (i.e. more than two species) interactions. It is also argued that ecological interactions is somehow incomplete and miss various and crucial nuances, and a enhancement of this framework is needed, including not only ecological outcome, but various attributes which are specificity, cost, contact dependencies, spatial and time dependencies, site, habitat, and compounds involved (Pacheco and Segrè, 2019).

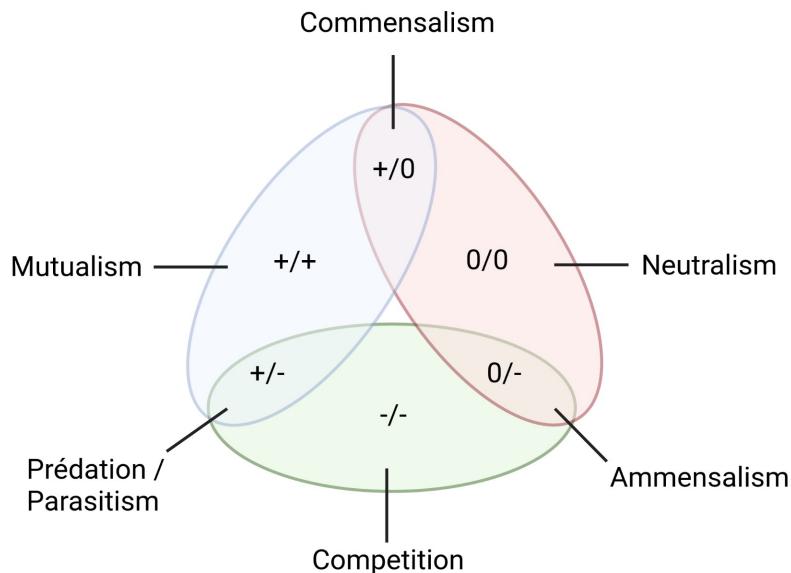


Figure 11: Ecological interactions and their outcomes for pairwised-species (+: beneficial, -: detrimental, 0: neutral). (after Zélé, Magalhães, Kéfi, & Duncan 2018, modified).

Box 6 : Ecological interactions (Coyte and Rakoff-Nahoum, 2019)

Competition occurs when species fight over a common and not sufficient resource(s) (of almost every conceivable kind), resulting in a negative outcome (-/-) for both. When a species is a stronger competitor than the other, the competition is defined as asymmetric. Depending on the intensity of the niches overlap, the weakest competitor disappears or adapts its realized niche to avoid competition (already mentioned on figure 10). Competition can either be indirect, for example via resource consumption (exploitative competition), or direct (interference competition), via fighting or secreting toxins, such as the type IV secretion systems in microbes.

Ammensalism is when one species has a behavior impacting negatively another species, without being affected in return (0/-). Ammensalism is often assimilated to asymmetric competition with a strong competitor, a typical example being the aforementioned toxin secretion capacities of some organisms (such as *Penicillium* secreting the well-known antibiotic, penicillin).

Predation is the action of predator consuming a prey, resulting in an asymmetric +/- outcome. For example, the bacteria *Bdellovibrio bacteriovorus* feeds on other bacteria.

Parasitism has a similar outcome to predation (+/-), but the parasite does not kill its host in order to feed itself. Instead, it might for example multiply inside it, like bacteriophages in bacteria. Depending on the parasite ecology, the host might be kept alive or killed after some time.

Mutualism (or cooperation) involves species supporting each other, resulting in positive outcomes for both (+/+). Most well-known mutualism cases involve organisms with widely different living requirements : plants and pollinators, plants and nitrogen-fixing microbes, lichens, corals... Cooperation is often costly, but mandatory to the species survival; returned benefits must then be higher than the costs to be evolutionary stable.

Commensalism is a situation where one species derives a benefit from another without affecting it in return. The outcome is then positive end neutral (+/0).

In the microbiome, metabolic dependencies (also often referred under the broader term “metabolic interactions”) are typically mutualistic and commensal i.e. bidirectional or unidirectional). They involve species growing on the exudates of another. Such interactions are called cross-feeding or syntrophies, and terms are often confused and used similarly. However, a clear classification has to be made and is described in section 2.3.3.5.

2.3.3.2 Time and spatial patterns affect existence, magnitude, and outcomes of ecological interactions

Interactions between organisms also depend on various attributes, space and time being among the most important. They must be taken into account to get more precise insights of an interactions, because they can involve unexpected mechanisms regarding the outcome (Pacheco and Segrè, 2019).

Time-dependent sharing of compounds or toxin secretion is known to modulate microbial interactions. For example, (Kelsic et al., 2015) modeled the coexistence of four antibiotic secreting and degrading bacteria to show how four organisms with varying degrees of antibiotic production. They could stably coexist in various temporal modes (stable equilibrium, limit cycles, or chaotic oscillations) without spatial separation (figure 12C).

Regarding space in the case of metabolic interactions, There are evidence suggesting its importance, despite relatively unknown mechanisms (Borer et al., 2020). The confinement of bacterial cells and their limited relocation capacities makes space critical for numerous processes (access to nutrients, metabolites exchanges, protection from pathogens or predators) (Stewart, 2003; Borer et al., 2018; Testa et al., 2019). Single species biofilms can engage cross-feeding between spatially distinct sub-populations, but cross-feeding interactions where one species depends on nutrients provided by surrounding other species imply a close proximity of partners. Such interaction leaves a signature in spatial organization of the assemblage (Dong and Fisher, 2019).

Species intermixing can also be strongly influenced by antimicrobial activities rather than cross-feeding interactions (figure 12B). Examples of such interactions involving the nuances brought by space coupled with antimicrobial activities can be mentioned. First, *Aggregatibacter actinomycetemcomitans* (*Aa*) and *Streptococcus gordonii* (*Sg*) are bacteria from the human oral cavity. *Sg* produces lactate, which is *Aa* preferred carbon source, but also secretes toxic hydrogen peroxide. *Aa* adapts then its dispersion at an optimal distance where it can assimilate lactate and detoxify hydrogen peroxide. Thus, spatial configuration results in a beneficial relation for *Aa* despite a clear antagonistic action from *Sg* (Stacy et al., 2014). In a second experiment, *E. coli* were exposed to a *Streptomyces* strain producing antibiotics. In a two species culture, the *E. coli* colonies grew outside the radius formed by this antagonistic *strain* but the addition of an antibiotic-degrading *Streptomyces* within the killing radius allowed them to grow in an area around this 3rd strain (figure 12A) (Kelsic et al., 2015).

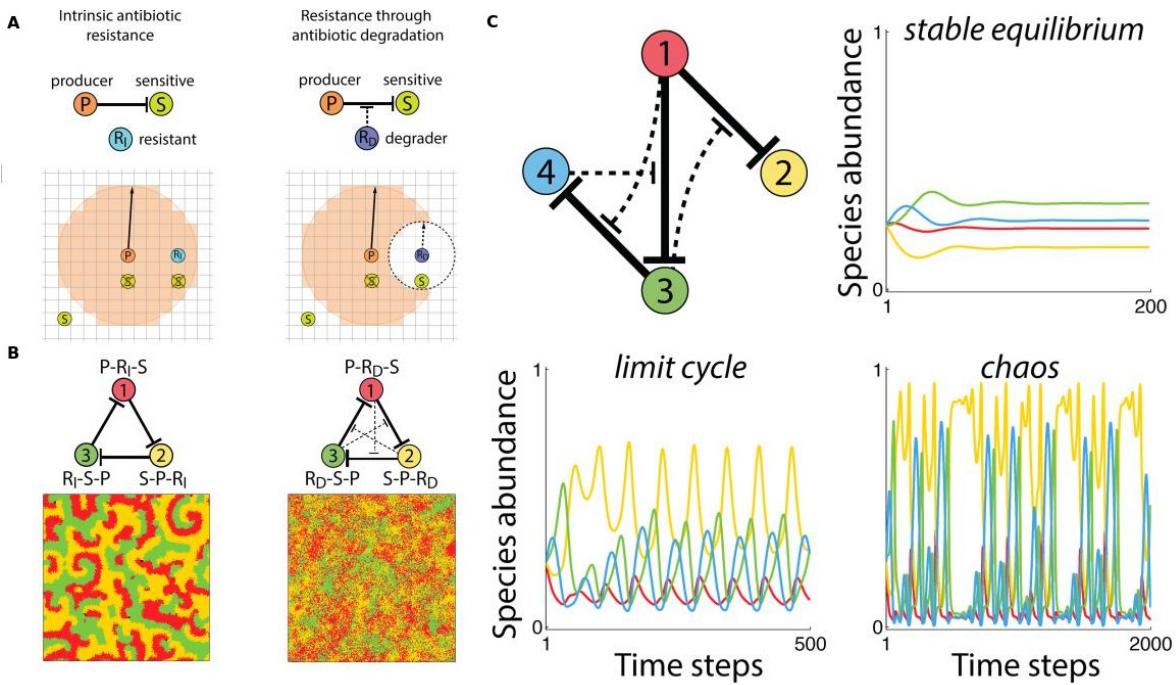


Figure 12: Illustration of the effects of space and time on microbial interactions. (A) An antibiotic-degrading bacteria can protect another bacteria from a 3rd bacteria's antibiotic secretions. (B) A three-species system with antibiotics secretion and degradation activities (right) display more spatial heterogeneity than a system with antibiotics secretion only (left). Such systems are also more durable in time. (C) Possible time patterns (abundances of four species) of a system with antibiotics secretion (from Kelsic et al. 2015, used with authorization).

Finally, microbes' spatial distribution and co-occurrences must be taken into account when looking for interactions. Most often, interactions are not known, but co-occurrences can be carried out from samples, and be used to select or discard putative interactions for further hypothesis testing and modeling. In the gut microbiota, some microbes were found to be differentially distributed in space (Welch et al., 2017). Such differences may weaken interactions by preventing the encounter of microbes. Moreover, in a spatially limited environment strong competition for space may bypass any cooperative effects (Coyte et al., 2015). This was highlighted in (Levy and Borenstein, 2013): putative interaction pairs were classified based on the co-occurrences of species. Microbes were on average more competitive against species with whom they commonly co-occurred compared to species with which they rarely associated, which might be a signature of habitat filtering, putting competition as a major driver of the microbial community.

2.3.3.3 Interactions existence, magnitude and outcome are environment-dependent

Interactions are environment-dependent (Pacheco and Segrè, 2019), a fact which is especially problematic when studying uncultivable organisms or when designing synthetic microbial communities. Indeed, some members may have metabolic dependencies that must be taken into account, or could be competing for the same nutrient. Other interactions might work differently according to the surrounding environment. Environmental factors are various, such as host species.

For instance, according to the host plant species, several bacterial taxa responded differently to the antagonistic activity of *Streptomyces* (Bakker et al., 2014).

An important environmental condition is the growth media which, according to its available nutrients, drives which compounds will be directly available and competed for. However they drive which compounds will be producible via microbes metabolism, later excreted and initiators of metabolic interactions (Heinken and Thiele, 2015; Magnúsdóttir et al., 2017). For example, when nutrients are limited, starved microbes compensate by engaging in behaviors that facilitate nutrient acquisition, notably by excreting molecules promoting cross-feeding (Fritts et al., 2021).

Interactions between organisms can dramatically change according to the environment, as demonstrated in (Zuñiga et al., 2019): the availability of nutrients (notably nitrogen) influenced the community stability by shifting members from cooperating (low nitrogen) with various metabolic exchanges to competing (high nitrogen). Environment-dependent metabolic interactions are more deeply discussed in chapter 2, regarding available nutrients, activatable metabolism, and compensatory metabolic interactions. In recent years, research accumulated evidences of metabolic interactions (also called metabolic dependencies) and started to question the importance of competition as a main driver of microbial communities (Pacheco and Segrè, 2019).

2.3.3.4 A wide set of methods to study microbial interactions

Metabolic interactions are far from being easy to decipher among a complex network of interacting microbes. There are various methods to find and quantify ecological interactions underlying the structure and dynamics of a microbiome. They range from *in vivo* analysis, with microbial abundances and co-occurrences obtained by specific DNA sequencing and quantifying, to *in silico* modelling of microbes metabolism or growth rates within time and environmental conditions. *In silico* methods have been intensively used for many advantages (data abundance, reduced costs...), under more and more system-biology-oriented approaches. Nevertheless, most of these methods are still vulnerable to mistakes via unmeasured external factors. Therefore, they are most often treated as hypothesis generators, from which the strongest predictions should be tested experimentally (Coyte and Rakoff-Nahoum, 2019).

3. Data and methods in Microbial ecology

Outlines :

Microbial ecology gained much more analysis power with the rise of all -omics technologies and their use in experimental and modelling approaches. In this short section, we give the basics of omics data, in order to properly introduce the second part of this general introduction, focused on metabolic cross-feeding and microbial system ecology (a system biology extension) used to disentangle metabolic interactions.

3.1 The -omics revolution

Microscopic observations and *in vitro* isolation were for long the only way to characterize microbes. Taxonomy and classification were thus limited, relying for instance on shapes or cell wall properties (i.e. gram-positive or gram-negative species), cellular organization, structure formed around plant roots. Microbial species and communities profiling made substantial progress alongside advances in molecular biology and DNA sequencing, until the maturation (initiated in 2005) of mass sequencing techniques, commonly called “next-generation sequencing” or “High-throughput DNA sequencing” (Heather and Chain, 2016). Such techniques allowed an immense jump ahead in microbial research, providing tremendous amount of “-omics” data and the information they carry (see box 7 and figure 13 from (Noecker et al., 2016), bypassing the limitations caused by many uncultivable (*in vitro*) organisms. For example, “shotgun metagenomics” is the untargeted (“shotgun”) sequencing of all (“meta”) microbial genomes in a sample. -Omics data can be used to get the taxonomic profile of a microbial community, its functional profile (by assigning sequences to physiological functions), or simply recover whole genome sequences. Another standard and widely used possibility is the identification of microbial species, thanks to genetic markers like high-throughput barcoding or 16S rRNA gene sequencing. In this case however, many species share the same markers and cannot be distinguished; they are then clustered in an “Operational Taxonomic Unit” (OTU), which share a fixed amount of sequence identity. Such abundance data can be for example used to study the dynamics of a community, under various environmental conditions or for competition/mutualism experiments. -Omics data also allow the reconstruction of metabolic pathways and networks, (Franzosa et al., 2015; Quince et al., 2017).

Box 7 : An non-exhaustive list of -omics data (Joyce and Palsson, 2006)

Genomics focus on the structure, function, evolution, mapping and editing of genomes, i.e. all DNA of an organism. After being sequenced, genomes are assembled, meaning all the sequence fragments are (computationally) aligned and merged in longer, contiguous sequences. Genomes are then annotated, which is the process to attach biological information to sequences, like gene prediction and protein-coding sequences. Annotation data are stored in various databases such as BIGG, AGORA, or MetaCyc., and can afterwards be used to annotate other organisms. The database choice matters : for example, BIGG references numerous strains of *E. coli* while AGORA records more general networks, and MetaCyc keeps only curated pathways of various organisms (Caspi et al., 2008; Jansma and Aidy, 2020). Genomics can be used for comparative studies (genome structure across species) or functional analysis (genes and proteins, often used with transcriptomics).

Transcriptomics is the sequencing of all RNA in a sample. It permits building a more precise functional profile of a community than with genomics, capturing a snapshot in time of genes expression, deciphering functions diversity, importance, and distribution among organisms. Transcriptomics can also reveal changes in functional activity in response to perturbations. As DNA, RNA can be sequenced with the so-called RNA-seq technologies. Metagenomics and metatranscriptomics do the same as the previous technologies, but for all species directly from a sample instead of one cultivated/isolated species.

Proteomics rely mostly on micro-arrays and mass spectrometry-based quantification of peptide mass and abundance in a sample, providing a more direct measure of protein activity than transcriptomics. Proteomics also permits the elucidation of protein structure, the detection of post-translational modifications, the discovery of new drugs for disease treatments, or the protein's interactions networks.

Metabolomics studies all compounds involved in cells or tissues metabolism (substrates, intermediates, products...), providing a snapshot of the physiological state. Exometabolomics emphasizes metabolites which are excreted in extracellular spaces. A major challenge in biology is currently to integrate this abundance of information from different sources into synthetic knowledge, to link the different levels of the systems (Vandenkoornhuyse et al., 2010; Franzosa et al., 2015).

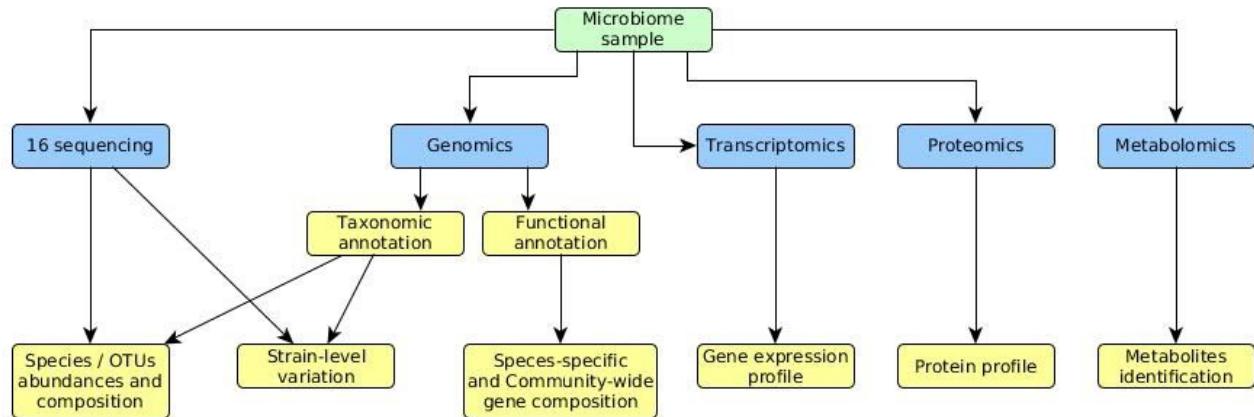


Figure 13: Schemes of microbiome analysis (composition, dynamics, and functions) with -omics data. Each -omic type targets different molecular levels (DNA, RNA, protein, metabolites ...) and has its own use. Depending on the research question, -omics data are preferred to others, but research is currently going towards the integration of all the information brought by multi-omics data (after Noecker, McNally, Eng, & Borenstein, 2016, modified).

3.2 Metabolic networks and system biology are the basis to study metabolic interactions

Metabolism is the set of chemical reactions taking place within an organism (Nielsen, 2017). Metabolism is composed of metabolic pathways (figure 14), structured as an intricate metabolic network (figure 15). Each pathway starts with initial reactants (nutrients from the environment, or products or byproducts of previous pathways) to produce final compounds, further used by the organism to maintain its balance and grow (Shah et al., 2021). Metabolic networks (also called genome-scale metabolic models, or GEM) are reconstructed from genomes' annotation, and incorporate all reactions, enzymes, compounds found in all recorded metabolic pathways, whether they are complete or not. Roughly, analyzing micro-organisms metabolic networks allow to predict their function, their homeostasis (stability and regulation), as well as their growth requirements and the compounds they can and cannot produce, hence potentially exchange to or receive from others under metabolic interactions processes (Larhlimi et al., 2011; Watson et al., 2015; Çakır et al., 2020).

Title: Serine biosynthesis
Organism: *Saccharomyces cerevisiae*

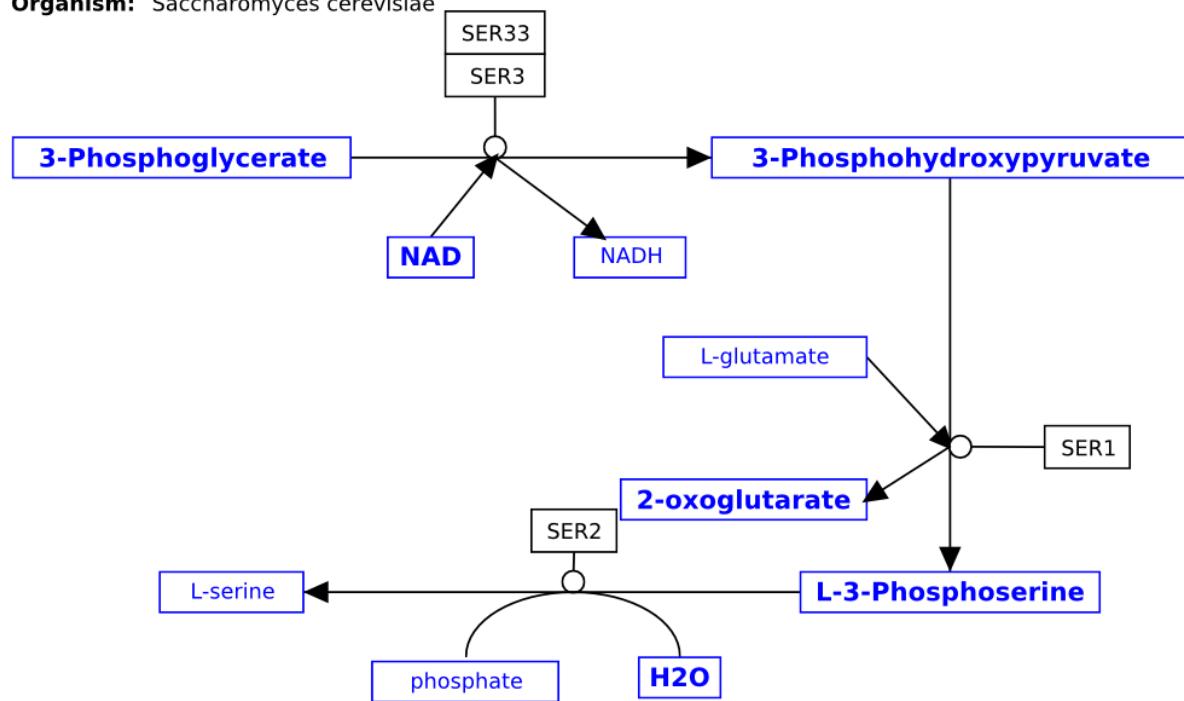
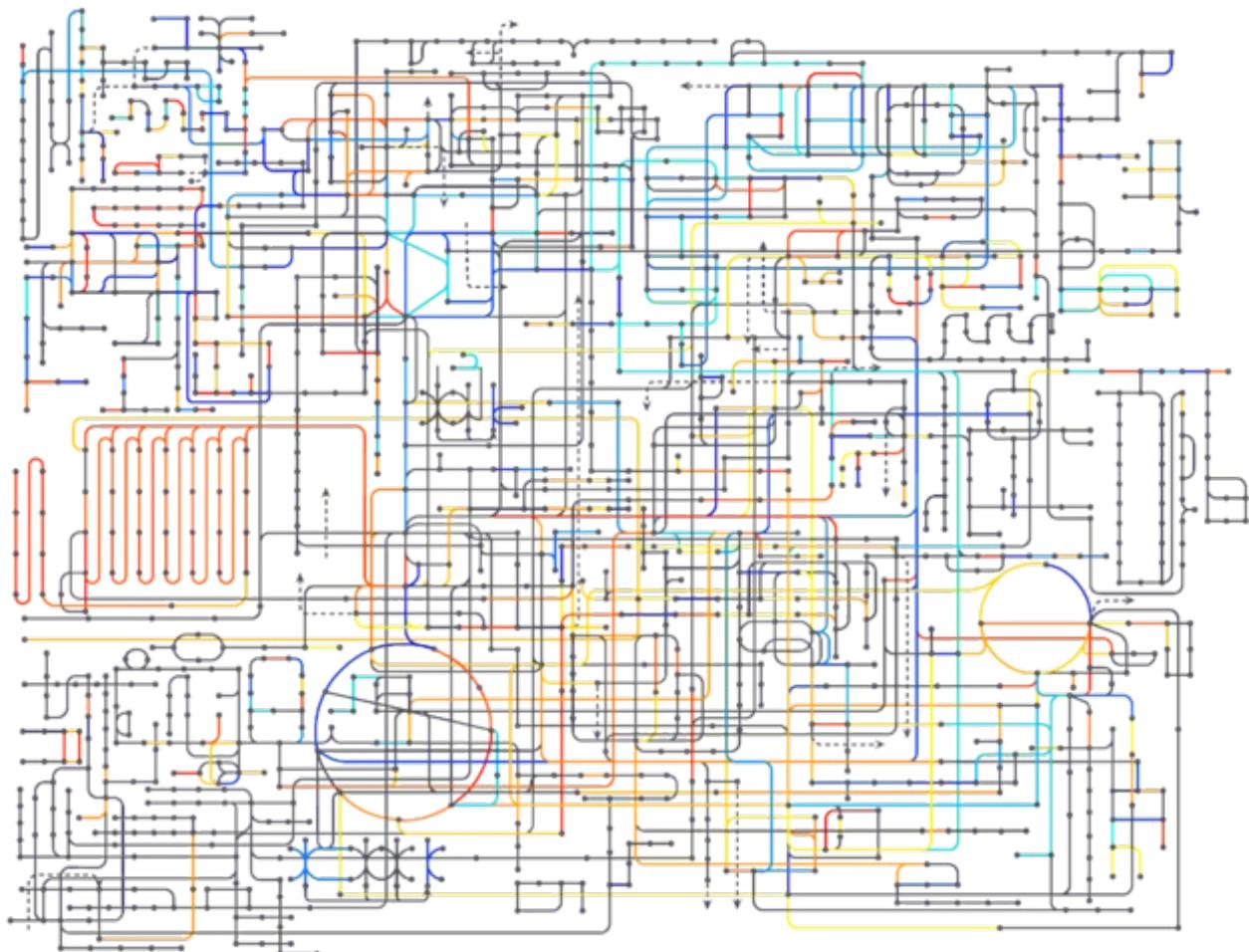


Figure 14: An example of metabolic pathway, here for serine biosynthesis in *Saccharomyces cerevisiae*. Chemical reactions are black arrows, reactants and products are blue squared. Involved genes are black squares (Creative Commons figure, from wikipathways.org).

In silico metabolic analysis is made possible with the use of computational and mathematical modeling, which are the basis of system biology. In systems biology, biological entities (molecules, cells, organisms, etc.) are seen as the connected components of networks, interacting together. The whole complex is then referred to as the “system”, and models attempt to explain, simulate and predict its behavior. Reconstructed metabolic networks are among the dominant approaches in system biology (Watson et al., 2015; Nielsen, 2017). In recent years, large volumes of data generated by -omics approaches furnished sufficient quantitative data to support massively such *in silico* system-based research, mainly on model organisms : human, mouse, yeast (Shahzad and Loor, 2012), or the brown algae *Ectocarpus siliculosus* (Tonon et al., 2011). However, omics approaches have methodological limits (incomplete genomes’ annotation, parameterization issues, statistical issues, computational limits (etc.)) (Vandenkoornhuyse et al., 2010) and large-scale models and networks do not necessarily allow the desired holistic, deep understanding of biological phenomena but rather give emergent properties sticking to a global overview. A combination of top-down and bottom-ups approaches is then necessary, notably for microbiome engineering (Lawson et al., 2019). Such combinations of approaches are shared by many. For instance, Vrancken et al., (2019) and Vandenkoornhuyse et al., (2010) stated getting a holistic and

mechanistic view of a microbiota would be extremely difficult in complex, natural communities, which is however the purpose of ecology. In Vandenkoornhuyse et al., (2015), the authors encourage the use of system biology and environmental -omics data with manipulation of small microbial communities to elucidate the interactions involved within the holobiont. Microbial systems ecology is then defined as the holistic study of microbial communities using systems biology approaches (Muller et al., 2018), and the study of microbial communities embraces a more systems-oriented approach, notably regarding metabolic interactions (see general introduction, part II) (Pacheco and Segre, 2019).



*Figure 15: A schematic part of the metabolic network of *Homo sapiens*. Each colored line represents an essential metabolic pathway, from seed reactants to final metabolites. Some pathways can cross each other (Creative Commons figure, Wikimedia Commons).*

Objectives of the PhD

Microbial communities have a great importance in multiple ecological processes, but the mechanisms underlying their structure, dynamics, and functions remain largely unknown. However since the rise of -omics techniques, tremendous amounts of data were collected and opened multiple research areas and directions. The early steps of microbial ecology were oriented towards the development of descriptive methods, which allowed to decipher spatial, time, host, genetic (...) patterns, similarity and dissimilarity between and within microbial communities. Microbial ecology still lacks a strong theoretical background but is now more and more structuring itself in order to move towards mechanistic research. Mechanistic approaches aim to decipher the inner structure of complex co-associations and ecological interactions networks of microbes giving emergent properties to microbial communities. Such organization goes notably with the maturation of microbial system ecology, which intensively use modelling of integrated -omics data, the principal framework being a constant shift between top-down and bottom-up designs in a design-built-test-learn cycle.

This PhD aimed to study metabolic dependencies of the root microbiota of *Arabidopsis thaliana*, based on annotated genomes and reconstructed metabolic networks of 193 isolated bacteria, picked in order to be representative of the host-plant microbiota.

A primary goal was to provide a framework and entry door to microbial system ecology, with a mini-review (general introduction, part II, submitted to Frontiers in Microbiology) focused on an overview of methods used to predict metabolic dependencies within microbes. The importance of metabolic dependencies is further discussed, as well as the nature of microbial ecology, which allows a crossroads between community ecology and system biology.

Then, a major work was dedicated to environment-dependent interactions (chapter 2, submitted to Microbiome). With a generic ambition, we explored with metabolic modelling the effect on bacteria's metabolism of growth constraints exerted by several rich and poor growth media. In addition, the effects of the addition of artificial root exudates in each medium was considered. Metabolic dependencies were studied with several angles. First (1), metrics dedicated to find putative metabolic dependencies were explored. Then (2), we considered the whole community and every possible exchangeable compound, and finally (3), minimal combinations of genomes able to produce targeted compounds (amino-acids, vitamins, and phytohormones).

This modelling approach opened a door to an experimental approach, in the aim to confront the previous metrics to the outcomes of bacterial co-cultures (chapter 3, draft to be submitted). Several SynComs of bacteria quadruplets were cultivated, and ecological interactions were inferred regarding the abundances of strains compared to their monoculture. *In fine*, competition seemed to dominate most Syncoms, but some Syncoms of interest were replicated afterwards in order to investigate further a potential exploitation of bacterial exudates by an *Achromobacter*. We also identified several methodological bias that could have masked the effects of cross-feeding.

Finally, we discuss the main interpretations of these results in terms of microbiota structure and dynamics via metabolic dependencies, the integration of these two different approaches, and on short and long term perspectives.

Part II. Metabolic cross-feeding with Microbial System Ecology to disentangle coexistence in microbiomes: A mini-review

In the previous part, we detailed all the main assembly rules of microbial communities, from abiotic to biotic factors, i.e. ecological interactions. Among them, microbe-microbe metabolic dependencies are more and more thought to be essential drivers of microbiotas structure and dynamics, unsettling the historical importance given to antagonistic interactions like competition. Such advances in research were allowed, mostly, by the arrival of Next-Generation Sequencing and integration of system biology in microbiology.

In this part, written as a mini-review (submitted to *Frontiers in Microbiology*), we first remind the importance of metabolic dependencies within a microbial community. We then provide details on how System biology is applied to microbiology, in a framework named Microbial System Ecology. Such framework aim to (1) reconcile modelling and experimental approach, which are used respectively for prediction and validation and (2) conjugate top-down and bottom-up approaches, used respectively to obtain an overview (mainly descriptive) of a community leading to more precise hypothesis, and on the other hand to get a mechanistic understanding of a particular process.

Microbial System Ecology to understand metabolic dependencies in microbiomes: A Mini-review

Victor Mataigne^{1,2}, Nathan Vannier², Philippe Vandenkoornhuyse^{1,*}, Stéphane Hacquard^{2,*}

1 Université de Rennes 1, CNRS, UMR6553 ECOBIO, Campus Beaulieu, 35000 Rennes, France

2 Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany

* Correspondance

Submitted to Frontiers in Microbiology

Abstract

Understanding how microorganism-microorganism interactions shape microbial assemblages is a key to deciphering the evolution of dependencies and co-existence in complex microbiomes. Metabolic dependencies in cross-feeding exist in microbial communities and can at least partially determine microbial community composition. To parry the complexity and experimental limitations caused by the tremendous number of possible interactions, new concepts from systems biology aim to decipher how the components of a system interact with each other to better connect the components of the system. The idea that cross-feeding does impact microbiome assemblages has developed both theoretically and empirically, following a systems biology framework applied to microbial systems, formalized as microbial systems ecology and relying on integrated -omics data. This framework merges cellular and community scales and offers new avenues to untangle microbial coexistence primarily by metabolic modeling, one of the main approaches used for mechanistic studies. In this mini-review, we first give a concise explanation of microbial cross-feeding. We then discuss how microbial systems ecology can enable progress in microbial research. Finally, we provide an overview of a microbial systems ecology framework mostly based on genome-scale metabolic-network reconstruction that combines top-down and bottom-up approaches to assess the molecular mechanisms of deterministic processes of microbial community assembly, that is particularly suitable for use in synthetic biology and microbiome engineering.

1 Introduction

Deciphering the assembly rules of microbial communities is vital for a mechanistic understanding of the general principles driving microbiome activity and functions (Vellend et al., 2014; Morrison-Whittle and Goddard, 2015). Microbial communities are governed by both stochastic and deterministic factors (Vellend, 2010; Stegen et al., 2012), and recent advances show

that deterministic processes largely contribute to shaping microbial community assembly (Ning et al., 2020; Xu et al., 2020). Ecological interactions including commensalism, competition, and mutualism contribute to the self-organizational properties of microbiomes (Stegen et al., 2013). However, how these different interactions act in concert to shape microbial assemblages remains poorly understood (Nemergut et al., 2013). Microbial communities are likely not only driven by antagonistic interactions but also by non-antagonistic symbioses, defined in 1879 by De Bary as the “living together of unlike organisms”, which are now recognized as central drivers of (co-)evolution. Symbiosis is often associated with obligate mutualism but is actually a continuum of interactions between mutualism and parasitism (Ewald, 1987; Drew et al., 2021), all implying dependency of one organism on another (figure 1A) (Raina et al., 2018). Among these interactions, metabolic dependencies by cross-feeding likely explain patterns in microbial communities, which can be seen as an intricate community of symbionts (Mas et al., 2016; Zomorrodi and Segrè, 2017; Amor and Bello, 2019; Coyte and Rakoff-Nahoum, 2019; Pacheco and Segrè, 2019; Seif et al., 2020; Zhu et al., 2020). Indeed, in community ecology, competition and related Gause’s competitive exclusion were previously considered to be the main drivers of community assembly but this role was questioned by observation of the unexpectedly complex microbial communities according to general ecology theories (Pacheco and Segrè, 2019). In parallel with other eco-evolutionary factors like viruses, which can maintain microorganism community richness according to, for instance, the ‘killing the winner hypothesis’ (Winter et al., 2010), cross-feeding is increasingly believed to play an important role in the complexity of microorganism communities (Zengler and Zaramela, 2018).

2 Metabolic cross-feeding as a major driver of microbiota assemblages

Microbial cross-feeding refers to the interaction between microorganisms in which molecules resulting from the metabolism of one microorganism are further metabolized by another (figure 1B&E) (Smith et al., 2019). Currently, it is not known if microbial cross-feeding is specific to particular compounds, environmental constraints, or species, limited to pairs of interacting organisms, or extended to several symbionts, or optional or obligatory for the survival of the microorganisms (Zengler and Zaramela, 2018). Also, different types of cross-feeding are recognized depending on whether they are unidirectional (one microorganism benefits from another) or bidirectional (both microorganisms benefit from each other’s secretions) or depending on which compounds are exchanged (figure 1B&E). A closely associated term is syntrophy, which also defines the consumption of an organism’s secretion by an auxotrophic organism (Smith et al.,

2019). The definition varies from an obligatory mutualistic metabolism (Morris et al., 2013) to optional (Hillesland, 2018). The compounds involved also vary, and are sometimes restricted to waste products (Oliveira et al., 2014) sometimes not (Stams and Plugge, 2009; Pande and Kost, 2017).

One example of known mutual cross-feeding is between *Rhodococcus ruber* and *Bacillus cereus*. *R. ruber* degrades a tetrahydrofuran, which results in acidic metabolites that are taken up by *B. cereus*, which, in return, regulates pH and secretes nutrients that are essential for *R. ruber* (Liu et al., 2019). Less specific cross-feeding can also occur. For instance, *Akkermansia muciniphila* degrades and ferments its host's mucus, leading to the production of oligosaccharides that are available for other microorganisms (Belzer et al., 2017). A hierarchy in the importance of microorganisms for the microbiota stability has also been demonstrated in relation to cross-feeding (Gutiérrez and Garrido, 2019) using a species-deletion approach in a consortium of 14 bacteria. Cross-feeding can also enable degradation of chains of complex molecules (Vet et al., 2020).

One key process is extracellular secretion of a wide range of “public goods”, including enzymes, proteins, byproducts, waste, co-factors, amino-acids, and vitamins. They benefit all the organisms in the community that are able to assimilate them (Croft et al., 2005; Yu et al., 2009; Seth and Taga, 2014; Rodionova et al., 2015; Cavaliere et al., 2017; Zengler and Zaramela, 2018; Fritts et al., 2021). Many microorganisms are auxotrophic for various metabolites, lack essential pathways or genes, and thus rely on extracellular sources (Mee et al., 2014), but still possess enzymes whose activity depends on these metabolites (Degnan et al., 2014).

Trying to discover if and how a microorganism secretes or assimilates a given extracellular compound is challenging (figure 1C) (Sung et al., 2017; D’Souza et al., 2018; Zengler and Zaramela, 2018). Moreover, ecological interactions are affected by temporal and spatial patterns (Kelsic et al., 2015), and by the organisms’ surrounding environment (Bakker et al., 2014). Notably, available nutrients control the metabolic activity of microorganisms, whether or not they depend on others (figure 1D) (Heinken and Thiele, 2015; Magnúsdóttir et al., 2017). For example, when nutrients are limited, microorganisms can compensate by engaging in behaviors that facilitate nutrient acquisition, notably by excreting molecules that promote cross-feeding (Fritts et al., 2021), even if they usually compete (Zuñiga et al., 2019). Another example of an environmental effect involves two mutants of *Pseudomonas stutzeri*. Depending on the pH, the mutants can shift from competition to strong cross-feeding of nitrite, which is a toxic compound at low pH (Borer et al., 2020).

2.1 Evolution and stability of cross-feeding

Energy saving could account for the origin of frequent occurrences of auxotrophy in microorganisms. First, the cost of producing certain metabolites is avoided by obtaining them from the environment (Zengler and Zaramela, 2018). Second, mutual cross-feeding has been shown to reduce the energetic cost of some metabolic pathways, for example amino-acids biosynthesis (Mee et al., 2014). Metabolic exchanges are thus a way to divide the cost of labor (Thommes et al., 2019). However it has been predicted that costless secretions may be numerous and represent sources of cross-feeding opportunities (Pacheco et al., 2019). Gene loss is the cause of auxotrophy, which may arise when a costly function can be performed by one or more other members of the community, and may cause selective pressure (Boon et al., 2014; D’Souza et al., 2014; Mas et al., 2016; Meijer et al., 2020).

Evolution of beneficiaries and providers via gene loss and public goods have been formalized in the Black Queen Hypothesis (BQH) (Morris et al., 2012a). This hypothesis relies on the fact that many vital genetic functions are leaky, thereby producing public goods for the community. BQH involves costly functions, so the benefits of losing it and relying on others are greater than those of than keeping it. Providers are hypothesized to be organisms that lost the race to profit from leakiness and are stuck in their role of providers (Morris, 2015). The BQH of evolution also explains an evolutionary trajectory of competition avoidance toward a steady-state equilibrium for the coexistence of microorganisms (Mas et al., 2016). Wastes are hypothesized to be detoxified by other organisms (Cavaliere et al., 2017). Selection via gene-loss would have then generated numerous commensal and even mutualistic dependencies between different microorganisms, leading to high species richness in communities but requiring a certain degree of stability (Boon et al., 2014).

Cooperative behaviors are considered to be unstable because of the constant threat of the emergence of organisms that benefit from the cooperative interactions but do not contribute to them (Cavaliere et al., 2017). Despite hypotheses like the BQH, it is still not clear how such interactions occur and are maintained (Zomorrodi and Segrè, 2017). When the auxotrophic organism is a mutant derived from a given microorganism population that relies on its conspecifics or symbionts without contributing in return, the dependency is classified as “cheating” and can be detrimental to non-cheating organisms (Ferriere et al., 2002). In plant-microorganism symbioses, it has been demonstrated that plants are able to sanction less cooperative microorganisms with a carbon embargo (Kiers et al., 2011). Research involving game theory, notably the prisoner’s dilemma,

helped explain these evolutionary and stability issues (Gore et al., 2009; Zomorrodi and Segrè, 2017).

2.2 The growing importance of metabolic cross-feeding compared to competition

Experimental results suggested that microbiota are dominated by competition (Coyte and Rakoff-Nahoum, 2019). However, results vary and although some studies suggest that microorganism communities are governed by antagonistic interactions and rarely cooperate (Biggs et al., 2017; Venturelli et al., 2018), others revealed rich networks of metabolic dependency interactions among microorganisms (Medlock et al., 2018). However, few true interspecies cooperation has been validated to date (Coyte and Rakoff-Nahoum, 2019). Niche differentiation and metabolic dissimilarity between co-occurring microorganisms could be explained by complementary biosynthetic capabilities rather than by competitive exclusion (Zelezniak et al., 2015). The fact that several bacterial taxa cannot be grown alone *in vitro* could result from such dependencies (Mas et al., 2016). Interestingly, it was observed that in microbial communities spread along a competitive-cooperative axis, the most competitive microorganisms were characterized by larger genomes and were mainly present in soil, while the most cooperative had smaller genomes and were present in both free-living and host-associated habitats (Machado et al., 2021). Many communities seemed to be engaged in a trade-off between competition and cooperation, echoing the trade-off between being metabolically independent and relying on others (Thommes et al., 2019).

Deciphering microbial interactions is a major challenge in microbiome research to enable the shift from descriptive approaches to a mechanistic understanding of microbiome assemblages. Such complex systems involving hundreds of interacting organisms make it difficult to determine which interactions primarily drive community stability or modulate shifts in assembly trajectories. In the following sections, we discuss the potential of microbial systems ecology (MSE) to disentangle the mechanisms of cooperation and co-existence in a microbiome. Microbial Systems Ecology: a crossroads between system biology and community ecology.

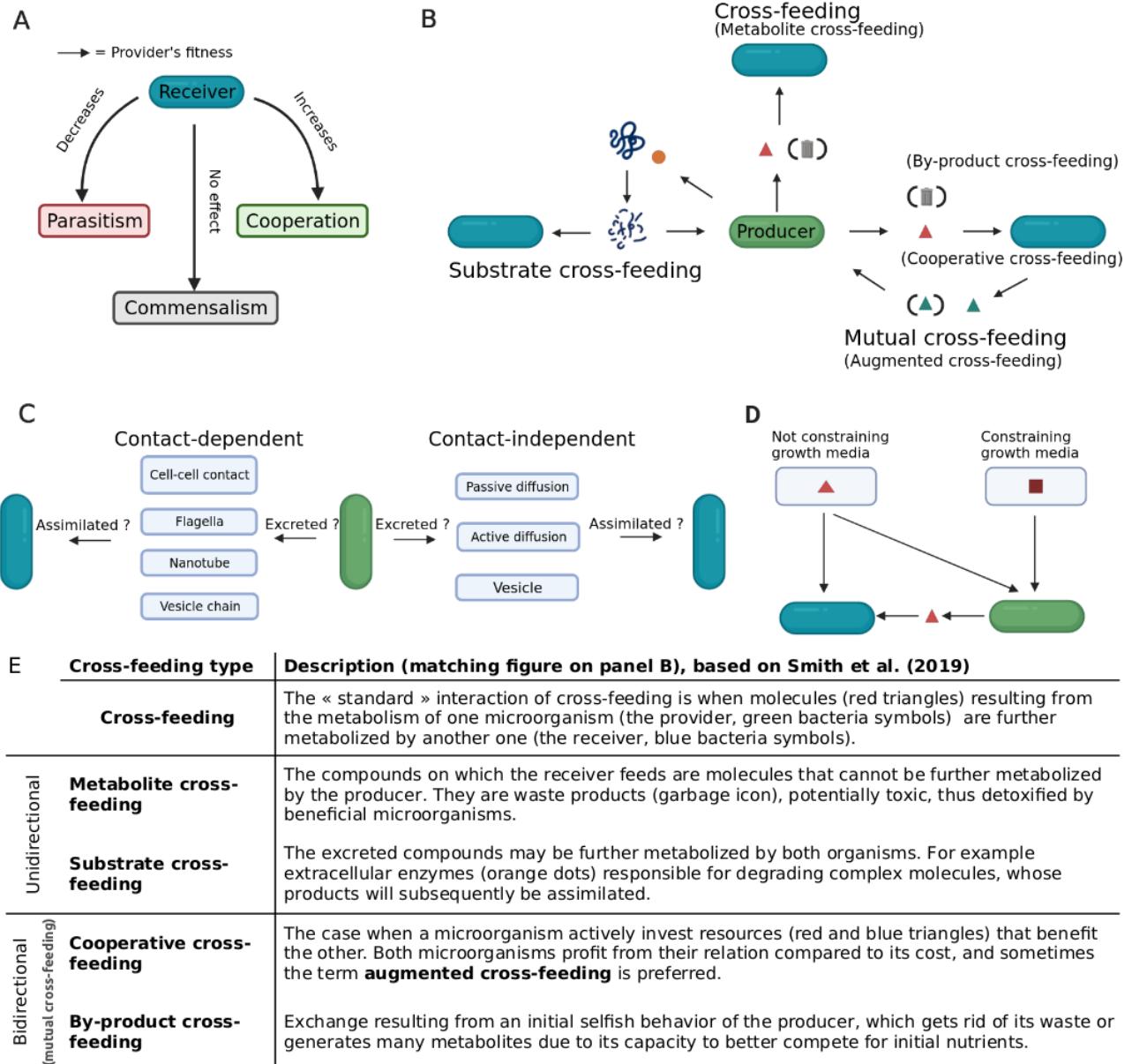


Figure 1: (A) Symbiosis (and hence cross-feeding) is not necessarily mutualistic. Depending on the effect of the receiver (blue bacteria symbols) on the provider's fitness, the interaction can go from cooperation to parasitism. (B) There are several subcategories of cross-feeding (Smith et al. 2019), depending on which compounds are exchanged and on the directionality of the exchange (mutual or not). (C) The existence of cross-feeding depends on the secretion, transport, and assimilation capacity of the public good (D'Souza et al. 2018). (D) Metabolic interactions are environment-dependent, notably regarding available nutrients. If a required nutrient (red triangle) is freely available in the growth medium, then cross-feeding is not indispensable for the receiver organism. Otherwise, when this particular nutrient is not available, but is synthesized by the producer from another nutrient (brown square), cross-feeding becomes obligatory for the receiver. (E) Cross-feeding categories are defined in panel B.

3 Microbial Systems Ecology: a crossroads between system biology and community ecology

Systems biology is the computational and mathematical study of interactions between the components of biological entities (molecules, cells, organs, organisms), considered as complex systems (Snoep and Westerhoff, 2005). Connecting components is preferred to characterizing isolated parts (Kitano, 2002b), because the latter are not sufficient to understand the behavior of the system as a whole. System biology involves a cycle of theory, modeling, testing hypotheses, followed by experimental validation. In addition to the structure of the system (gene interactions, biochemical pathways, etc.), biological systems must integrate dynamics and robustness components, i.e., how they behave over time under varying conditions, as well as their sensitivity to perturbations (Kitano, 2002b; Alon, 2006). Omics approaches produce sufficient quantitative data to support simulation-based research, leading to genome-scale modeling to analyze the cell function properties of the system, mainly based on graph theory (Kitano, 2002a; Rodriguez et al., 2019). Research includes reconstruction of metabolic networks, transcriptional regulatory networks, interactome networks, hormone signaling (etc.), for various applications including crop protection or sustainable agriculture, therapies for obesity, diabetes, and inflammatory bowel disease, or conservation biology (Amor and Bello, 2019; Rodriguez et al., 2019; Vázquez-Castellanos et al., 2019).

MSE is defined as the holistic study of microbial communities using systems biology (Muller et al., 2018). In microbiology, the cellular and the individual levels are often intertwined: the components of the system are cells and are also individuals of different microbial species and components of the community, creating an ideal crossroads which, in systems biology, is used to analyze populations and communities. MSE includes numerous approaches to study assembly rules, co-existence, trophic networks (etc.) in microbial communities. Such communities are seen as networks of networks: i.e., community members consisting of collections of interwoven molecular networks (Muller et al., 2018). MSE is mainly based on the construction of predictive models using a corpus of computational methods that make it possible to mine large amounts of data, notably to predict putative interactions or phenotypes under different growth conditions (Franzosa et al., 2015; Bordron et al., 2016). Nevertheless, most of these methods are still vulnerable to mistakes via unmeasured external factors. They are therefore often treated as hypothesis generators, of which the strongest have to be tested experimentally (Coyte and Rakoff-Nahoum, 2019; Goyal et al., 2021).

Considerable efforts have been made to analyze and model microbiomes and predict microbial interactions (Li et al., 2016; Knight et al., 2018; Kumar et al., 2019) notably based on predicted metabolism by identifying keystone genes and functions and by identifying the microorganisms' ecological niches. Genome-encoded metabolism can reveal fundamental niches while resource usage, realized niches, and their overlaps between species can be inferred from transcriptomes, proteomes, and metabolomes (Mee et al., 2014; Muller et al., 2018) thus making it possible to identify core and specific metabolism or to predict metabolic interactions. For example, in a set of five bacteria, such models found that species-specific metabolism is related to secondary metabolism, and metabolic interactions were required to perform copper bioleaching (Bordron et al., 2016). Despite the immense and untapped potential of -omics, niche inference remains a challenging task due to niche multi-dimensionality, the complexity of trophic interactions, and fluctuating environmental conditions (Muller et al., 2018).

3.1 Microbial Systems Ecology approaches and framework

3.1.1 Metabolic network reconstruction

Once an organism's genome has been sequenced and annotated, its metabolic network can be inferred (Mendoza et al., 2019). Metabolic networks are often referred to as “genome-scale metabolic models” (GEMS) gathering all the metabolic capacities of an organism, linking chemical reactions, reactants, products, and enzymes needed to reconstruct metabolic pathways (Jansma and Aidy, 2020). GEMs can predict cell behavior under various conditions (notably nutritional): which metabolic functions organisms are capable of achieving, which compounds can be produced, or what are the growth requirements of a particular network. However, the main limit is that GEMs are mainly drafts, and their reliability depends to a great extent on how well annotated the organism concerned already is. This applies to only a few dozen well-known organisms including humans, the mouse, *Arabidopsis thaliana*, yeast, some bacteria (Shahzad and Loor, 2012). Under-investigated organisms produce more general GEMs, because specific genes are less annotated, resulting in gaps or incomplete pathways, which is problematic when attempting to establish precise functional profiles (Jansma and Aidy, 2020). Indeed, it has been demonstrated that many GEMs are limited to well-conserved, primary metabolic pathways rather than secondary metabolic pathways, thus limiting the representation of the organisms they model (Monk et al., 2014). Such problems can be overcome with additional steps like gap-filling and manual curation (Prigent et al., 2017), but these are subject to false positives when working with unknown organisms (Henry et al., 2010; Frioux et al., 2020).

When data on stoichiometric reactions are available, metabolic networks can be enhanced through quantitative analysis of metabolite fluxes within the network. After considering available nutrients, fluxes of metabolites within and between pathways are computed to maximize an objective function, such as biomass production. A standard approach is flux balance analysis (Bordbar et al., 2014). However, the objective function is often difficult to define, and such methods require high-quality GEMs. What is more, they still only provide a static view of the community; models that incorporate metabolic modelling, dynamics of species abundance and of concentrations of metabolites over time are an active field of development (Muller et al., 2018; Vrancken et al., 2019). For a recent review of computational tools dedicated to the reconstruction and analysis of metabolic networks, we recommend the one by Mendoza et al., (2019).

Simulations of GEMs under environmental constraints are used to identify potential competition for nutrients, to predict cross-feeding or ecological niches, with applications in metabolic engineering (Heinken and Thiele, 2015; Magnúsdóttir et al., 2017; Frioux et al., 2018; Muller et al., 2018; Mendoza et al., 2019). These approaches enable more direct quantification of interactions than techniques that rely on natural communities *in vivo*. However, problems increase with the number of species studied simultaneously, and precise metabolic modeling rapidly becomes impractical for natural communities because of the tremendous number of possible configurations (Coyte and Rakoff-Nahoum, 2019). Specific approaches are dedicated to considering multiple species at once, for example, community flux balance analysis (Khandelwal et al., 2013) but generally speaking, existing methodological limits (Vandenkoornhuyse et al., 2010) do not produce the necessary holistic understanding of microbiota but rather give an only slightly more than general overview of emergent properties, or are limited to a small fraction of a community. Metabolic networks are thus mostly used in bottom-up (reductionist) approaches (Shahzad and Loor, 2012), but it is important to also take top-down (global) approaches or combinations of both into consideration (Lawson et al., 2019) (figure 2).

3.1.2 The microbial systems ecology framework calls for shifts between top-down and bottom-up approaches

In MSE, the study of complex systems like microbiomes uses both top-down and bottom-up approaches within a design-build-test-learn process that is particularly suitable for microbiome engineering and synthetic biology (figure 2), where the optimum and minimum combinations of organisms are investigated in order to perform a biological function (for an exhaustive explanation and review, see Lawson et al., (2019)). Such a process works in cycles, where the design and build phases are adapted to the functions targeted, and the test and learn phases are used to correct any errors and to optimize the system.

Top-down approaches start from a complete microbial community (or at least a sufficiently big and representative set of microorganisms) and aim to discover signature patterns of underlying biological mechanisms (figure 2 A&B). Top-down approaches are basically descriptive and were developed using many multivariate statistics, meta-omics, and experimental data, to capture key microbiome functions or effects or particular environmental variables rather than prioritizing which organism or pathway is at play behind an observed phenotype (Ramette, 2007; Shahzad and Loor, 2012; Lawson et al., 2019). Most of our knowledge about the gut microbiome was obtained using top-down approaches and helped discern dysbiosis patterns associated with diseases (Bashan et al., 2016; Amor and Bello, 2019). For example, one method involves clustering the members of a community according to their metabolic functions, and/or building co-occurrence networks to identify coexistences and to propose hypotheses to explain the origin of the coexistence (Faust and Raes, 2012; Layeghifard et al., 2017). To achieve that goal, simple metrics computed from metabolic networks are used to compute metabolic overlap, metabolic interaction potential, or the functional distance between organisms (Zelezniak et al., 2015; Russel et al., 2017), leading to the formulation of hypotheses about ecological processes involved, including metabolic interactions. Top-down approaches offer a macro-scale framework but miss intricate details, consequently limiting both holistic and mechanistic views of complex (i.e. natural) communities (Vrancken et al., 2019).

This limit is offset by bottom-up approaches that ignore the whole system and start from single microorganisms to build simple sub-communities to deduce the functional properties that could emerge from a small subsystem, and then gradually increase model complexity (figure 2 C&D) (Amor and Bello, 2019; Lawson et al., 2019). Bottom-up approaches use proficient computational and mathematical modeling (for details see Vrancken et al., (2019)), notably based on GEMs, for

example with constraint-based analytics able to directly identify combinations of GEMs able to produce a compound that cannot be produced by single genomes (Frioux et al., 2018). Such approaches have (for example) been used to predict mutualism and competition in relatively big microbial consortia (Friedman et al., 2017; Kong et al., 2018). Overall, cooperative interactions (including metabolic dependencies) are often key components of bottom-up designs in synthetic biology (Amor and Bello, 2019) and in general, core metabolism is a reliable starting point, as it captures carbon and energy metabolism (Lawson et al., 2019).

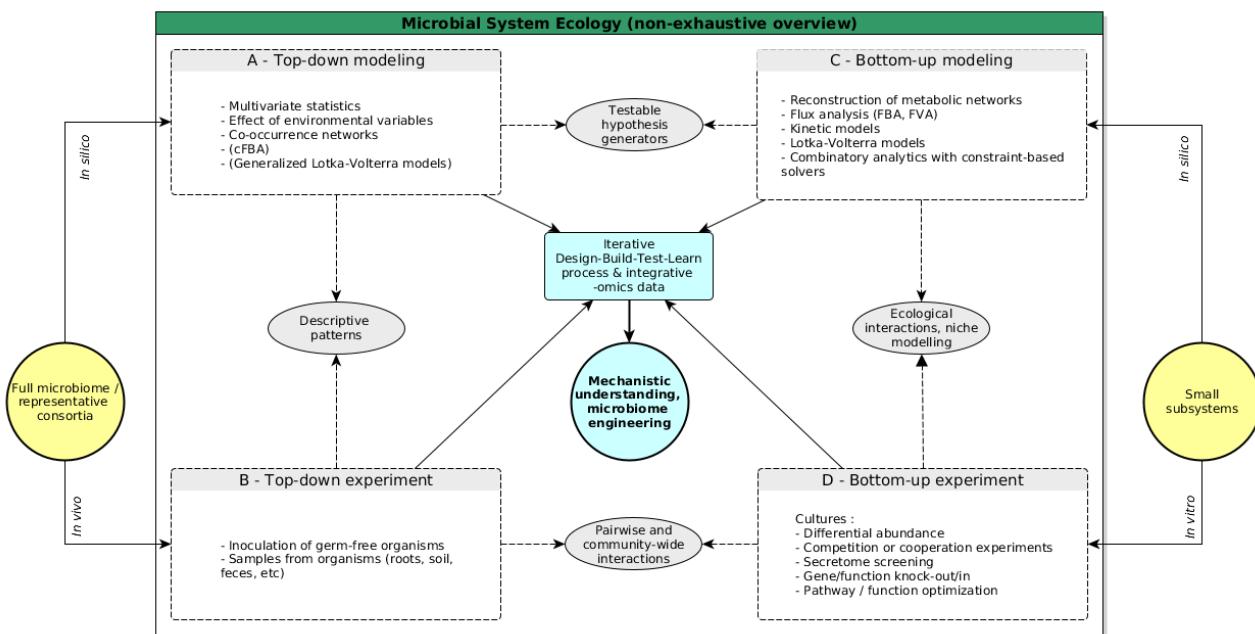


Figure 2: Schematic view of top-down and bottom-up approaches in microbial systems ecology. The list of methods, techniques, and goals is not exhaustive. In this framework, deciphering the structure and dynamics of a microbiota implies continuous and iterative shifts between approaches, either top-down / bottom-up and in silico / in vitro / in vivo. Top-down modeling (A) intensively uses -omics data obtained from high-scale top-down experiments involving numerous species (B). For example, top-down models can use descriptive and multivariate statistics to detect structural and time patterns in species abundances, or cluster microorganisms in functional groups. Both can subsequently be correlated with their co-occurrences or modeled with generalized Lotka-Volterra models (respectively based on relative abundances and growth rates with an interactions matrix), which are also used to model the potential influence of a microorganism on others. In bottom-up modeling (C), a reductionist approach is preferred, and small subsystems of microorganisms are analyzed in more detail, with emphasis on modeling how they putatively interact. Most models are based on reconstructed metabolic networks, which are crucial to predict interactions such as nutrient competition or exchange. Software based on constraint-based programming exist to rapidly find combinations that can then be further modeled using flux analysis or regular Lotka-Volterra models. Putative interactions must be tested when possible (D). Each approach and method used contributes its own knowledge and should be completed with other knowledge. Approaches must be chosen based on the research goal: microbiome engineering, synthetic biology, deciphering assembly rules of the community with a mechanistic and holistic view (etc.). Methods and techniques are provided as examples and do not claim to be exhaustive (see Shahzad and Loor (2012), Fransoza et al. (2015), Amor and Bello (2019), Lawson et al. (2019), Lloyd-Price et al. (2019), Vrancken et al. (2019) for more).

4 Conclusion

Disentangling ecological processes within a microbial community is the only way to obtain a mechanistic view of its composition, stability, productivity (etc.). Ecological interactions, particularly cross-feeding, must thus be taken into account in any microbial ecology project, notably in synthetic biology and microbiome engineering, with many applications including human health and sustainable agriculture (Toju et al., 2018; Henriques et al., 2020). With this goal in view, microbial systems ecology frameworks are being developed to unify top-down and bottom-up approaches in an iterative design-build-test-learn cycle (Lawson et al., 2019). Still, MSE should be used cautiously to avoid being drowned under hundreds of irrelevant models. Whenever possible, predictions of an MSE framework should be tested experimentally (Röling and Van Bodegom, 2014; Muller et al., 2018; Vázquez-Castellanos et al., 2019), and in return, experimental observations should improve models. To build reliable and in-depth knowledge, efforts should focus on a few aspects, such as GEM quality (in order to go beyond research on conserved, well-known metabolic pathways), integrate -omics data (Franzosa et al., 2015), niche modeling, and microbial secretome with exometabolomics (Jacoby and Kopriva, 2019).

Acknowledgments and funding

This work was supported by a grant from the French Ministry for Research and Innovation, by a grant from the CNRS (EC2CO), and also by a starting grant to S. Hacquard from a European Research Council (MICRORULES 758003), the Max Planck Institute, the Cluster of Excellence on Plant Sciences (CEPLAS), and the ‘Priority Programme: Deconstruction and Reconstruction of the Plant Microbiota (SPP DECRyPT 2125)’, both funded by the Deutsche Forschungsgemeinschaft.

Competing interests

The authors declare no competing interests.

Chapter 2 – Multi-genomes metabolic modelling predicts functional inter-dependencies in the *Arabidopsis* root microbiome

Structure, dynamics, and ecological interactions in microbial communities can be investigated and predicted *in silico* with numerous modelling methods, formalized as Microbial System Ecology (MSE). In this chapter, following the previous development around metabolic dependencies and MSE, genome-scale reconstructed metabolic network were used to find putative metabolic dependencies among a set of 193 bacteria from the *Arabidopsis thaliana* root microbiota. Reconstructed metabolic networks were notably used to predict the number and composition of metabolites producible by each genome. Three hypothesis were tested :

- The number and composition of predicted producible compounds (thus metabolism) is clustered by phylogeny.
- Available nutrients apply severe nutritional constraints on genomes' predicted metabolism.
- Metabolic cooperation via cross-feeding between genomes counter nutritional constraints.

All hypothesis were validated. Microbial taxa displayed highly differentiated metabolism in terms of predicted producible compounds, and phylogenetically distant genomes had a higher and more diverse production of metabolites, as well as reduced sets of shared metabolites. Nutritional constraints dramatically reduced the per-genome predicted number of producible metabolites. However, putative cooperation by metabolites cross-feeding acted as a counter to these constraints. At the whole community scale, the added-value brought by cooperation rose the predicted number of metabolites to a similar level as the unconstrained-one. At the scale of simple subsystems (pairs of triplets of genomes), there were numerous combinations able to produce previously unproducible targeted metabolites.

Multi-genomes metabolic modelling predicts functional inter-dependencies in the *Arabidopsis* root microbiome

Victor Mataigne^{1,2}, Nathan Vannier², Philippe Vandenkoornhuyse^{1,*}, Stéphane Hacquard^{2,*}

1 Université de Rennes 1, CNRS, UMR6553 ECOBIO, Campus Beaulieu, 35000 Rennes, France

2 Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany

*Corresponding author, victor.mataigne@univ-rennes1.fr philippe.vandenkoornhuyse@univ-rennes1.fr

Submission planned to Microbiome

Abstract

Metabolic dependencies among microbes likely explain co-existence in microbiota. In this *in silico* study, we explored genome-scale metabolic models (GEMs) of 193 bacteria isolated from *Arabidopsis thaliana* roots. We analysed their predicted producible metabolites under simulated nutritional constraints, including ‘root exudate-mimicking growth media’ and assessed the potential of putative metabolic exchanges of end and by-products to parry those constraints. We found that the genome-encoded metabolic potential is quantitatively and qualitatively clustered by phylogeny, highlighting metabolism differentiation between taxonomic groups. Random, synthetic combinations of increasing number of strains (SynComs) indicated that the number of producible compounds by GEMs increased with average phylogenetic distance, but most SynComs were centered around an optimal phylogenetic distance. Moreover, relatively small SynComs could recapitulate the capacity of the whole community due to metabolic redundancy. Inspection of 30 specific end-product metabolites (i.e. target metabolites: amino acids, vitamins, phyto-hormones) indicated that a majority of the strains had the genetic potential to produce almost all of these targeted compounds. Their production was predicted (1) to depend on external nutritional constraints, and (2) to be facilitated by nutritional constraints mimicking root exudates, suggesting the key importance of nutrients availability and root exudates to modulate the number of producible metabolites. An Answer-set-programming solver allowed to identify numerous combinations of strains predicted to depend on each other under severe nutritional constraints to produce these targeted compounds thus indicating a putative sub-community level of functional redundancy. This study predicts a metabolic restriction caused by available nutrients in the environment. By extension, it highlights the importance of the environment for niche potential, realization,

partitioning, and overlap. The results also suggest that metabolic dependencies and cooperation among root microbiota members is likely relevant to compensate for environmental constraints and to maintain co-existence in complex microbial communities.

1 Introduction

Microorganisms are known to colonize macroorganisms, establishing a dense network of interactions and contributing to essential functions maintaining their host homeostasis (Turner et al., 2013). These functions are various, ranging from protection against pathogens to nutrient uptake, or resistance to stresses such as heat or drought (Berendsen et al., 2012; Rodriguez et al., 2019). Together with other factors such as temperature, pH, oxygen, nutrients (Hacquard et al., 2015) or priority effects (Toju et al., 2018), the host itself is a niche which influences its microbiota composition (Compan et al., 2019; Griggs et al., 2021). In plants, soil can be seen as the reservoir of microorganisms from which microorganisms are recruited to compose the root microbiota (Trivedi et al., 2020; Xiong et al., 2021) and where root exudates play an active role (Haichar et al., 2008; Badri and Vivanco, 2009). Last, microbe-microbe interactions are essential in shaping the structure and dynamics of a microbiota (Konopka, 2009; Hassani et al., 2018; Pacheco and Segrè, 2019), resulting in a dense network of interactions (Sung et al., 2017).

Understanding the diverse interactions between microbes represent a critical step for a holistic, community-level understanding of microbiota functioning. These microbe-microbe interactions spread on a spectrum going from competition to cooperation (Coyte and Rakoff-Nahoum, 2019), depending on time, space, other species presence, and energetic cost (etc) (Pacheco and Segrè, 2019). Two main competing theories explain microbial assemblages. The niche differentiation theory states that phylogenetically similar species are more likely to compete with each other due to their shared functional traits and resource overlap, leading to less probable co-existence (Lam et al., 2020). The habitat filtering theory suggests that dominant species exhibit similar functional traits, because their presence is determined by environmental parameters (Levy and Borenstein, 2013). Until now, the relative importance of competition and cooperation remains unclear (Johnson et al., 2012; Coyte and Rakoff-Nahoum, 2019). Different approaches often give conflicting results : several computational approaches predicted many cross-feeding possibilities while some *in vitro* experiments rather highlighted competition (Coyte and Rakoff-Nahoum, 2019). However, metabolic interactions and particularly metabolic dependencies are reported to play a major role in maintaining community diversity, stability and in explaining microbial co-existence (Morris et al.,

2012a; Mee et al., 2014; Lovley, 2017; Sung et al., 2017; Zomorrodi and Segrè, 2017; Seif et al., 2020). Extracellular metabolites can thus play a major role in microbial community assembly (Fritts et al., 2021), and metabolic dependencies among strains might explain why some microbes cannot be cultured in standard laboratory conditions (Bernstein et al., 2019).

Microbial system ecology approaches are now regularly used to model complex systems such as ecological processes (Faust et al., 2018; Muller et al., 2018; Kumar et al., 2019; Mataigne et al., submitted). The acquisition and analysis of -omics data, coupled with modelling approaches, allow to computationally predict an organism's resource usage, biosynthetic capabilities, deficiencies, and growth across various conditions, notably available nutrients (Muller et al., 2018; Bernstein et al., 2019), which are herein referred as "nutritional constraints" (table 1). These models rely on the reconstruction of metabolic networks (GEome-scale Metabolic models (GEMs)) from annotated genomes (Feist et al., 2009; Henry et al., 2010). Thus, predicting fundamental niche overlaps and competition between members of a microbial community becomes possible (Alneberg et al., 2020). Studying phylogenetic structure of microbial communities also permitted to detect correlations between the phylogenetic signal and metabolism (Easson and Thacker, 2014; Goberna and Verdú, 2016; Li et al., 2017; Aguirre De Cácer, 2019).

We questioned *in silico* (1) how phylogeny shapes GEMs both at single strain and (2) small random combinations of strains (SynComs) scales, (3) how strong is the effect of the constraint applied by available nutrients (including root exudates) on GEMs', and (4) how metabolic cooperation can balance these nutritional constraints. We tested 4 hypotheses. -(i) unconstrained metabolism (see definition in table 1) is highly clustered by phylogeny, meaning that predicted producible metabolites are differentiated or overlapped between strains according to their taxonomy (H1); (ii) combinations of GEMs would have more producible metabolites than single GEMs, depending on the phylogenetic similarity between the corresponding bacteria (H2); (iii) available nutrients have a potential impact on the metabolism of bacteria, leading to a reduction of producible metabolites from unconstrained to constrained metabolism (H3); (iv) metabolic cooperations are frequent and likely compensate for nutritional constraints by allowing the production of specific key compounds (H4), under the strong postulate that every compound produced by a bacterium can be shared with others. This postulate is at least partially supported by several studies which analyzed or predicted bacteria and plants secretomes (Bednarek et al., 2010; Baran et al., 2015; Orsi et al., 2017). To address these hypotheses we analyzed a collection of genomes of bacterial strains isolated from the roots of *Arabidopsis thaliana* (Bai et al., 2015) and used system biology

approaches to predict *in silico* the genomes functioning. Metabolism of each bacteria was predicted with GEMs reconstructed from genomes' annotation.

2 Materials and Methods

Table 1 provides vocabulary and definitions and Figure 1 adduces a graphical summary of what has been done regarding data and metrics acquisition.

Table 1 : Description of the metrics used. The term “community” means either the **full community** or a random subsample of strains (**SynCom**) (but precise terms are employed when necessary). The acronyms in bold in the table are used throughout the text.

Metric	Description
Nutritional constraint	Available nutrients on which a GEM can rely on (i.e. the initial reactants of the whole network). Nutritional constraints are modelled with simulated growth media. In this paper, an “ unconstrained ” GEM represents its metabolic potential, i.e. all the metabolites it encodes and can produce in theory.
Predicted Produced Metabolites (PPM)	The list (number and composition) of all metabolites predicted to be producible by one or several GEMs at once (also referred as a meta-GEM), under a nutritional constraint or without constraint. Such a metric is used to summarize the unconstrained and constrained (by available nutrients) metabolism inferred from genomes .
Core Predicted Produced Metabolites (CPPM)	The part (number and composition) of a community PPM which is producible by every GEM individually in a set of GEMs.
Targeted Predicted Producible Metabolites (TPPM)	A set of metabolites on which a part of this study is focused. Their ability to be produced by one or several GEMs is analyzed (number and composition), under a nutritional constraint or not.
Community's added value	The part of the PPM (number and composition) of several GEMs which is only producible by metabolic interactions within a community (i.e., not producible by any single GEM).
Average phylogenetic distance	The average of all pairwise phylogenetic distances between pairs of strains in a synthetic subsample of strains (SynCom). The full community also has its average phylogenetic distance.

2.1 Genomes data

We used 193 annotated bacterial genomes (Bai et al., 2015) all isolated from *A. thaliana* roots collected in Cologne soils (Germany), and selected to build a taxonomically representative core set of bacteria of the host-plant (Bai et al., 2015). Annotated genomes (Bai et al., 2015; Wippel et al., 2021) were downloaded from the At-SPHERE database (<http://www.at-sphere.com/>). Both taxonomy and phylogeny of the whole set of genomes (Bai et al., 2015) were used. The phylogenetic tree was inferred by maximum likelihood (Bai et al., 2015) from a multi-alignment of

31 bacterial AMPHORA (Wu and Eisen, 2008) genes obtained with Clustal Omega v1.2.1 (Sievers et al., 2011) passed into FastTree v2.1 (Price et al., 2010).

2.2 Reference database

The reference to link genomes' annotation to metabolism was the MetaCyc database, a collection of organism specific Pathway/Genome Databases (PGDBs). MetaCyc contains ~2500 metabolic pathways from many organisms (Caspi et al., 2008). Two criteria motivated this choice. First, the database has been manually curated. Second, our metabolic networks reconstruction tools (*mpwt* and *AuReMe* (Aite et al., 2018; Belcour et al., 2020) are pre-configured to work with this curated database.

2.3 Metabolic networks (GEMs) reconstruction

Metabolic networks of each genome were simulated with genome-scale metabolic models (GEMs) reconstructed with an automated command-line version of PathwayTools (Karp et al., 2002, 2011, 2020) using the *mpwt* program of the metage2metabo tool suite (Belcour et al., 2020), then converted in padmet and sbml format with AuReMe and padmet-utils (Aite et al., 2018). GEMs in *sbml* format were parsed with the python lxml package when needed. All GEMs were drafts, used without gap-filling or manual curation procedure. Those steps are usually required to improve the quality of a GEM (Thiele and Palsson, 2010), but are likely to introduce false positives, particularly with poorly known organisms, hiding potential metabolic dependencies. Consequently, we chose to rely on drafts of GEMs, preferring false negatives (due to flaws in genomes' annotation) than false positives.

2.4 Metrics from genomes and GEMs

In order to detect patterns between strains and metabolism, a set of metrics (complete definitions are provided in table 1) were used and applied on single GEM (i.e. single strains) and on random combinations (“SynComs”) comprising 2 to 20 GEMs. The Python API of “Miscoto scope” (Frioux et al., 2018) was used to compute all the Predicted Produced Metabolites, (PPM) for single GEM, SynComs, and the full community under simulated nutritional constraints. AuReMe reports were parsed to record constraints-free PPM. Phylogenetic distances were computed based on the phylogenetic tree with the Python package *ete3* (Huerta-Cepas et al., 2016). Genomes sizes were available in the annotation data. For each SynCom was also computed the Core Predicted Produced

Metabolites (CPPM), using sets data structure in Python 3 and AuReMe and Miscoto scope outputs (Figure 1). The full community's added value was also computed with Python sets.

2.5 Targeted Predicted Producible Metabolites (TPPM)

TPPM are compounds for which the capacity of production was studied, for single GEM and for the whole community (i.e. meta-GEM, where all GEMs can leak and exchange any compound). The production of TPPM from GEM can be computed under various nutritional constraints with simulated growth media (see the dedicated section below). There were 30 TPPM in this study : 17 amino-acids (Serine, Alanine, and glutamic acid were excluded because present in Artificial Root Exudates, which were part of the growth media), 8 B vitamins (thiamine diphosphate, riboflavin, nicotinate, (R)-pantothenate, pyridoxine, biotin, tetrahydro-folate, adenosylcobalamin) and 5 phytohormones (auxin, salicylic acid, abscisic acid, ethylene, jasmonic acid). Amino-acids were chosen for their fundamental, ubiquist biological importance. Vitamins were chosen according to their importance in metabolism. Phyto-hormones were chosen in regards to the root-associated trait of the studied microbial community. In addition, these metabolites biosynthesis pathways are relatively well-known, reducing the risk of false negatives in reconstructed metabolic networks. When under a nutritional constraint (see below), TPPM production was computed for each GEM with the Python API of “Miscoto scopes” (Frioux et al., 2018). Without nutritional constraint, TPPM production was assessed with their absence/presence in AuReMe reports.

2.6 Nutritional constraint (growth media) modeling

Nutritional constraints were set up with the simulation of various growth media. We used two online resources to choose the growth media : MetaCyc (previously mentioned), which contains few growth media with their detailed list of nutrients, and the KOMODO database (Oberhardt et al., 2015). KOMODO references a huge number of growth media, providing their composition, as well as the constituents MetaCyc IDs. We modelled 9 growth media (5 poorly nutritive with essentially mineral nutrients and 4 highly nutritive, with more carbon sources, table 2), in which a common list of cofactors (Kim et al., 2015) were added. Each medium has an ‘Artificial Root Exudates’ (based on Baudoin et al., (2003) enriched version (i.e. 22 media in total) (Table 2). Growth media, except rich media, were chosen according to two conditions : (1) none of the TPPM should be in a medium's nutrients, and (2) the nutrients mustn't contain any “mixture compound”, such as tryptone or yeast extract, whose composition is not characterized in MetaCyc. ‘Rich’ media did not fully match these conditions because they contained vitamins. Among them, LB-lennox-enriched

medium contains tryptone and yeast extract. All growth media were used to investigate a range of nutritional constraints on TPPM, but some were discarded when analyzing the production of TPPM they were incompatible with (table 2).

Table 2 : Summary table of the used growth media for nutritional constraints. Since rich media always contain some TPPM, the set of studied growth media varies according to the analysis: TPPM were either analyzed all at once, without vitamins, or only phyto-hormones, according to the content of the media. Detailed media composition can be found on gitlab.

Growth media	Type	Contains TPPM	With Mixture	Analyzed TPPM
M63	Poor	No	No	All
M9				
mineral medium				
hydrogen oxidizing				
MBM				
basal	Rich	Vitamins	Tryptone (amino-acids) + Yeast extract	Amino-acids, phyto-hormones
phb pyruvate				
MMJS				
LB-lennox enriched		Vitamins + amino-acids		Phyto-hormones

2.7 Putative GEMs combinations for metabolic interactions

For each nutritional constraint (i.e. each medium), Miscoto mincom (Frioux et al., 2018) (a version implemented into the metage2metabo (Belcour et al., 2020) package) was used with the whole microbial community as input (GEMs in *sbml* format). This tool uses Answer Set Programming, a declarative approach oriented toward combinatorial problem-solving (Collet et al.; Laniau et al., 2017). It rapidly found all the simplest combinations (called “solutions”) of GEMs able to produce as many specified TPPM as possible under growth constraints. More complex solutions (with more GEMs) are ignored. TPPM were either all of the aforementioned compounds, non-vitamins, or phyto-hormones only, in accordance with the aforementioned condition “no TPPM in growth media” (table 2). Splitting TPPM was also important because results are strongly TPPM-dependent : for example, GEMs with an essential, but rare reaction for the production of a given TPPM will be very represented, potentially masking alternative possible combinations for other TPPM. Results of each run were stored in a *json* format.

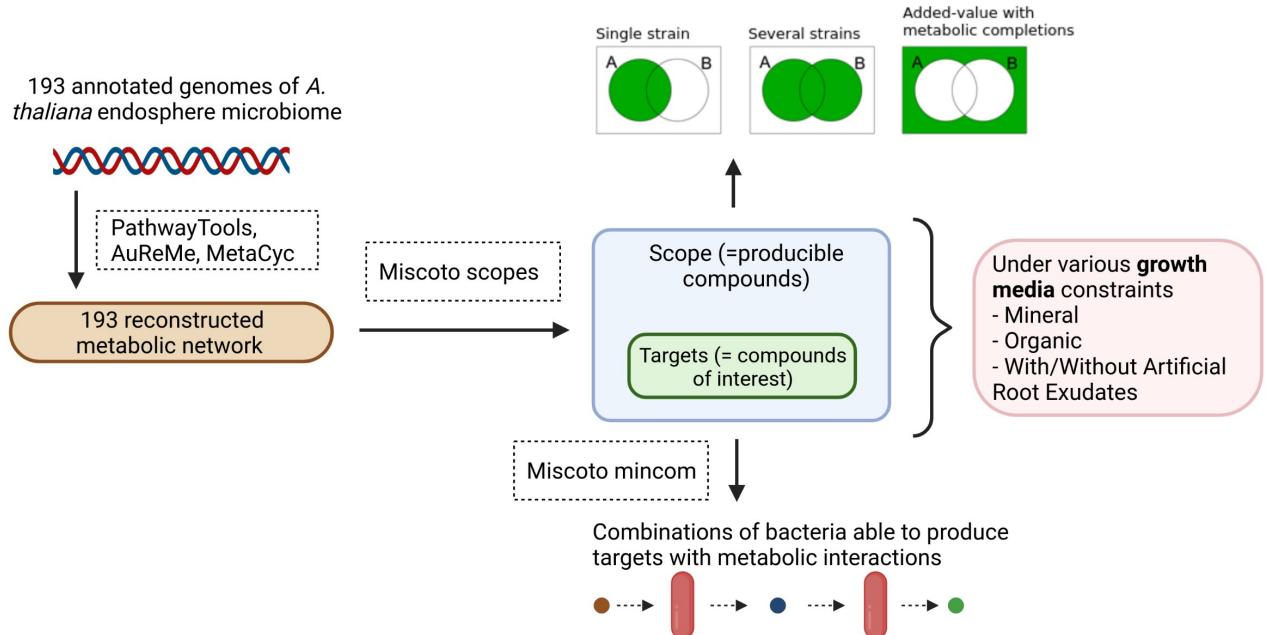


Figure 1: Genome sequence data processing and metrics acquisition. Our analysis relies on reconstructed metabolic networks, one per genome, for which PPM and TPPM production were computed, under several nutritional constraints. PPM can also be computed for communities of several cooperating GEMs : in that case, the part of the PPM and TPPM producible only by the community ('added-value') is also returned. Last, minimal combinations of GEMs able to produce as many TPPM as possible were computed.

2.8 Quasi-Poisson GLMs

The correlations between number of PPM/CPHM and SynCom size, genome size, and phylogenetic distance were computed (without nutritional constraint) on random SynComs. First, for each number of members (variable 'SynCom size') going from 2 to 20 (with a step of 1), 500 SynComs were built by randomly picking strains from the initial pool of 193 strains, without replacement. First, independence between explanatory variables were assessed (supplementary figure IV D). Then two Generalized Linear Models based on a quasi-poisson distribution (in order to counter overdispersion of the response metrics, supplementary figure IV, C1 to C4) were built, modelling the response of number of PPM and CPHM of SynComs according to average phylogenetic distance, average genome size, and SynCom size. A supplemental polynomial (degree 2) term of phylogenetic distance was added to model the curve of the response metrics. :

$$Y \sim P(\mu_i, \theta)$$

$$\mu_i = \exp(\beta_0 + \beta_1 P_i + \beta_2 G_i + \beta_3 S_i)$$

With i being a SynCom, Y being its number of PPM or CPHM, P its average phylogenetic distance, G its average genome size, S its number of strains (i.e. GEMs). The maximum SynCom size to put in the model was 12 strains. This size was chosen for its position at the start of the plateau reached

by all the metrics values or variance (supplementary figure IV, B1 to B4). This was also the size whose bigger Syncoms started to show no significant difference from one SynCom size to the next when tested with many bootstrapped subsamples (for each SynCom size: 500 iterations of 50 strains each, one-sided Wilcoxon, Mann & Whitney tests, supplementary figure I). The models' residuals were slightly biased and non-normal, caused by the uneven spreads of values in different SynComs sizes (figures S5 and S6). Since R2 are not automatically returned with the models we used, they were computed with the following formula : 1 - Residual deviance / Null deviance (which are available in models' R summaries). For computing time and resources, SynCom's PPM added value was left aside for this analysis.

2.9 Other statistical analyses

Tests of the effect of taxonomy on metrics distributions and growth media on the number of PPM were performed using non-parametric tests (Wilcoxon, Mann & Whitney rank-sum tests). Corresponding effect sizes were computed with the Cliff's delta method. Principal Coordinate Analysis (PCoA) were performed on Jaccard distance matrices, with the *pcoa* and *vegdist* function of the R packages ape and Vegan (Oksanen et al., 2020). Permanovas on the distance matrices were performed with the adonis function associated with a multivariate analogue of Levene's test for homogeneity of variances (PERMDISP2 procedure), with the *betadisper* function of the same package. Growth media dendrogram was built by hierarchical clustering with the R base *hclust* function (with the default 'complete' method'), after computation of Bray-Curtis distances (with the *vegdist* function) based on the composition of the corresponding full community's PPM (qualitatively, i.e. which compounds are producible under which nutritional constraint, by how many GEMs). Tests implying taxonomic effect excluded Bacteroidetes and Firmicutes phyla because of their small sample sizes (4 and 7 strains respectively). Significance threshold was fixed at 0.01.

2.10 Scripting

The organization of data acquisition and links between tools inputs and outputs (Figure 1) was made with homemade Python 3 scripts. All Miscoto outputs were stored in *json* format, and the relevant data they contained (PPM and TPPM, number of genomes producing a TPPM under a given nutritional constraint ...) was parsed and stored as *csv* tables. Figures and data analysis were performed with R 4 with the *ggplot2* package (Wickham, 2016) and Python 3 with the *matplotlib* and *seaborn* packages. Scripts and data are available at

https://gitlab.com/mataivic/article_metabolic_modelling_thaliana_microbiome

3 Results

3.1 A link between genome-predicted unconstrained metabolism and phylogeny

We first tested how phylogeny structured the distributions of the different metrics, under the hypothesis that GEMs unconstrained metabolism differs along phylogeny (H1). Unconstrained metabolism corresponded to the situation where all the putative genes carried by a genome are functioning (table 1). Only Actinobacteria and Proteobacteria distributions were statistically tested (i.e. the number of strains was insufficient for other phyla). Smaller genomes sizes and number of PPM were observed in Actinobacteria compared to Proteobacteria (figure 2B, $p=0.0038$ and $p<0.0001$, with effect sizes of -0.27 and -0.48). The bigger the genomes are, the higher is the number of PPM and TPPM (figure 2C). The Bacteroidetes in the culture collection ($n=4$) displayed small genomes and number of PPM, while the Firmicutes ($n=7$) number of PPM and size of genomes were similar to the other phyla. Differentiation in PPM composition was also detected among phyla based on PERMANOVAs ($p<0.001$, $R^2=0.213$, p (permdisp) = 0.005) with well separated groups (including Bacteroidetes and Firmicutes) observed on PCoA (figure 2D). When considering the 30 TPPM only, this effect remained significant ($p<0.001$, $R^2=0.175$, p (permdisp) = 0.2325), but between-group differentiation was reduced (figure 2D, supplementary figure II D). The same patterns are conserved at the class level (supplementary figure II). This suggests a strong metabolic differentiation between phyla at the whole GEMs scale, but a more conserved metabolism when considering the TPPM.

3.2 SynCom unconstrained PPM are greater than that of single strains and rapidly reaches saturation

We extended the previous single-GEM approach to random SynComs in order to analyze the metabolic capacities (number of PPM and CPPM, without nutritional constraint) of merged GEMs (i.e. fully cooperative with all possible metabolic exchanges) (H2). The correlations of number of PPM and CPPM to SynCom size ($n=2$ to 20 members on plots, $n=2$ to 12 in GLMs), genomes size, and phylogenetic distance were explored. The three of them were significantly correlated both with the number of PPM ($R^2=0.86$ diagnostic plots on supplementary figure V and CPPM ($R^2=0.82$, diagnostic plots on supplementary figure VI).

SynComs size strongly increased SynComs PPM (coefficient=0.36, $p<2e-16$) and CPPM (coefficient=-0.65, $p<0.001$), but this effect was more pronounced for small-size SynComs.

Notably, data from SynComs with many strains largely overlapped, plateauing at about 2,000 producible metabolites and 400 core-metabolites. Increasing SynCom size up to 193 strains would result in less and less variations in SynComs, until reaching the full community with a PPM of 2383 and a CPPM of 263. Interestingly, increasing SynComs size rapidly returned numbers of PPM and CPPM close to the full microbial collection (figure 3) likely due to metabolic redundancy. Roughly, SynComs composed of more than ~12 GEMs, displayed numbers of PPM and CPPM closer to the values of the full 193-members community than the values of the smallest SynComs (figure 3).

Phylogenetic distance increased SynComs PPM (polynomial coefficients=1.32 and -1.83, $p<0.001$) and decreased SynComs CPPM (polynomial coefficients=-4.37 and 1.2, $p<0.001$) (figure 3.A&B). However, PPM reached a peak at a phylogenetic distance of ~1.3 for small SynComs then decreased (Figure 3A; Figure 4A). This decrease turned into a plateau at higher SynCom sizes (Figure 3A). CPPM among GEMs decreased then reached a plateau. This highlighted more and more diverse metabolism and less and less shared metabolism between strains. Most SynComs were concentrated around this phylogenetic distance rather than spread equally along all possible distances, particularly for big SynComs (figure 3 and S4, A1 to A4). This corresponds to the values of most combinations of Proteobacteria and Actinobacteria (supplementary figure III A to B) and explained by the fact that these two phyla were dominant in the dataset.

Genomes sizes significantly increased both PPM (coefficient=0.23, $p<0.001$) and CPPM (coefficient=0.46, $p<0.001$) (figure 3C&D). Thus, genomes sizes compensated for the negative effect of phylogenetic distance on metabolic redundancy. At a fixed phylogenetic distance, SynComs with bigger genomes have both a higher metabolic diversity and a bigger core metabolism than SynComs with smaller genomes (supplementary figure III C1&C2). SynComs were also aggregated around an optimal value, linked to the corresponding optimal phylogenetic distance and the taxonomic composition of SynComs.

The combination of GEMs of phylogenetically distant strains is positively correlated to metabolic diversity and negatively correlated with metabolic redundancy. However our results displayed many supplementary patterns we did not predict : most combinations of GEMs are aggregated around the same mean phylogenetic distance, a few GEMs are sufficient to approximate the full community metabolism, and big genomes both increase metabolic diversity and redundancy.

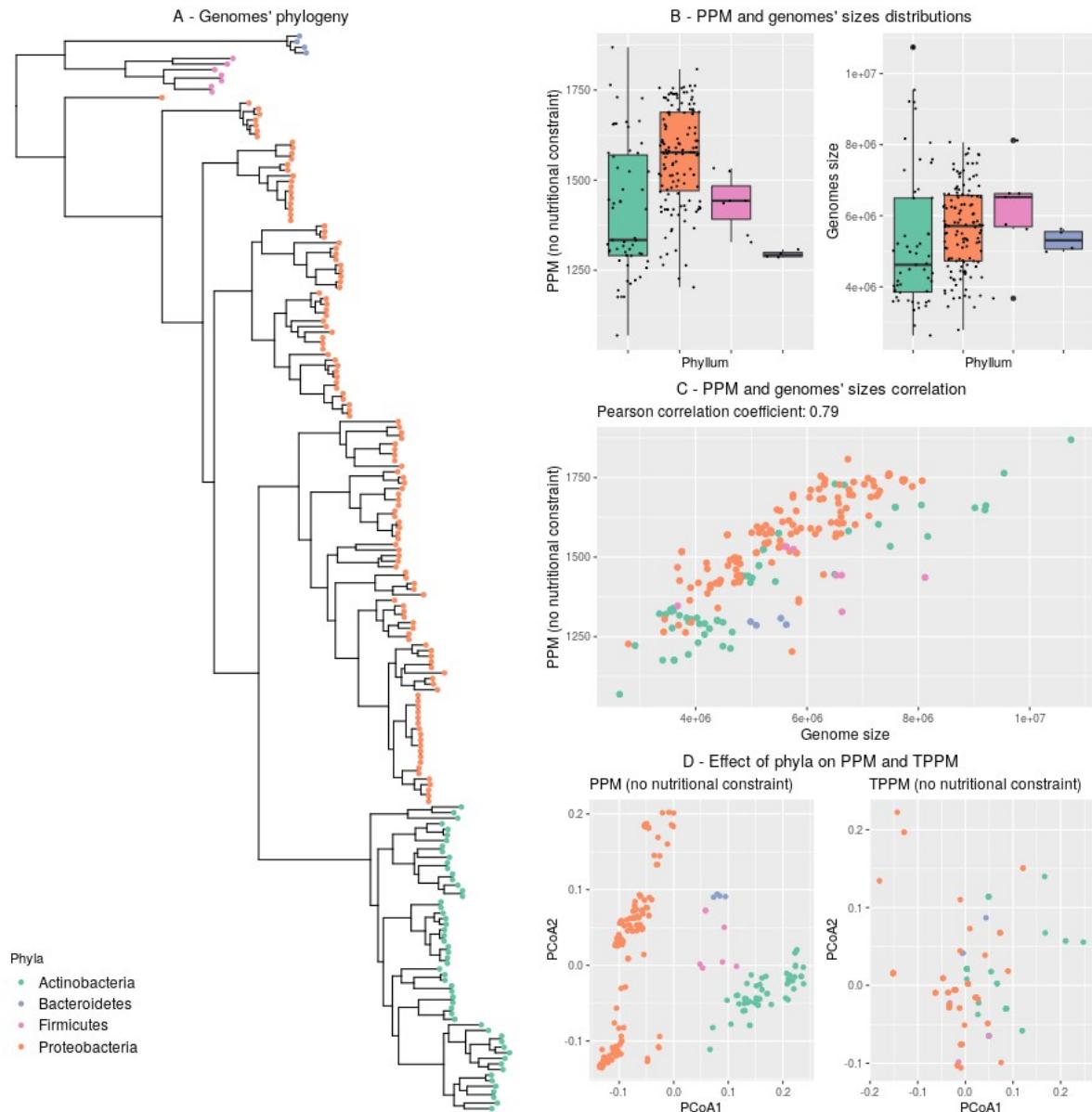


Figure 2: Description of the 193 genomes collection from *A. thaliana* roots microbiota (A) Phylogenetic tree (maximum likelihood from a multi-alignment of AMPHORA genes in Clustal Omega) (B) Boxplots displaying the quantitative effect of Phyla on genomes sizes and distributions of the number of PPM. (C) Dotplots of the number of PPM and TPPM as a function of genomes size. (D) PCoA displaying the qualitative effect of phyla on metabolites' production (i.e. which compounds are produced by which taxa). Across the panels, Colors match genomes' phyla.

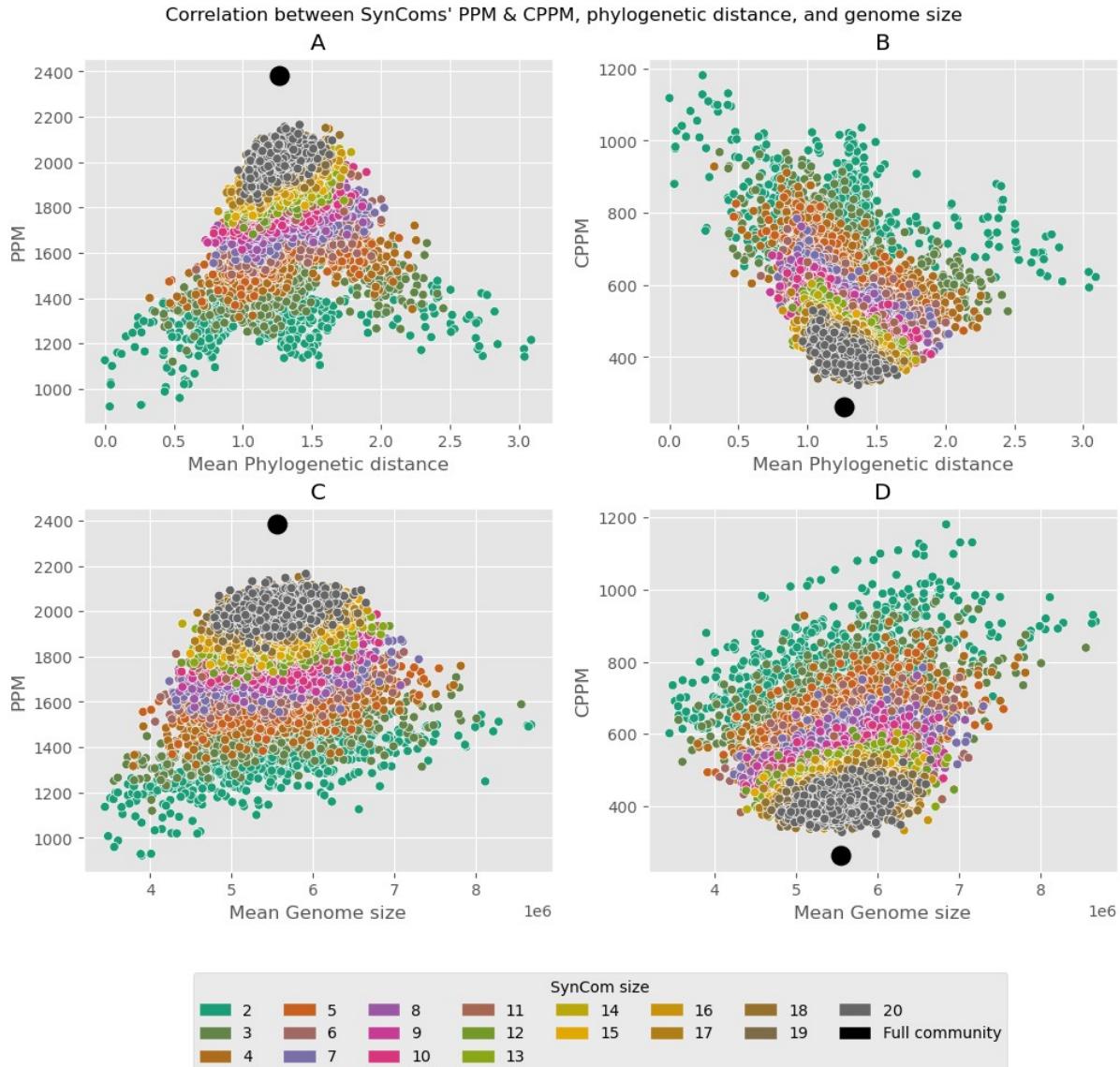


Figure 3: Correlations between explanatory metrics (mean phylogenetic distance (A&B), and mean genome size, (C&D)) and response metrics: number of PPM (A&C), and CPPM (B&D). These correlations were used for the quasi-Poisson GLM models. In each panel, the black dot corresponds to the full community (193 genomes) values. The number of PPM increases rapidly with SynCom Size, mean phylogenetic distance, and mean genome size, then starts plateauing slowly towards the full community value. (see also supplementary figure III, notably panels B1 to B4). The number of CPPM was negatively correlated with SynCom size and mean phylogenetic distance, and positively correlated with mean genome size.

3.3 PPM and TPPM number and composition depends on nutritional constraints

In this analysis, we predicted the shift between unconstrained and constrained GEMs under the effect of available nutrients (referred as ‘nutritional constraints’, exemplified here by the simulated growth media composition, table 1) (H3). “Poor media” refer to severe constraints with essentially mineral nutrients, while “rich media” refer to more permissive constraints, with more carbon sources (see methods).

At the single GEM scale, PPM under nutritional constraints were dramatically reduced compared to PPM of unconstrained GEMs (all $p < 0.001$ with Holm correction, figure 4A, right plot). PPM under poor media were significantly lower than PPM under rich media ($p < 0.001$, Cliff’s delta effect size = -0.8978 without ARE and $p < 0.001$, Cliff’s delta effect size = -0.5102 with ARE; figure 4A). Media supplementation with compounds artificially mimicking exudates composition of plant roots (ARE) is predicted to significantly increase the number of PPM compared to the non-supplemented media (green vs. orange colour in figure 4A, table 3, $P < 0.001$). Notably, PPM compositions were more similar (i.e. similar sets of producible compounds) across media containing ARE, irrespective of the media poor/rich initial aspect (Figure 4B). Hence, the addition of ARE is predicted to unlock the production of the same metabolites across media. Most of the GEMs have the potential to produce most of the TPPM (27 out of 30 being producible by single GEM within the full dataset), but cannot complete the entire pathways under most of the nutritional constraints. The supplementation of growth media with ARE was predicted to increase the number of TPPM producible by single GEM (figure 4D).

At the full community scale, the added-value brought by metabolic exchange between all GEMs increased the community’s number of PPM at similar values regardless of the nutritional constraints (figure 4C). Depending on the media type (poor/rich) and the absence/presence of ARE, the community added-value increased the number of PPM from 22% to 140%. Thus, GEMs functioning is limited by available nutrients that condition which reactions can be activated, but metabolic exchanges between all GEMs of a community can largely compensate for growth constraints, whether they are severe or not. The next part of the analysis was to focus on how smallest assemblages can also compensate for growth constraints.

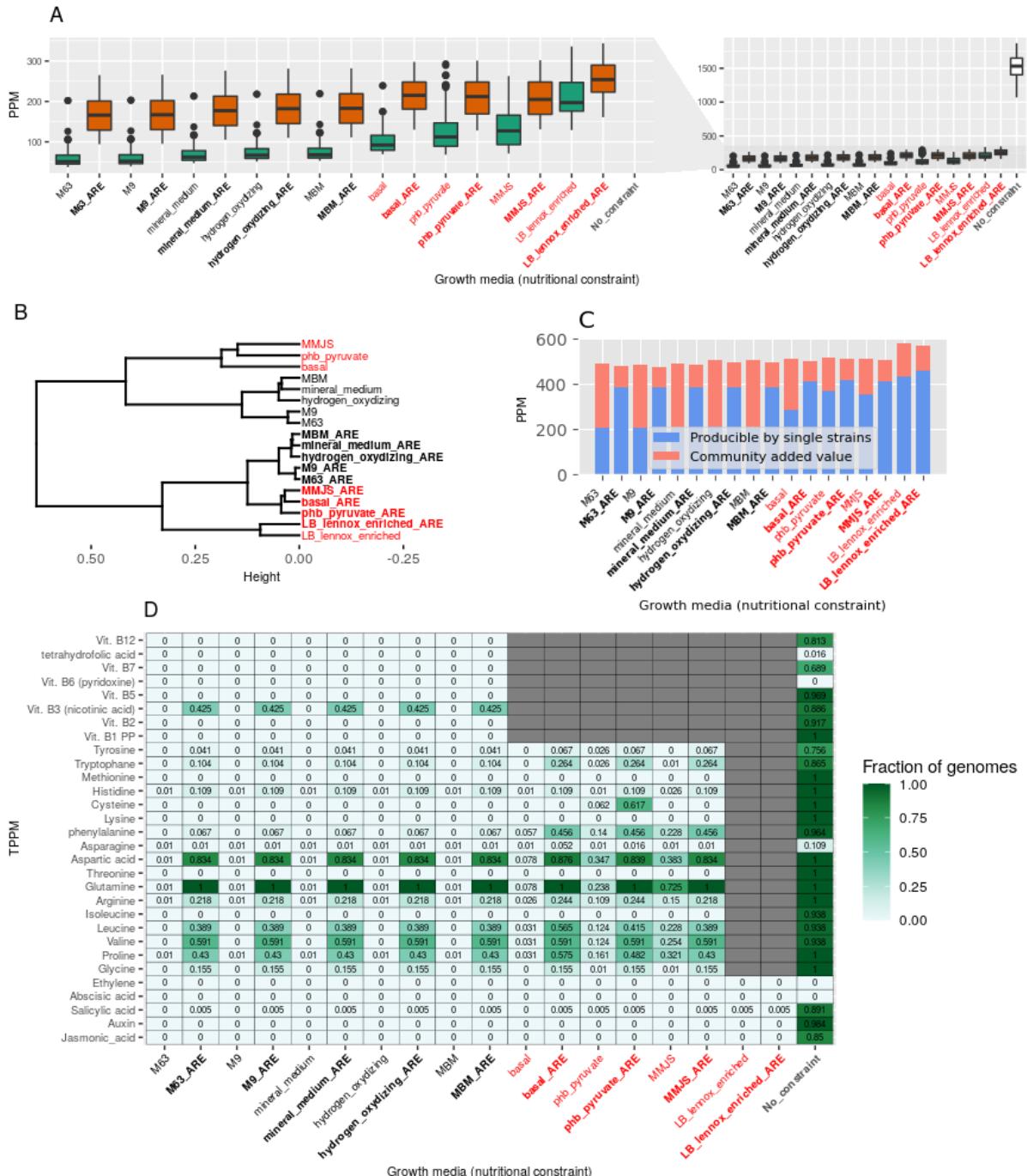


Figure 4: Effect of nutritional constraints on PPM and TPPM. (A) Boxplots of GEMs' number of PPM according to the growth media, compared to no applied constraints. (B) Hierarchical clustering ('complete' method, Bray-Curtis distances) of growth media according to the composition of the community's PPM (without community's added value). (C) Full community's number of PPM according to the growth media, with the added-value brought by metabolic cross-feeding. (D) TPPM production per growth media. Decimal numbers indicate the fraction of GEMs capable of producing the compound. Gray cells correspond to growth media already containing the TPPM, which are thus irrelevant.. On all figures axes, poor media are labeled in black, rich media in red, ARE-enriched media in bold and standard media in plain text

Table 3 : Wilcoxon rank-sum tests on the effect of ARE on the number of PPM.

Media (with and without ARE comparison)	Mann & Whitney p value	Effect size (Cliffs' delta)
M63	P<0.001*	-0.9786
M9	P<0.001*	-0.9787
mineral medium	P<0.001*	-0.9788
hydrogen oxydizing	P<0.001*	-0.9787
MBM	P<0.001*	-0.9788
basal	P<0.001*	-0.9723
phb_pyruvate	P<0.001*	-0.785
MMJS	P<0.001*	-0.7081
LB lennox enriched	P<0.001*	-0.523

3.4 Simplest SynComs are predicted to produce TPPM through metabolic exchanges

We used ‘Miscoto mincom’, an answer-set programming solver designed to automatically find, under a simulated nutritional constraint, all the simplest (i.e. smallest) combinations of GEMs that complete the metabolic pathways to produce the specified TPPM (30 in total) (H4). Combinations of GEMs are considered as a meta-GEM where an incomplete pathway in a single GEM can be completed by another, thanks to exchange of intermediate products.

Possible combinations were numerous and always small. Most of the time, 2 or 3 GEMs were sufficient to produce the TPPM. The sets of combinations were also regularly identical from one media to another (figure 5A), suggesting the important effect of identical nutrients among growth media and shared reactions among GEMs. The total number of GEMs involved in combinations varied according to the TPPM included in the analysis. For example 67 GEMs were returned when all TPPM were considered under poor media (11 under rich media), while the whole set of 193 GEMs (for a few media) were returned when only amino-acids or vitamins were considered (figure 5A).

Globally, rich media (amino-acids and phytohormones TPPM) returned very few combinations, meaning they allow more GEMs to produce TPPM without predictable mandatory metabolic dependencies or cooperation (figure 5A). For example, there were less than 10 minimal

combinations of two different GEMs able to produce 12 amino-acids, while poor media returned dozens to thousands of combinations of two or three GEMs. The addition of ARE also decreased the number of simplest combinations (for example from 2288 to 82 for rich media with vitamins as TPPM), with the exception of the situation “all TPPM + poor media” (figure 5A, first row). In such a situation, the ARE-enriched poor media contained more combinations (724 for ARE-enriched poor media, 448 for standard poor media, for 14 producible TPPM).

GEMs combinations increased the number of producible TPPM compared to single-GEMs capacities under several growth constraints (figure 5B). Under poor media (i.e severe nutritional constraints), 6 to 7 supplementary amino-acids were predicted to be producible by thousands of combinations of two or three GEMs compared to individual GEMs (often 6 amino-acids). Interestingly, for rich media all amino-acids were predicted to be producible by single GEMs, but none could produce themselves the 17 targeted amino-acids, resulting in a few combinations of strains predicted to exchange end products instead of intermediate metabolites (figure 5B). Among vitamins and phyto-hormones, only nicotinic acid (vitamin B3) and salicylic acid were producible, with required metabolic exchanges between 2 GEMs under the constraints of poor media.

The frequencies of GEMs in solutions was highly unequal with few GEMs occurring in several hundreds solutions (figure 6A). The other GEMs were much less frequent with only a few occurrences. Thus, for all TPPM at once, the majority of solutions can be aggregated with a set of 7 GEMs (supplementary table S1), belonging to strains of the phyla proteobacteria and, surprisingly, firmicutes (despite being very underrepresented in the dataset). When splitting TPPMs by category, the majority of solutions can be aggregated with a set of 15 GEMs (supplementary table S1). We did not find any clear correlation between a GEM frequency and its corresponding genome’s size (figure 6B). In fact, the high frequency of these particular GEMs were driven by their strong contribution to the production of a few particular TPPM. We noted the case of salicylic acid (supplementary figure VII). In MetaCyc, the bacterial [salicylic acid pathway](#) is composed of two reactions depending on the availability of chorismate (which has a longer pathway), but other reactions out of this pathway are also recorded in the database. The first reaction is encoded into 5 GEMs only, which are part of the aforementioned 7 main GEMs. The second one is encoded into 47 GEMs. Most of the combinations for the completion of this pathway of salicylic acid are then built with those GEMs. We noted that it differs with the total of 74 GEMs returned by the solver, which highlights the production of salicylic acid by the reactions in other pathways than the main one described above (as described in Lefevere et al. (2020) in plants and Mishra et al. (2021) in plants

and bacteria). One GEM was remarkable as it was the only one predicted to encode the complete salicylic acid pathway: the *Pseudomonas* identified as 'Root569', and could putatively produce salicylic acid by itself under the condition of being placed under the constraint of a rich medium.

Hence, many small combinations of GEMs could compensate for diverse nutritional constraints by exchanging metabolites, unlocking the production of several TPPM unreachable by single GEM.

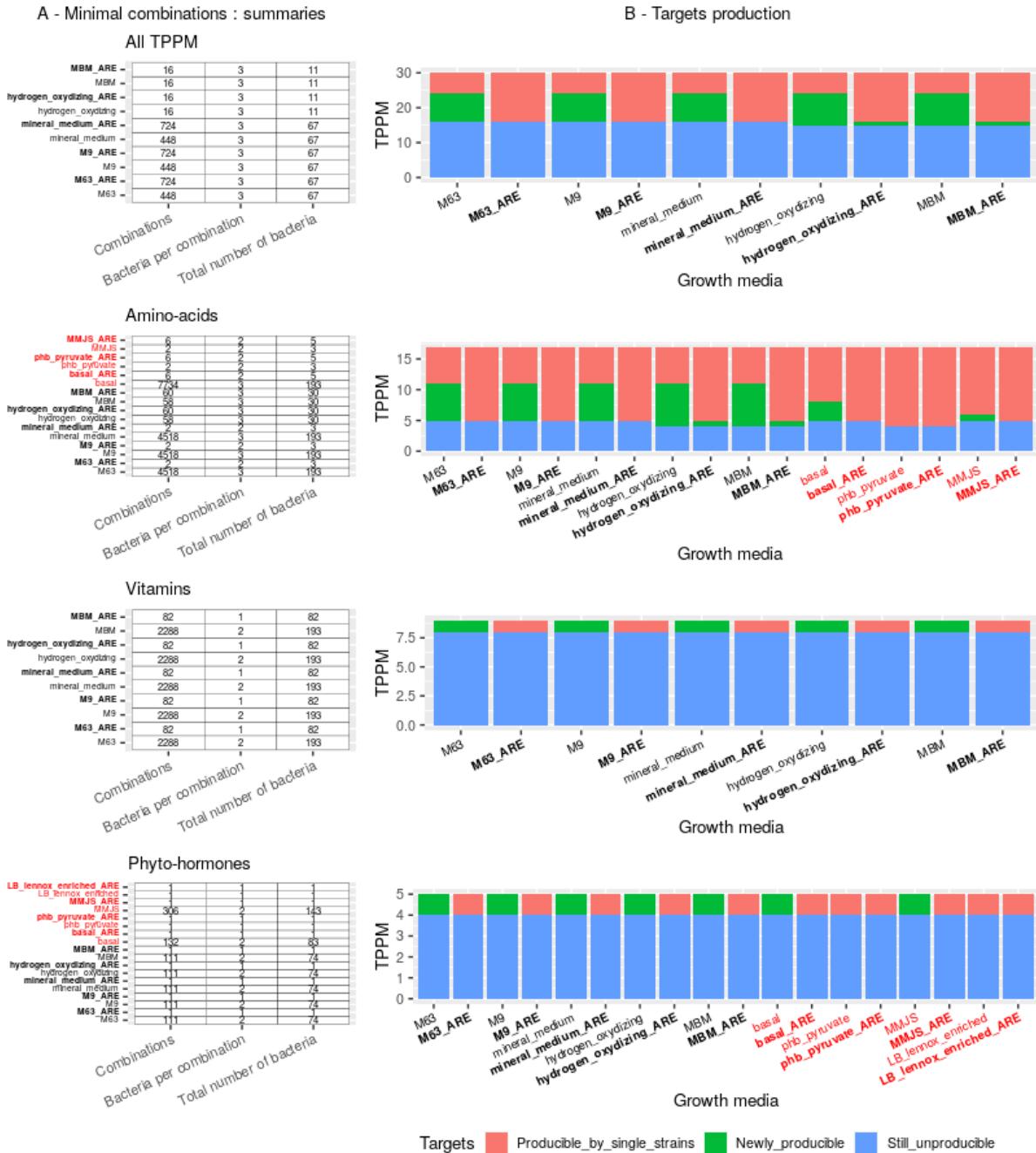


Figure 5: Minimal combinations of GEM to produce TPPM. (A) Summary of the number of combinations of GEMs able to produce the TPPM, for all TPPM simultaneously, amino-acids only, vitamins only, and phyto-hormones only. (B) Details of producible and unproducible TPPM with single GEM capacities (red) and combinations added-values with metabolic completions (green). TPPM and nutritional constraints are ordered in the same way as for panel A. Results are given for each nutritional constraint (black : poor; red : rich), without (plain) or with (bold) ARE. The list of nutritional constraints vary according to the considered TPPM, in accordance with the no-TPPM-in-media condition (i.e. growth media containing TPPM in their composition were excluded when necessary). On all figures axes, poor media are labeled in black, rich media in red, ARE-enriched media in bold and standard media in plain text.



Figure 6: Frequencies of GEMs in minimal combinations to produce TPPM. (A) 20th most occurring GEMs in all possible minimal combinations for all TPPM, only amino-acids, only vitamins, and only phyto-hormones. (B) Occurrences of GEMs in minimal combinations as a function of their genomes sizes. Text-annotated dots are the most occurring GEMs. There was no relation between a genome size and its frequency in combinations.

4 Discussion

4.1 Fundamental ecological niche signature in GEMs

The analysis of the PPM of the 193 GEMs of the *A. thaliana* culture collection demonstrated that phylogenetically related bacterial species are expected to share more similar metabolism than distant species. The first hypothesis (H1) was validated, and is consistent with the literature (Jaffe et al., 2020). Previous research attempted to predict ecological traits from genomic and metabolic information (Barberán et al., 2014; Alneberg et al., 2020), highlighting differentiation between taxonomic groups and hierarchical conservation within groups (Martiny et al., 2015). However finding fundamental niche signatures in GEMs remains a challenging task (Muller et al., 2018), and complementary trait-based approaches have been used to determine ecological attributes or correlate overlapping niches with phylogeny (Barberán et al., 2014; Fahimipour and Gross, 2020).

4.2 Phylogenetic distance, similarity and complementarity, antagonism and cooperation in SynComs

SynComs allowed a more diverse metabolism when they contain phylogenetically distant strains, our second hypothesis (H2) was validated, along with many unpredicted results (discussed below). Several studies used metrics similar to ours and obtained comparable results (Zeleznik et al., 2015; Russel et al., 2017; Hester et al., 2019; Lam et al., 2020), highlighting a correlation between metabolic similarity/dissimilarity and phylogenetic distance. Phylogenetically distant bacteria can be predicted to have less metabolic resource overlap and a higher potential to cooperate (Hester et al., 2019), while phylogenetically closer are expected to compete (Lam et al., 2020). Based on similarity and dissimilarity, other studies attempted to go further and interpret whether niche differentiation or habitat filtering were at play. Results are conflicting and research is consequently currently far from a consensus. For example, some studies found that species in the gut microbiome tend to co-occur more frequently with their competitors, thus highlighting habitat filtering (Levy and Borenstein, 2013), while it was shown that increased phylogenetic relatedness was correlated to competitive exclusion among bacterivorous protist species (Violle et al., 2011), favoring niche differentiation.

Rather than favoring niche differentiation or habitat filtering, SynComs unconstrained metabolism question the aggregation of most SynComs around a putative optimal phylogenetic distance, along with a metabolic diversity peak for smaller SynComs (figure 3A). The peak is

probably due to the taxonomic composition of SynComs with a high phylogenetic distance. These Syncoms all incorporate a Firmicutes or a Bacteroidetes (supplementary figure III A&B), which both display lower metabolic capacities than the other phyla, causing an inevitable decrease of PPM compared to other combinations. The aggregation of SynComs around the same phylogenetic distance is however more interesting. Closely related bacteria have similar needs, and thus encounter strong resource competition (Sung et al., 2017) despite many cross-feeding opportunities. On the contrary, very distant bacteria avoid competition, i.e. are adapted to different niches to colonize, and have a limited overlap in their needs. The observed optimal phylogenetic distance in our results could be interpreted as a trade-off between metabolic similarity and dissimilarity. Such optimum niche overlap would limit competition for resources among closely related species while being sufficient to promote exchanges of metabolites. as previously described in (Machado et al., 2021). However, the existence of microbial communities at both ends of the competitive-cooperation spectrum is also highlighted, mostly dependent on the environment (soil, free-living, or host-associated environments).

4.3 Metabolism is nutritional-constraint dependent

Sets of constraints were modeled using different growth media compositions. The third working hypothesis (H3), stating that the available nutrients have a deep impact on the metabolism of bacteria, was validated.

GEMs displayed low metabolic capabilities under severe growth constraints, alleviated by ARE which is in line with the nature of the genome collection used, isolated from *A. thaliana*'s roots. This result suggests that exudates from plant roots are important determinants determining the range of producible metabolites in the root microbiome. This was particularly visible for TPPM: the same set of TPPM was produced when ARE were added, regardless of the growth medium. This could either reflect the fact that our TPPM are core, essential compounds that most bacteria rely on and are supposed to metabolize. TPPM would create similar metabolic patterns among GEMs, activated by ARE.

These findings highlight the importance of nutrient availability for the bacterial community's metabolic profile (Erlandson et al., 2018). However, there is currently no consensus on the effects of all nutrients on microbial communities. For example, despite nitrogen being a strong predictor of metabolism, its enrichment has various impacts on soil microbiota's diversity and composition (Cui et al., 2021). Root Exudates are known to manipulate microbial communities but the effect of only a

few compounds were recently elucidated (Musilova et al., 2016; Voges et al., 2019; Jacoby et al., 2021).

The high difference observed between unconstrained and constrained metabolism could also exhibit the difficulty to correctly model an environment through the simulated nutritional constraints. Indeed, some TPPM were unproducible in any media, despite full completion rates within several GEMs (such as auxin and B12 vitamin). However, it could also reflect the fact that most growth media are unadapted to all these organisms, which many are known to be difficult to grow *in vitro* (Mas et al., 2016). Hence, metabolic cooperation would be especially relevant with such microorganisms. To parry *in vitro* culture issues, attempts of reverse ecology exist, for example by computing the nutrients set required by a metabolic network to produce biomass (Borenstein et al., 2008; Zelezniak et al., 2015; Laniau et al., 2017; Hester et al., 2019). Such approaches allowed the computation of overlapped and differentiated growth-requirements of several organisms, strengthening the study of ecological niche and metabolic interactions.

4.4 Metabolic dependencies are predicted to be major drivers of microbial communities structure

Metabolic exchanges were found to be essential to improve GEMs' metabolic capacities, both at the whole community and for combinations of a few GEMs, validating our fourth hypothesis (H4). However, we did not expect such high differences between poor and rich media, nor the counterintuitive effect of ARE on very poor media with many TPPM.

4.4.1 Metabolic exchanges are nutritional-constraint dependent and compensate severe growth constraints

Rich growth media and ARE unlocked more PPM and TPPM for single GEMs, with little metabolic cooperation required (figure 5). Indeed, growth constraints are likely weaker in rich media since most nutrients are available, thereby unlocking many reactions and their associated metabolic pathways. Reciprocally under poor media, i.e. severe growth constraints, single GEMs were not self-sufficient and more metabolic exchanges were required to produce some TPPM. This was visible by the high difference in the number of possible combinations of GEMs able to produce TPPM. However, when considering all TPPM at once on very poor media (M9, M63, mineral medium), the supplementation with ARE behaved differently and increased the number of predicted metabolic interactions between GEMs (figure 5A), which was unexpected because they are supposed to improve single GEMs autonomy. We concluded that under very severe nutritional constraints, an extremely small subset of GEMs are able to produce intermediate products and

cooperate. In this case ARE unlocked enough chemical reactions in other GEMs to compensate for the constraints, without permitting the full completion of TPPM's pathways, hence increasing the number of combinations. The combined effect of nutrients availability in soil and nutrients secreted by the host might then be a strong conditioner of metabolic interactions. In support of this statement, Klitgord and Segré (2010) found that there is always a way to predict a growth medium inducing metabolic interactions between pairs of seven species. However, they failed to predict a viable media for individual species, highlighting again the importance of cooperation. Finally, at the full community scale, the major compensation predicted by cooperation between all GEMs (figure 4C) also highlights the importance of metabolic cooperation to counter strong nutritional constraints.

4.4.2 Minimal combinations of GEMs reflect functional redundancy for targeted compounds

The results obtained herein allowed to predict the putative cooperation between GEMs able to produce relevant TPPM of the root microbiome (amino-acids, vitamins, phyto-hormones). For most TPPM categories, a number of putative combinations were returned when strong nutritive constraints, with or without ARE were applied. These results are coherent with previous research such as Frioux et al. (2018) and Thommes et al. (2019). The number of solutions predicted echoed recent research which predicted a large range of metabolites that can be secreted without cost, generating countless cross-feeding opportunities (Pacheco et al., 2019). Even if these results are strongly linked to the considered TPPM and constraints, it can be emphasized that combinations of bacteria are predicted to be able to co-metabolize to complete core, ubiquitous metabolic pathways. This observation can be interpreted as a community-level functional redundancy (Frioux et al., 2018) and as an insurance of completion of metabolic processes for a range of environmental constraints, for nutrition or interaction with host plants (Lebeis et al., 2015). These putative redundancies in metabolic completions might be key for stability processes under variable environmental constraints. Beside these core metabolic functions, it would be interesting to extend the analysis of minimal communities for secondary, more specific metabolism.

The software used, *Miscoto mincom* (Frioux et al., 2018), worked in a way where most parsimonious solutions only, the simplest combinations of GEMs fitting the applied constraints, are calculated. However, these putative solutions do not imply that more complex combinations shouldn't exist to produce the TPPM. Finally, we did not explore all solutions to make the distinction between mutualistic cross-feeding or one-way auxotrophies. One strain could be the

final producer of a TPPM by taking advantage of the secretion of another strain, without mutualistic exchanges, or several exchanges might be required. Both interacting behaviors likely coexist, even at the level of a single bacteria, depending on the other bacteria. Up to date the knowledge on these behaviors among co-existing bacteria is lacking. Additional hypotheses in the bacterial secretomes might fit better what is actually occurring in living systems.

4.4.3 Genomes size effects remains unclear

Bacteria with big genomes are usually considered as generalist species with wider niches (Sriswasdi et al., 2017). They indeed have a higher unconstrained metabolism (higher PPM) thus likely a higher probability to possess uncommon, important reactions involved in the production of the chosen TPPM. Antagonistic bacteria are also more likely to have larger genomes, linking antagonistic and generalist strategies (Russel et al., 2017). On the contrary, bacteria with reduced genomes are more likely to be involved in metabolic interactions due to their reduced set of reactions (Giovannoni et al., 2014; Gil and Peretó, 2015). In our results, bacteria with larger genomes exhibited both a higher metabolic similarity and complementarity than others (supplementary figure III C1&C2), highlighting a putative reservoir of functions which compensated for low phylogenetic distances, without many hints about their orientation towards cooperation or antagonism. The absence of correlation between GEMs' corresponding genomes sizes and their frequencies in putative metabolic interactions did not permit either to establish a link between generalist/specialist behavior and cooperation or auxotrophies provider.

4.4.4 Few strains are enough to reach the community's potential

As SynComs sizes increased, their metabolism quickly became similar (Figure 3 and S4). According to the results, the full community's unconstrained metabolism can be approximated with a few dozen GEMs only (in link with H2). In addition, most of the combinations for the production of TPPM involved a reduced set of GEMs. It echoes many studies which identified core microbiota composed of a reduced pool of species (Toju et al., 2018) to perform and/or optimize a biological function (Niu et al., 2017), analyze the impact of core strains on the whole microbiome (Bodenhausen et al., 2014; Lebeis et al., 2015), or to study host colonization processes (Bai et al., 2015). Other studies supported that a mirobiota's functional stability is kept regardless of the strains pick, as long as each functional group is picked (Louca et al., 2016).

5 Conclusions and prospects

Metabolic diversity and similarity were detected along genomes taxonomy. From the multi-genomes metabolic modelling analyses performed herein we predicted functional inter-dependencies and highlighted a long lasting ecological paradigm, the trade-off between competition and cooperation. We also found that putative metabolic interactions are numerous and constraint-dependent showing community level interlinkages, and cooperation permitting to buffer nutritional constraints. Such numerous interactions highlighted the importance of richness and diversity in microbial communities for community-level functioning. Altogether, these results bring hints of knowledge on the path to decipher microbial interactions in a microbiota applicable beyond the limits of the set of genomes used to produce the results.

In this view, deeper and more realistic genome-based modelling approaches mobilizing exchanges' costs, fluxes analysis, the use of a constant of nutrient depletion through time, would possibly provide a closer look on the community-level genomic toolbox to respond to constraints and to decipher evolutionary and behavioral responses to these constraints either under fluctuating or constant environments. All these approaches leading to putative responses will have to be combined to experimental validations. In this aim, the integration of multi-omics data, including secretome (Fritts et al., 2021), a current frontier of research is needed to break down padlocks on the interoperability of the data.

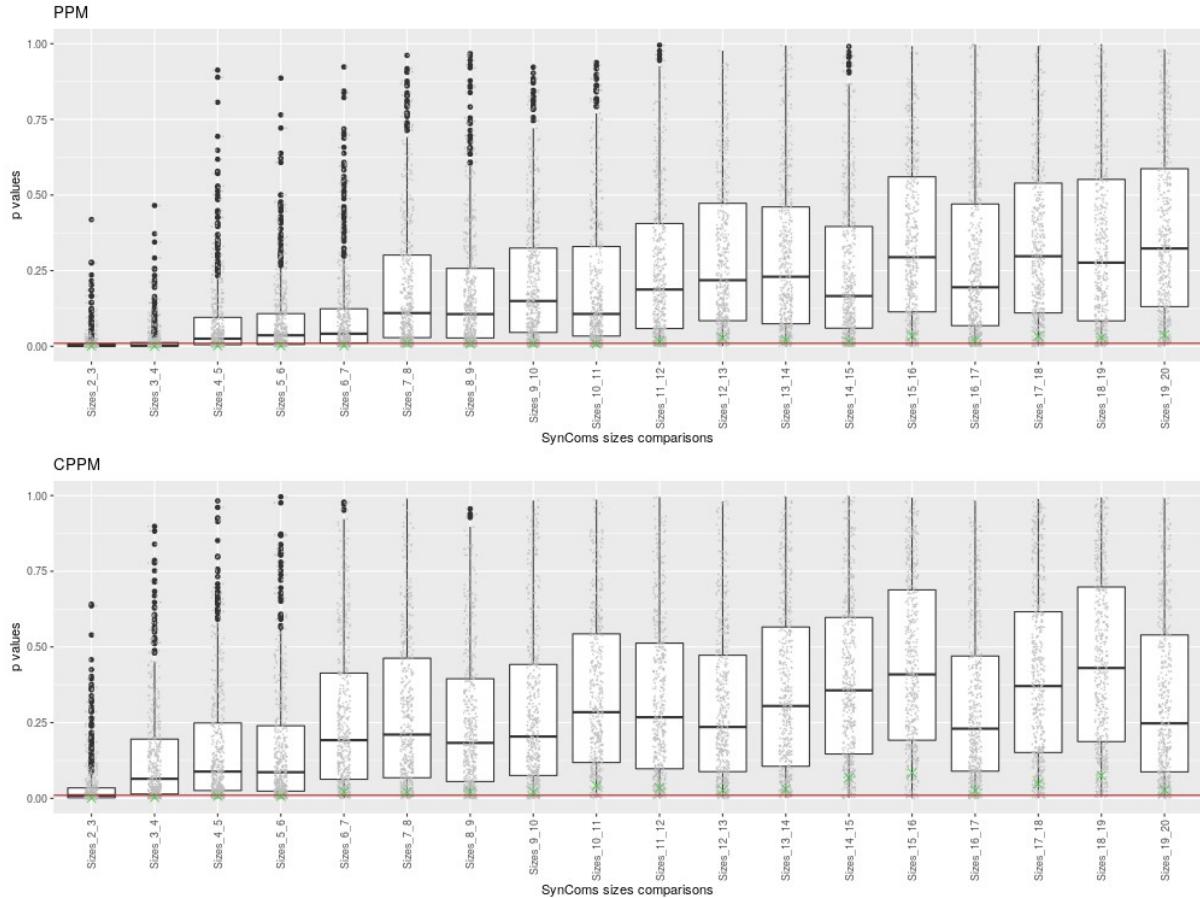
Acknowledgments and fundings

This work was supported by a grant from the French Ministry for Research and Innovation, by a grant from the CNRS (EC2CO), and also by a starting grant to S. Hacquard from a European Research Council (MICRORULES 758003), the Max Planck Institute, the Cluster of Excellence on Plant Sciences (CEPLAS), and the ‘Priority Programme: Deconstruction and Reconstruction of the Plant Microbiota (SPP DEC RyPT 2125)’, both funded by the Deutsche Forschungsgemeinschaft.

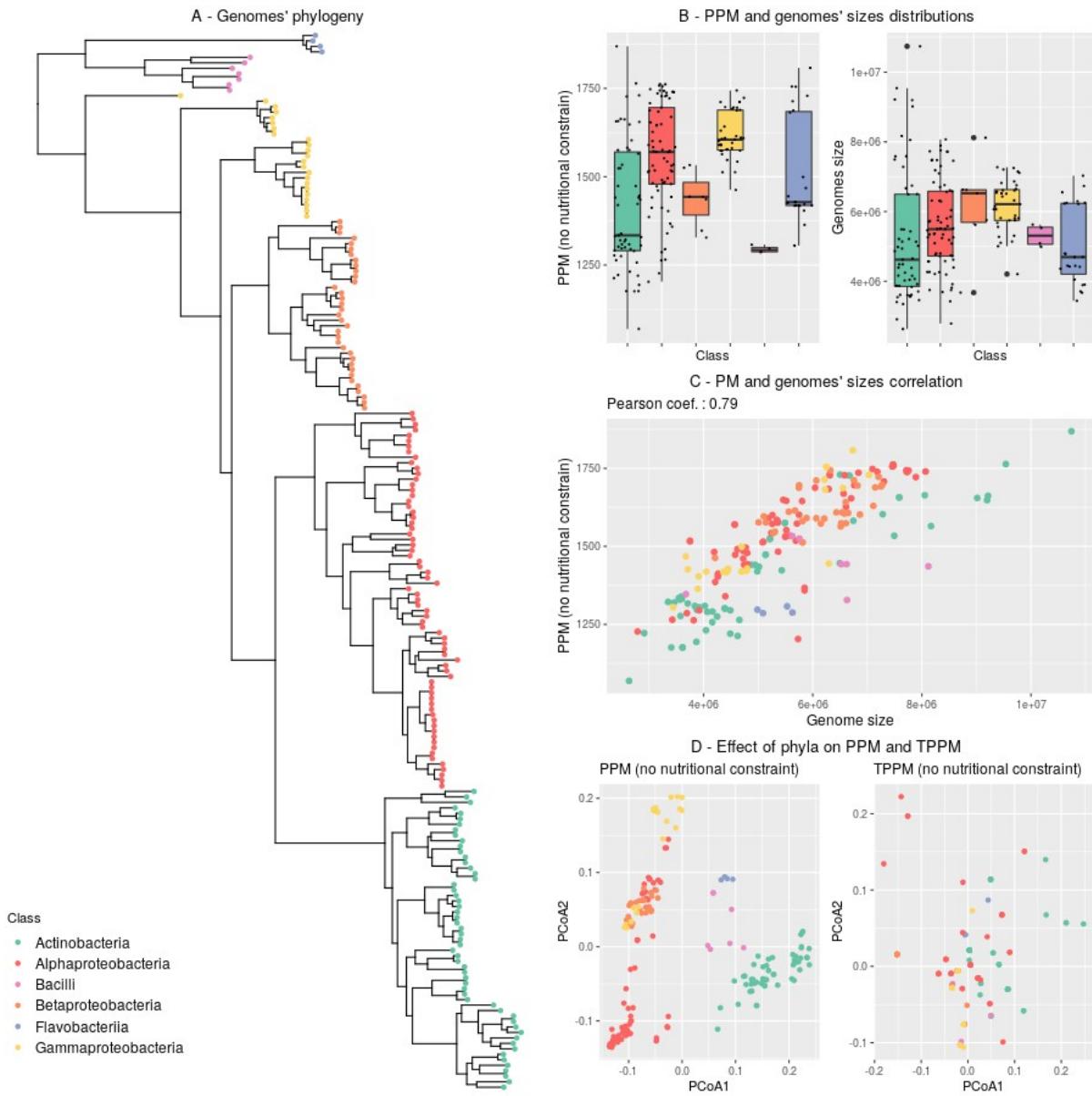
Competing interests

The authors declare no competing interests.

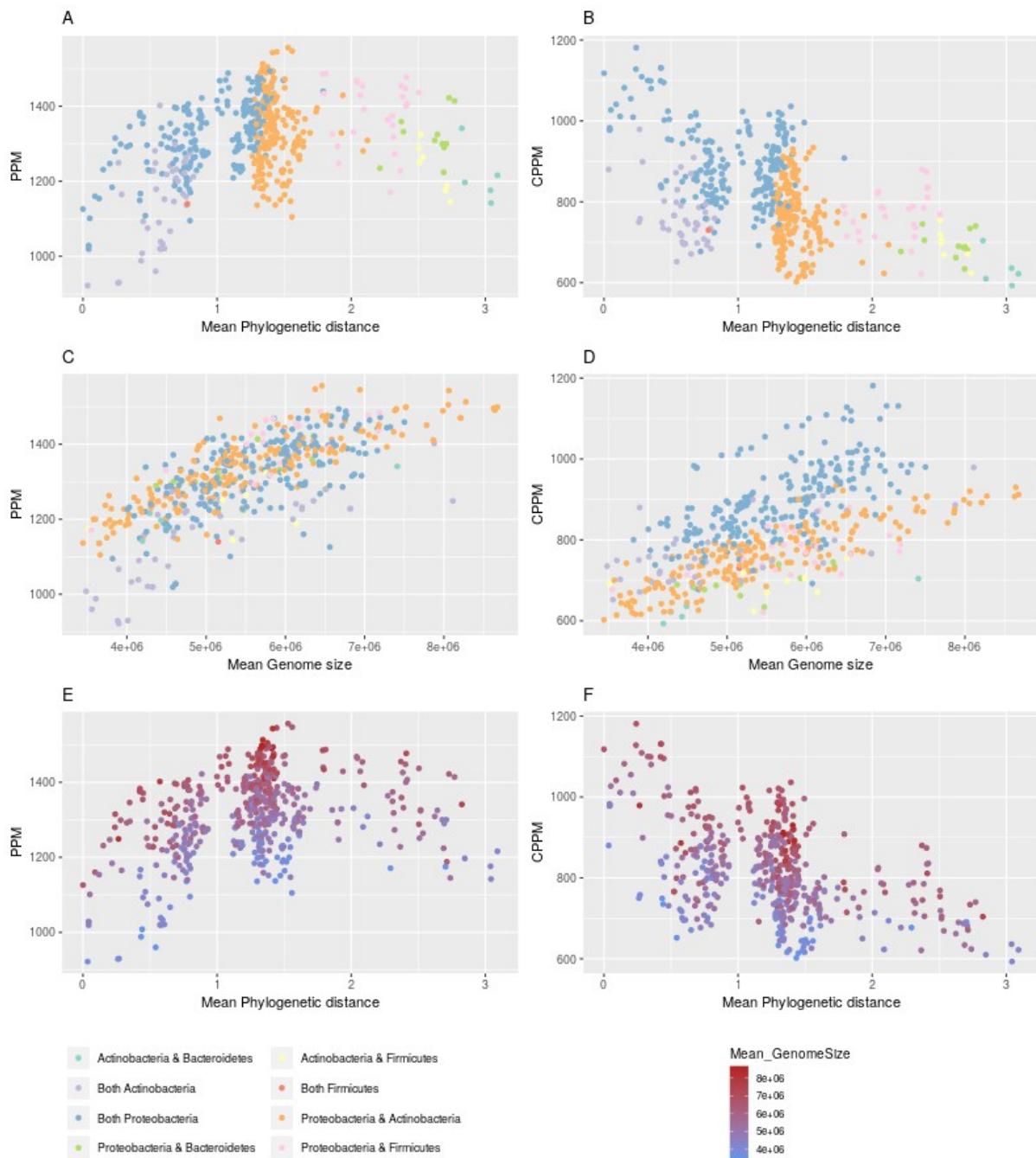
Supplements



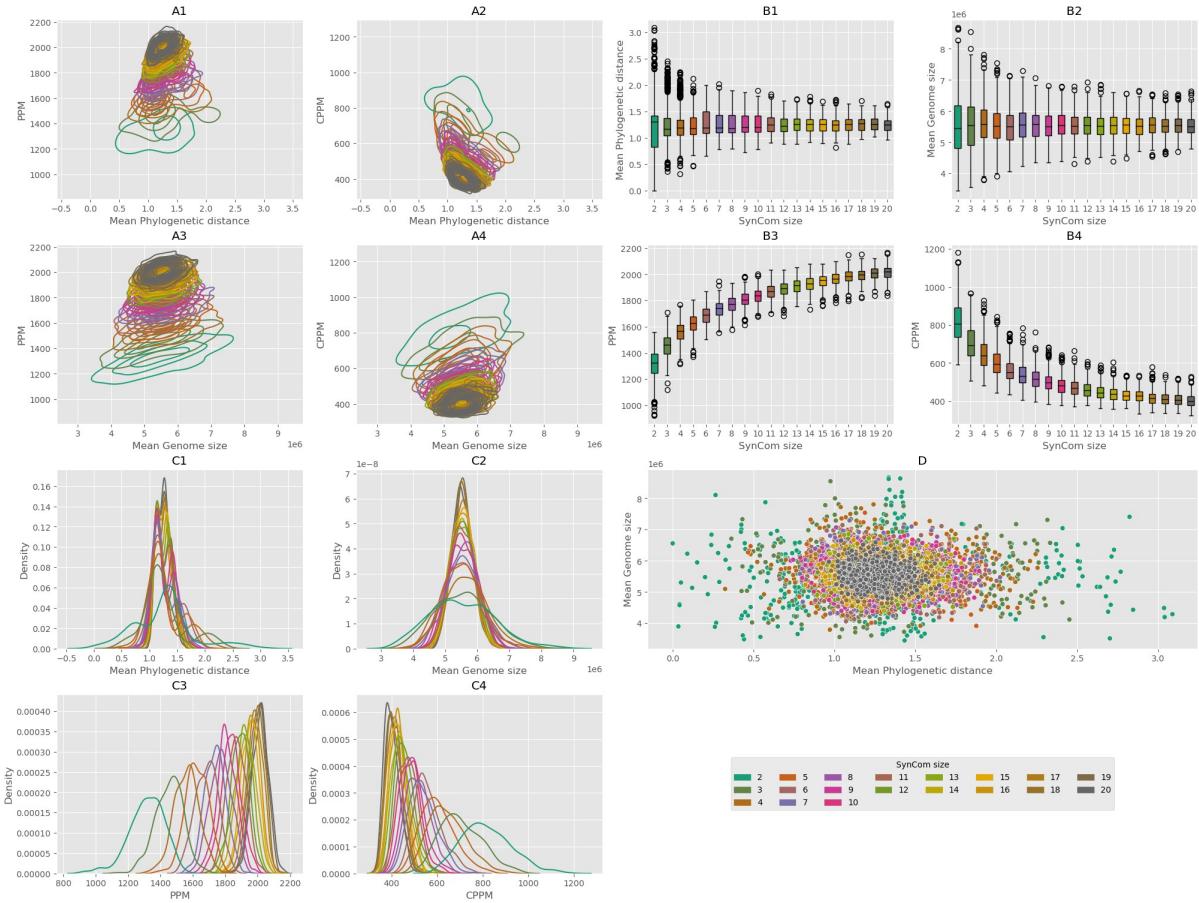
Supplementary figure I: Method for fixing the maximum Syncom size to inject into Poisson GLMs, we compared every $(s, s+1)$ pairs of SynComs' PPM and CPPM, s being a SynCom size in $[2, 20]$. For each $(s, s+1)$ pair, 200 pairs of random subsets of size $n=50$ SynComs were taken, and their PPM and CPPM were tested with Wilcoxon, Mann & Whitney tests. Boxplots of the 200 p values are displayed for all size comparisons and for PPM (top) and CPPM (bottom). Red lines are $p=0.05$ and green crosses are 1st quantiles. SynComs of 12 strains were chosen as a limit because it was the minimal size at which less than 10% of the p values were under 0.05 for PPM and CPPM.



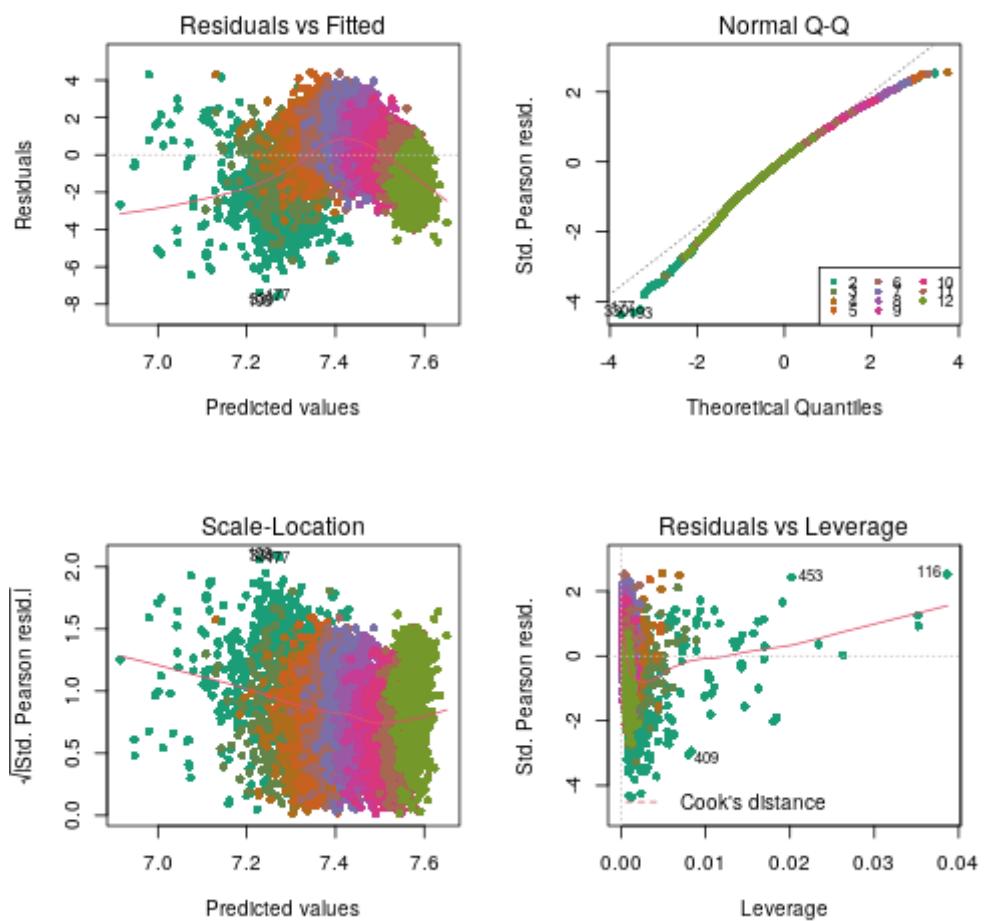
*Supplementary figure II: description of the 193 genomes collection from *A. thaliana* roots microbiota (A) Phylogenetic tree (maximum likelihood from a multi-alignment of AMPHORA genes) (B) Boxplots displaying the quantitative effect of class on genome sizes and PPM distributions. (C) Plots of the producible metabolites or TPPM as a function of genome sizes. (D) PCoA displaying the qualitative effect of phyla on metabolites' production (i.e. which compounds are produced by which taxa). Colors match strains classes.*



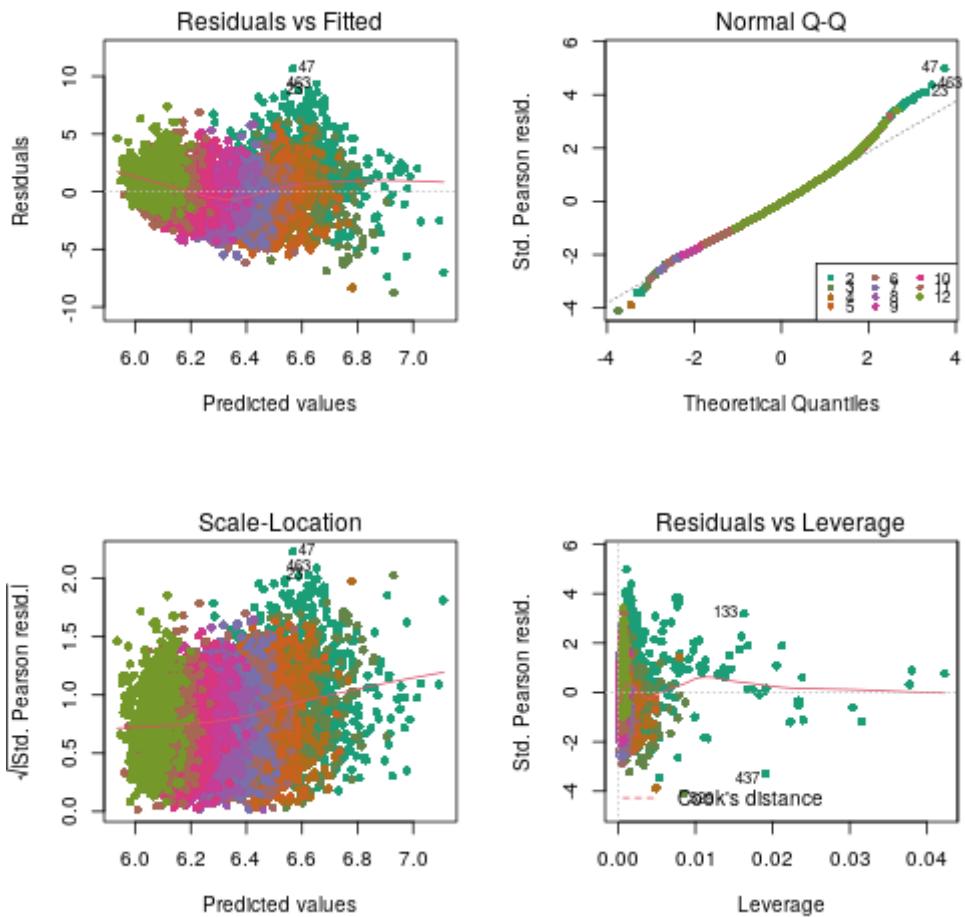
Supplementary figure III: Details of the correlations between explanatory metrics and response metrics for SynComs with two strains only. (A-B) Patterns of SynComs' taxonomic composition in the correlation of PPM and CPPM to phylogenetic distance. Proteobacteria and Actinobacteria combinations are responsible for the PPM peak. (C-D) there is also a taxonomic signal among the response of PPM and CPPM to genome sizes. (E-F) Effect of genome sizes on SynComs' PPM and CPPM. SynComs with a bigger average genome size have both bigger PPM and CPPM than SynComs with a smaller average genomes size. For each plot, only SynComs with 2 strains are represented



Supplementary figure IV: Details on phylogenetic and metabolic metrics. A1 to A4 : density plots show that for each SynCom size, most values are concentrated around a narrow range. B1 to B4 : Boxplots showing the reach of a plateau (in terms of values and/or variances) for each metric. Only data for SynComs' size below the plateaus were kept in regression models (size 2 to 12 strains). C1 to C4 : distributions of the different metrics, split by SynCom size. D : phylogenetic distance and genome size are not correlated, making their use as independent variables valid in the quasi-poisson regression.



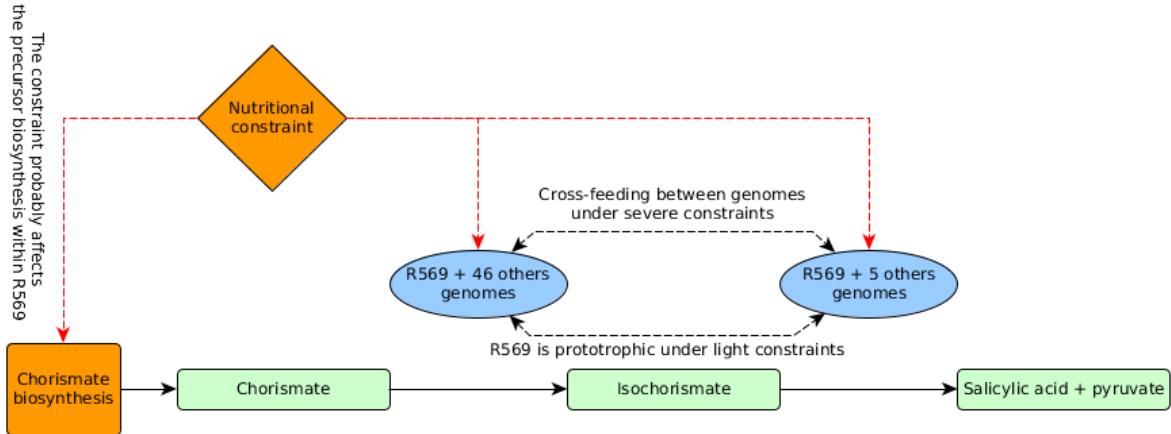
Supplementary figure V: Diagnostic plots of the quasi-poisson GLM modelling the response of PPM in SynComs.



Supplementary figure VI: Diagnostic plots of the quasi-poisson GLM modelling the response of the CPPM in SynComs

Supplementary table I : Taxonomy of most occurring GEMs in the putative combinations of GEMs permitting TPPM production through metabolic exchanges. The TPPM categories in which each GEM is the most involved are mentioned.

Identifier	Involved TPPM	Phylum	Class	Order	Family	Genus
Root11	All	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus
Root131	Amino-acids, vitamins	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus
Root569	Phyto-hormones (salicylic acid)	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Root401	Phyto-hormones (salicylic acid)	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Root329	Phyto-hormones (salicylic acid)	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Root68	Phyto-hormones (salicylic acid)	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Root71	Phyto-hormones (salicylic acid)	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Root381	Amino-acids	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	Bosea
Root483D 1	Amino-acids	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	Bosea
Root404	Vitamins (B3)	Proteobacteria	Betaproteobacteria	Burkholderiales	Rhizobacter	NA
Root52	Vitamins (B3)	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Paenibacillus
Root198D 2	Vitamins (B3)	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Duganella
Root336D 2	Vitamins (B3)	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Duganella
Root351	Vitamins (B3)	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Massilia
Root418	Vitamins (B3)	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Massilia



Supplementary figure VII: Supplementary figure 7 : A schematic view of the salicylic acid biosynthesis pathway. There are only two reactions, being possessed respectively by 47 and 6 GEMs, Root569 being the only GEM with a complete pathway. Under not constraining growth media, Root569 has the capacity to produce salicylic acid by itself, but lost this ability under severe nutritional constraints. In such cases, the set of strains has to exchange intermediate metabolites to produce salicylic acid. An hypothesis would be that Root569 is incapable of producing chorismate under severe growth constraints, while other strains are.

Chapter 3. Cross-feeding and predicted metabolic diversity promote coexistence in *A. thaliana* root microbiota

In this chapter, we attempted to correlate *in silico* patterns from the previous chapter with abundances of co-cultured bacterial strains of *Arabidopsis thaliana* root microbiome. 41 synthetic communities (SynComs) were designed along a gradient of predicted producible metabolites. SynComs size was set to 3 strains, according to the Answer Set Programming results of the previous chapter. A fourth, shared strain (*Mesorhizobium sp*) was added to all SynComs for a total of 4 strains. SynComs were cultivated in a liquid M9 minimal media enriched with Artificial Root Exudates, and samples were taken at four different timepoints (48h, 72h, 96h, and late 7 days timepoint). Strains monocultures were sampled after 7 days only. Among the SynComs with display apparent cross-feeding patterns, two were replicated, in a liquid M9+ARE media again then in a split-system device.

We hypothesized that SynComs with a bigger set of predicted producible metabolites would display more cooperation patterns (i.e. strains with better growth in SynComs than their monoculture) and less competition

Competition was quite common, with some strains dominating SynComs and others being apparently excluded, when compared to their monocultures. However, absolute abundances seemed to confirm that competition is reduced in SynComs with a high metabolic diversity, possibly due to less niche overlap. Enhanced growth of strains in SynComs was less common than expected, and without apparent link with the gradient of metabolic diversity. However there was a particularly interesting pattern involving an *Achromobacter sp* in three different SynComs.

The constant strain *Mesorhizobium sp* was a slow grower and was apparently excluded (i.e. very low abundances), but harbored growth in one SynCom. Replication of these cultures confirmed that *Achromobacter sp* was able to grow better when cultivated with other strains, likely due to an ability to feed on bacteria's secretion.

The split-system protocol suggested an ability of *Mesorhizobium sp* to profit from nutrient flow without contact to the provider(s), thus a possible avoidance of contact-dependent competition. When cultivating *Achromobacter sp* in the split-system, results suggested the same ability. However, these results need replication with more statistical robustness to be confirmed.

Cross-feeding and predicted metabolic diversity promote coexistence in *A. thaliana* root microbiota

Victor Mataigne^{1*}, Nathan Vannier², Philippe Vandenkoornhuyse^{1*} and Stéphane Hacquard^{2,3}

¹Université de Rennes 1, UMR 6553 ECOBIO, Rennes, France

²Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany.

³Cluster of Excellence on Plant Sciences (CEPLAS), Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany

*Corresponding author,
 victor.mataigne@univ-rennes1.fr
 philippe.vandenkoornhuyse@univ-rennes1.fr

Abstract

Deciphering the intricate network of interactions between microorganisms is crucial to understand the assembly rules of microbial communities. Using strains isolated from *A. thaliana*'s roots, we reconstructed metabolic networks and predicted *in silico* the metabolites producible by a range of microbial metabolic networks combinations. We then explored the outcome of 41 4-strains bacterial co-cultures representing a diversity gradient of predicted producible metabolites. A constant strain (*Mesorhizobium sp*) was included in every SynCom in order to analyze more precisely its interaction profile. Each strain was also cultivated alone in order to compare abundances in co-cultures and monocultures. SynComs with a high number of producible metabolites were expected to enhance strains growth through cross-feeding. In most SynComs, one or two strains were dominating in terms of abundances. Also, most SynComs seemed to harbor competition, with many strains harboring a reduced growth compared to their monocultures, some being potentially excluded because of very low abundances. A few other strains seemed unaffected and, interestingly, competition seemed to decrease when reaching the higher end of the gradient of metabolites diversity. Interestingly, some strains displayed an improved growth in SynComs compared to their monoculture, suggesting a dependency to the other SynComs members, but with no apparent link with the gradient. An *Achromobacter sp* strain, for which the results were replicated, appeared to benefit from cross-feeding in different SynComs, suggesting a low specificity to the provider. By replicating the results involving this strain, we identified that this growth benefit depended on the provider growth. The use of a dedicated growth system suggested that such a benefit could occur without contact through soluble compound(s) secreted in the environment, but remained to be statistically confirmed. The same protocol on the constant *Mesorhizobium sp* in a SynCom of particular interest suggested (without statistical confirmation)

that these strains responded differently, *Achromobacter* sp doing better in mixed co-culture, and *Mesorhizobium* sp being isolated from other strains, likely because it could benefit from the flow and avoid contact-dependent competition. Our results highlight the strong prevalence of competition in small co-cultures experiment but also the existence of strong dependency of some strains on cross-feeding. However, the difficulty to detect more cross-feeding with our initial protocol might be due to various issues, such as unmeasured antagonistic interactions. For instance, cross-feeding could be a more underlying interaction, softening competition, for instance to maintain a low-level metabolism. More sophisticated models and adapted experimental protocols were thus suggested as perspectives.

1 Introduction

Given the microbial community complexity, a current important question of microbial ecology is to decipher how ecological interactions as competition and cooperation shape the structure and dynamics of microbial communities (Boon et al., 2014; Hansen et al., 2020). After an initial consensus prioritizing competition, the relative importance of competition and cooperation have been regularly questioned (Mas et al., 2016; Coyte and Rakoff-Nahoum, 2019; Pacheco and Segrè, 2019). Cooperative processes were recognized as central to microorganisms' ecology despite inherent stability issues (West et al., 2006; McNally and Brown, 2016; Rakoff-Nahoum et al., 2016). Recently, the rapid development of -omics technologies allowed tremendous progress in microbial research, both for experimental and modeling approaches (Aguiar-Pulido et al., 2016; Hamilton et al., 2017; Li et al., 2017; Muller et al., 2018; Lloyd-Price et al., 2019; Jansma and Aidy, 2020). For instance, the reconstruction of organisms' metabolic networks with GEnome-scale metabolic Models (GEMs) from annotated genomes permit to predict metabolic exchanges (Muller et al., 2018). Also, the development of mathematical modelling in microbiology offers novel ways to model the evolutionary dynamics of cooperation (Lewin-epstein and Hadany, 2020). Such methods are notably employed with host-associated microbes and for the design of microbial factories with synthetic biology, targeting gut and plant microbiotas with applications for health (Heintz-Buschart et al., 2016) and sustainable agriculture (Ray et al., 2020).

Various approaches, including top-down, bottom-up experimental and *in silico* approaches tried to decipher the determinants of microbial communities assembly. These approaches are formalized as microbial system ecology (Mataigne et al., submitted), and provided massive descriptive data and identified many patterns and a few mechanisms involved in the organization of microbial

communities (Knight et al., 2018; Amor and Bello, 2019; Lawson et al., 2019). However, even if such results allowed massive steps forward into microbial ecology, we still miss detailed mechanistic understanding of ecological processes. Notably, different approaches regularly lead to contradictory results. For instance, *in silico* models generally predict many metabolic interactions through cross-feeding (Lawson et al., 2017; Smith et al., 2019), while standard experimental approaches with co-cultures generally struggle to validate cross-feeding or cooperation between microbial consortia, highlighting mostly competitive processes (Foster and Bell, 2012). Despite a few known cases of validated cross-feeding (Lin et al., 2018; Bui et al., 2019; Liu et al., 2019; Blasche et al., 2021), cooperative interactions remain difficult to highlight experimentally, except under particular environmental conditions or experimental protocols. Indeed, cross-feeding has been shown to occur only under some circumstances, such as various stress relievers (Liu et al., 2019) particular nutrient composition (D’Souza and Kost, 2016), or spatial configuration. Indeed, in other situations, stronger interactions can occur and mask cross-feeding (Goldford et al., 2018). For example, in (D’Souza and Kost, 2016), amino-acids auxotrophic *E. coli* strains had a significant fitness advantage in an amino-acids supplying growth media. Despite highlighting the importance of metabolic dependencies, these results did not confirm any metabolic exchanges between organisms, the auxotrophic strains relying only on the composition of their surrounding environment. Hence, in order to reconcile modelling and experimental approaches to find microorganisms effectively feeding on the secretions of others, studies are currently encouraged to combine modelling and experimental approaches. *In silico* models with dedicated metrics should inform hypothesis building with *in vitro* experiments as hypothesis validators (Lawson et al., 2019). Following this scientific ambition, we herein combined modelling and experimental approaches. We focused on a regularly cited pattern, stating that *in silico* models often predict that the metabolic diversity of a microbial community (i.e. producible metabolites) is correlated with high phylogenetic distance (thus strains’ dissimilarity) and is an indicator of potential cross-feeding and less competition between bacteria (Zelezniak et al., 2015; Russel et al., 2017; Hester et al., 2019; Lam et al., 2020). Cross-feeding would take place due to the high number of potentially exchangeable metabolites, and competition would be reduced because of less overlap of ecological niches due to high phylogenetic distances.

In this study, we considered small consortia of bacterial strains and tested how a simple metric as the number of predicted producible metabolites (‘PPM’) can be used to predict coexistence, and possibly cross-feeding, between bacterial strains. Hence, we hypothesized that SynComs with high PPM would allow a better co-existence between strains due to more cross-feeding interactions and

less overlapping niches, thus reduced competition. Strains benefiting from cross-feeding were expected to display higher abundances when cultivated in SynComs than cultivated alone ('monocultures'), while competing strains would display the opposite. SynComs with less competition were expected to harbor an improved global growth. Hence, we designed 41 combinations of 4-strains synthetic communities (SynComs) along a gradient of PPM (predicted by GEMs), which also follows phylogenetic relatedness. In total, 83 bacterial strains of a culture-collection representative of *Arabidopsis thaliana* root microbiota were used. We noticed however a lack of statistical power for some results, which incites carefulness for all interpretations.

2 Material and methods

2.1 Culture collection and genome-scale metabolic models

We used a culture collection of 193 strains of *Arabidopsis thaliana* root microbiome available at the Max Planck Institute for Plant Breeding research (Köln, Germany), for which annotated genomes were also available on www.at-sphere.com (downloaded on October 2018) (Bai et al., 2015). For each genome, a draft genome-scale metabolic model (GEM) was reconstructed (Mataigne et al., in prep) with an automated version of "Pathway Tools" (Karp et al., 2002, 2011, 2020) (the "mpwt" program from the metage2metabo tool suite (Belcour et al., 2020) and "AuReMe" (Aite et al., 2018). The reference database was MetaCyc (Caspi et al., 2008) (the default database, which contains almost 2500 curated metabolic pathways from many organisms). There was no gap-filling or manual curation procedure, since these organisms are relatively unknown, and we preferred to keep false negatives rather than introduce false-positives.

2.2 Construction of the predicted produced metabolites gradient

We generated 50,000 random SynComs of size N=4 strains using Python 3. Each strains had its reconstructed GEM, for which we had computed the number of compounds it can produce ("Predicted producible Metabolites" abbreviated in "PPM") with "Miscoto scope" (Frioux et al., 2018). PPM must be computed under a simulated nutritional constraint, which was a minimal media M9 with Artificial Root Exudates (ARE) supplements (supplementary file F1). Hence, SynComs' PPM were simply the union of its strains' PPM, plus the supplemental PPM obtained by metabolic exchange between strains (under the strong assumption that everything can be exchanged).

The PPM of the 50000 SynComs ranged from 112 to 481 metabolites and was then divided into ranks matching the 15-quantiles, in each 3 SynComs were picked (figure 1). We chose to use

quantiles instead of a linear division of the PPM range in order to have the same number of random SynComs per rank, facilitating the presence of a constant strain in each rank (i.e. present in every picked SynCom). This constant strain was referred to as Root695 (*Mesorhizobium sp*). The picked SynComs were also the closest possible to the minimum, median, and maximum phylogenetic distance of their gradient's quantile. Phylogenetic distances were computed with the Python package *ete3*, from a tree computed by (Bai et al., 2015). The SynCom size of 4 strains (3 strains + constant strain) was chosen according to previous modelling results (Mataigne et al., in prep), where many SynComs of 3 strains were predicted to be able to produce amino-acids by metabolic cross-feeding, in which the constant strain was added. SynComs composition within the gradient of predicted producible metabolic compounds were not fully random. SynComs containing strains displaying identical 16S rRNA were discarded in order to allow a correct mapping after sequencing (see below).

Four SynComs out of the 45 picks (supplementary file F2) were discarded post-sequencing due to contaminations, resulting in a final number of 41 SynComs mobilizing 83 bacterial strains (supplementary file F4). *In fine*, the chosen SynComs spread along a gradient from 235 to 438 PPM (figure 1).

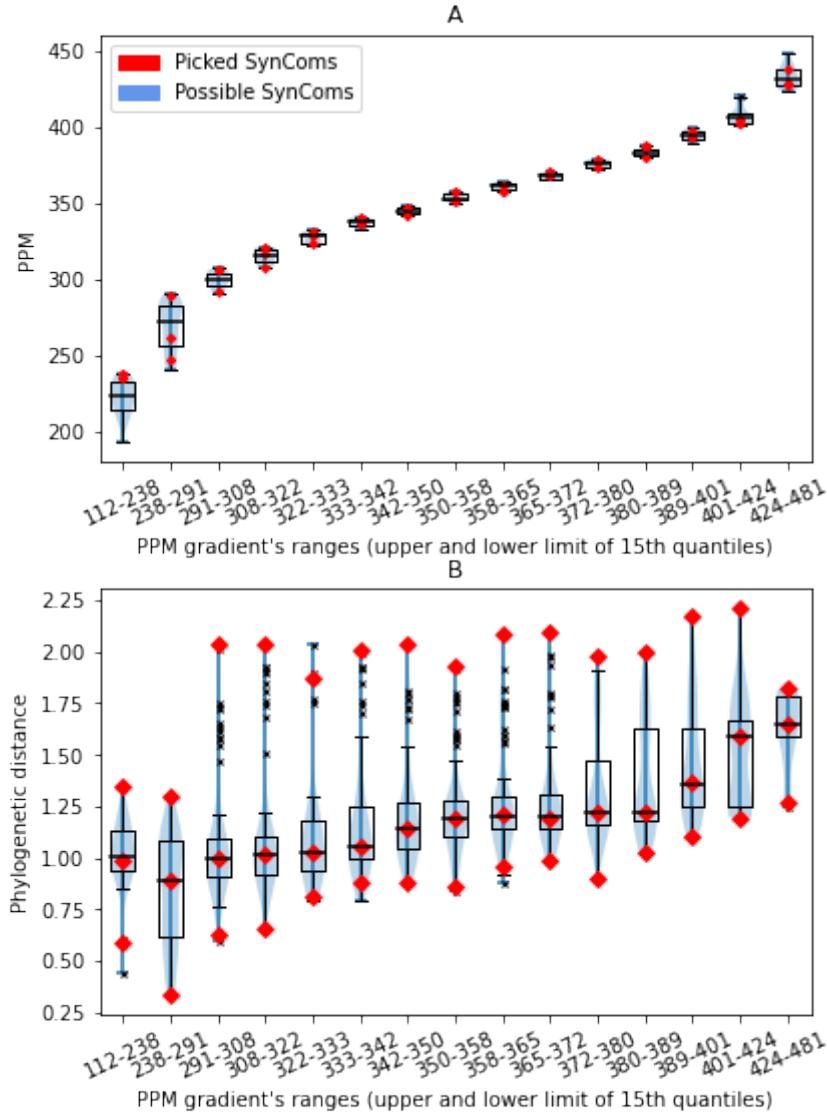


Figure 1: The gradient of PPM from which SynComs were picked. (A) The gradient of PPM was built on 50000 random Syncoms (each of size=4 bacteria, in blue). The gradient was divided into 15 ranks following the PPM 's 15th-quantiles to have the same number of random SynComs per rank. Three SynComs (red) were picked per rank. In fine, picked SynComs created a gradient going from 235 to 438 PPM. (B) Picked SynComs followed the minimum, median, and maximum of the distribution of phylogenetic distances of each rank.

2.3 Bacteria cultures

Individual strains were cultured in Petri dishes on 50% TSB with agar. After 48h, each SynCom was constructed by colony picking the four strains into 160 μ L of minimal M9 media amended with artificial root exudates (ARE, Baudoin et al., (2003), see composition in supplementary file F5). 10 μ L were transferred in 990 μ L of liquid culture (M9+ ARE), in triplicates (figure 2). Samples were disposed (in a random order) into an incubator at 20°C and 180 rpm. Sampling was done by taking 30 μ L, at 5 time points: 0h, 48h, 72h, 96h, and 7-days after inoculation. Each sample was

stored at -80°C. Monocultures of single strains were set up with a single replicate and a late 7-days time-point.

Four Syncoms found to display potential patterns of cross-feeding were replicated and also cultured in pairs. The experimental protocol was identical, except minor adjustments.

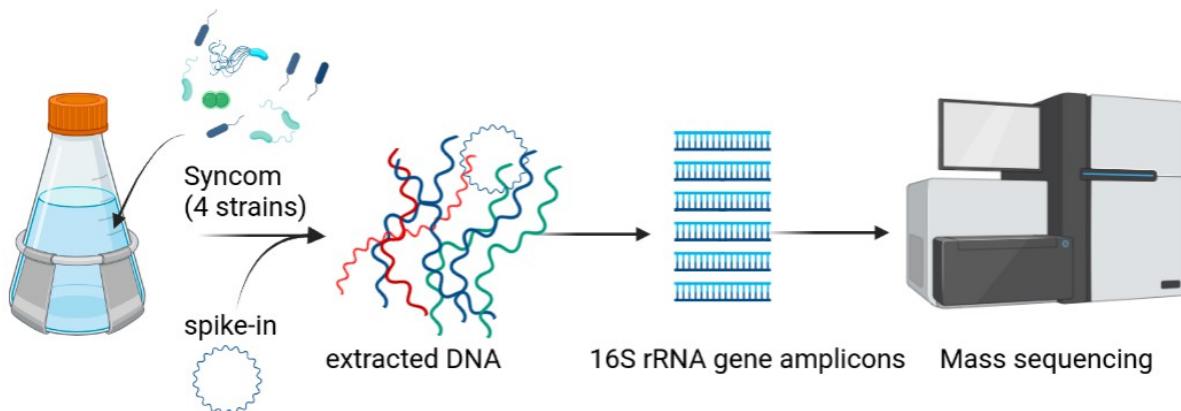


Figure 2: Experimental pipeline used to analyse the strains growth within each Syncoms. Quantitative analyses were made possible by the use of a reference (spike-in) added in each culture before DNA extractions. Figure drawn with BioRender.

2.4 Cultures in split-system

To validate the cross-feeding observed in four SynComs and to test whether it is dependent on contact between strains, we used a split-system device, made of two compartments separated by a semipermeable membrane (figure 3). The strains of interest (Root565 and Root695) were isolated in the first compartment, while the other strains of the SynCom were placed in the second compartment. A negative control was included by letting the second compartment empty (referred as “mock” throughout the text). The compartments were filled with 250ml of M9+ARE inoculated with each strain pre-cultivated in liquid 50%TSB cultures. TSB cultures were centrifuged and the pellet was washed three times with M9+ARE. The OD was adjusted to 0.2 with M9+ARE and 300µL of each strain adjusted culture was inoculated into the split-system. Each compartment was sampled at three time-points (24h, 48h, 72h) to measure optical density. .

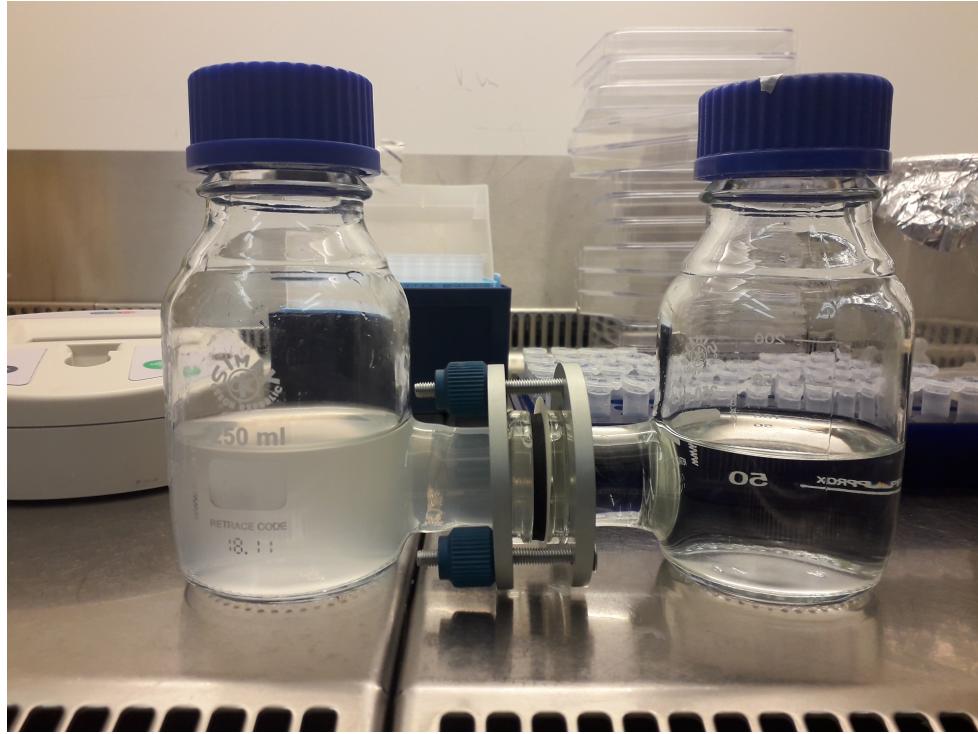


Figure 3 : Compartmented microbial cultures (split system device) to experimentally observe the consequence of not being in physical contact. Here we aimed at observing the importance of bacterial excretions for the growth of a given strain. The compartment to the left is separated from the compartment to the right by a semipermeable membrane. In the picture, one compartment contained a strain of interest (the receiver of the putative cross-feeding interaction, to the right) while the other contained the rest of the SynCom (the providers)

2.5 DNA extraction, purification, and sequencing

After transferring all samples on 96-well plates (10 µL per sample), DNA extraction was done with two buffers (15µL and 30µL, separated by 30min at 95°C). 5.82µL (10x more for the second set of cultures) of an artificial spike-in plasmid (Guo et al., 2020b) per 3.5mL of buffer was added. This spike-in plasmid allowed to have strains' abundances relative to a reference value, thus mimicking absolute abundances, making samples comparable for the data analysis. A first PCR was then performed with 1µL of each sample with 799F and 1192R primers (0.75µL both), with buffer (2.5µL), MgCl₂ 0.5µL), BSA 3% (2.5µL), dNTPs (0.5µL), BIORON DFS-Taq (0.4µL), H₂O (16.1µL) for a total mix of 25µL.

Digestion to remove primers and enzymes was done by adding a digestion mix to the PCR mix (for each plate-well) made of antarctic phosphatase (1µL), ExoI (1µL), and antarctic phosphatase buffer (3µL) and incubating at 37°C for 30min, then 85°C (enzymes deactivations) for 15min. The mix was centrifuged 10min at 3000rpm and 3µL of the supernatant was used for a second PCR to barcode the samples on the 2 strands (i.e. 1 barcoding PCR primer per plate and 1 barcoding primer

per well, resulting in a plate/well combination for each sample). Negative controls of PCR were verified with 5 μ L of each sample on gel electrophoresis.

Purification was performed with the Ampure magnetic bead kit. For each sample, the initial mix was 10 μ L of PCR product with 10mL of EtOH and 20 μ L of beads, incubated at room temperature for 5 min, then on a magnetic rack for 3 min followed by a washing step with 70 μ L of EtOH, for 1min (repeated one time). After removing EtOH and 5min drying, 50 μ L of nuclease free H₂O were added to each sample to elute the amplicons. The supernatants were collected.

After pooling all samples together at equimolar concentration, the amplicon sequencing library was sequenced using a 2x300 sequencing kit on a MiSeq (Illumina) sequencer following the manufacturer recommendations (figure 1).

2.6 Formatting sequencing data

Fastq MiSeq sequencing files were demultiplexed into a reads abundances table (number of reads of each strain+spike-in+contaminants) per sample (727 samples in total) with Python scripts, bash scripts, and usearch (Edgar, 2010). Fourteen samples with 0 spike-in reads were discarded. The 16S RNA sequences database was divided into databases per sample, each sample file being mapped on a database containing the expected SynComs and the spike-in. For each sample, sequences for which usearch returned no hits were mapped a second time, on the full database containing all the 83 used strains. The matched sequences were considered as contaminants. Sample counts were normalized according to spike quantity in the sample. After all the trimming and filtration steps, 541 samples of SynComs among which, 473 contained the constant strain) and 43 samples of monocultures were kept.

2.7 Data analysis

Abundances in reads of each sample were normalized to the spike-in abundance. The linear regression with `SynComs_total_abundance~SynComs_PPM` was performed at the 7-days timepoint, where all SynComs' were stable. The response variable was log2 transformed, and one datapoint was excluded because of a negative log value (the number of abundance reads was lower than the spike-in reads, thus inferior to 1). Statistical testing of the abundance difference of Root565 and Root695 in different SynComs or paired cultures were done with t-tests (the best suited test for very small samples; distributions were assumed to be normal). All statistical analyses were done in R.

3 Results

3.1 Reduced competition in SynComs with high PPM

Our hypothesis stated that SynComs with more PPM were more likely to coexist, compared to SynComs of lower PPM. The total abundances (4 strains at once) of SynComs (at the final time point) were significantly and positively correlated with their PPM ($p=5.5e^{-7}$, $R^2=0.215$, figure 4). Such a better global growth linked to high PPM suggests less nutrient competition between strains of SynComs with high PPM, thus less overlapping niche between phylogenetically distant strains (because PPM is correlated with phylogenetic distance).

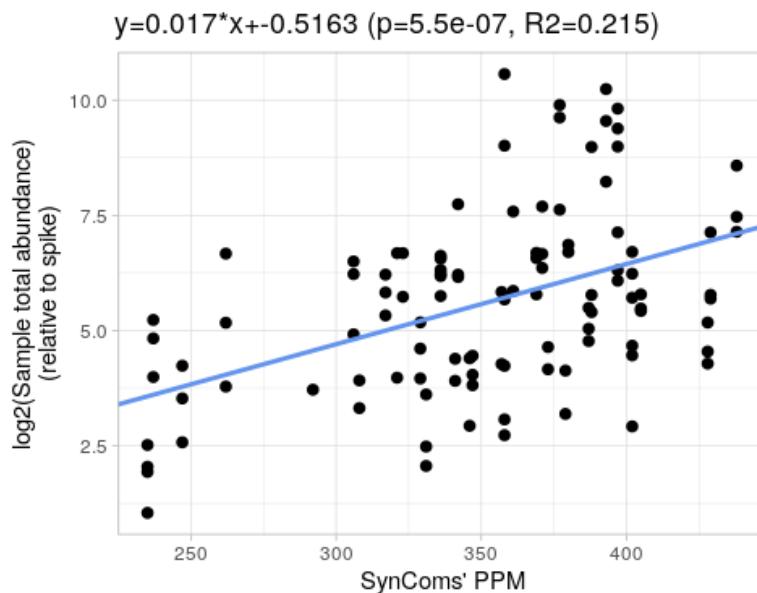


Figure 4 : Plot of Syncoms' abundances in reads (base 2 log-transformed) as a function of their PPM in the corresponding samples. All data are from the 7 days time-point. Because all replicates are independent, they are not summarized as means of triplicates.

3.2 Cases of cross-feeding identified among many competitive situations

Despite the softened competition highlighted in SynComs with high PPM, most Syncoms (examples : C27, C16, C24, C6, C22, etc., supplementary figure II) displayed dominance of one strain with high abundances and/or growth rates compared to the other SynCom' members. Four SynComs displayed two dominant strains (C37, C15, C44, C18, supplementary figure II). Dominated strains had very low abundances and growth rates close to 0. Unfortunately, contaminations caused the loss of about a half of monocultures, limiting co-cultures and monocultures comparisons. Among the tested strains that displayed growth, 17 grew relatively

equally in SynComs and monocultures, thus apparently insensitive to competition or cooperation (examples : Root918/C32, Root65/C16, Root935/C24+C2, Root186/C10, Root275/C4, etc., supplementary figure II). Compared to available monocultures, 11 strains (out of 43 monocultures) had lower abundances in SynComs (examples: Root901/C28+C43, Root68/C33, C3/Root420, Root1204/C35, etc., supplementary figure II). 30 strains did not grow and had abundances extremely close to 0 at all time-points. 5 of these strains harbored slightly better growth in their monoculture (supplementary figure II). Strains with improved growth in SynComs compared to their monocultures were rare (supplementary figure II). For example, Root558 in C6 displayed such a pattern, and was the only growing strain in the SynCom. Root267 in C22 and R935 in C44 also displayed such patterns, apparently neutral to other strains, but growths' standard deviations were too high to be sure.

3.3 An *Achromobacter sp* with an apparent benefit from cross-feeding in three SynComs

The most interesting case of improved growth was the one of Root565 (*Achromobacter sp*) in SynComs C37, C31, and C15 (supplementary figure II, figure 5 A&C&E). In SynCom C37, this strain grew better than in its monoculture (~2 times more), while the other most abundant strain (Root267 -*Acidovorax sp*-) was inhibited (~1.6 times less). Root565 growth was dramatically higher in C31 (~4 times more), and also possibly in C15 (~3 times more, but a doubt remained about this tendency because of a high standard-deviation). In C31, Root50 (*Sphingomonas sp*) was poorly abundant but unaffected. The absence of Root318D1 (*Variovorax sp*) monoculture in C31 did not permit to infer the outcome for this putative provider, and in C15 the other dominant strain, Root473 (*Variovorax sp*) did not seem affected. Additional cultures confirmed that Root565 had a strongly improved growth in all the three SynComs compared to its monoculture (figure 5A). Those growth differences were statistically significant in most cases and time-points (supplementary table 1).

Paired co-cultures indicated that all strains used as providers significantly improved Root565 growth (supplemental table 2), with Root473 (SynCom C15) permitting the maximum growth improvement (25 times more when comparing the means at the 96h time-point). Root267 and Root318D1 (both in SynCom C37) were the second most enhancing strains, followed by Root50 (SynCom C31, figure 5C). In each case, the differential growth increased at each time point. This preliminary experiment and the very interesting results obtained need to be confirmed in an additional independent experiment with more replicates. The abundance of Root565 was correlated

to the abundance of its paired co-cultured strain ($p=3.84e^{-20}$, $R^2=0.843$, with all replicates and all time-points, figure 5E). Indeed, Root267 and Root983 grew well on monocultures (figure 5F). Note that neither the pair Root565/Root29 and the monoculture of Root29 were analyzed because no reads of Root29 were detected.

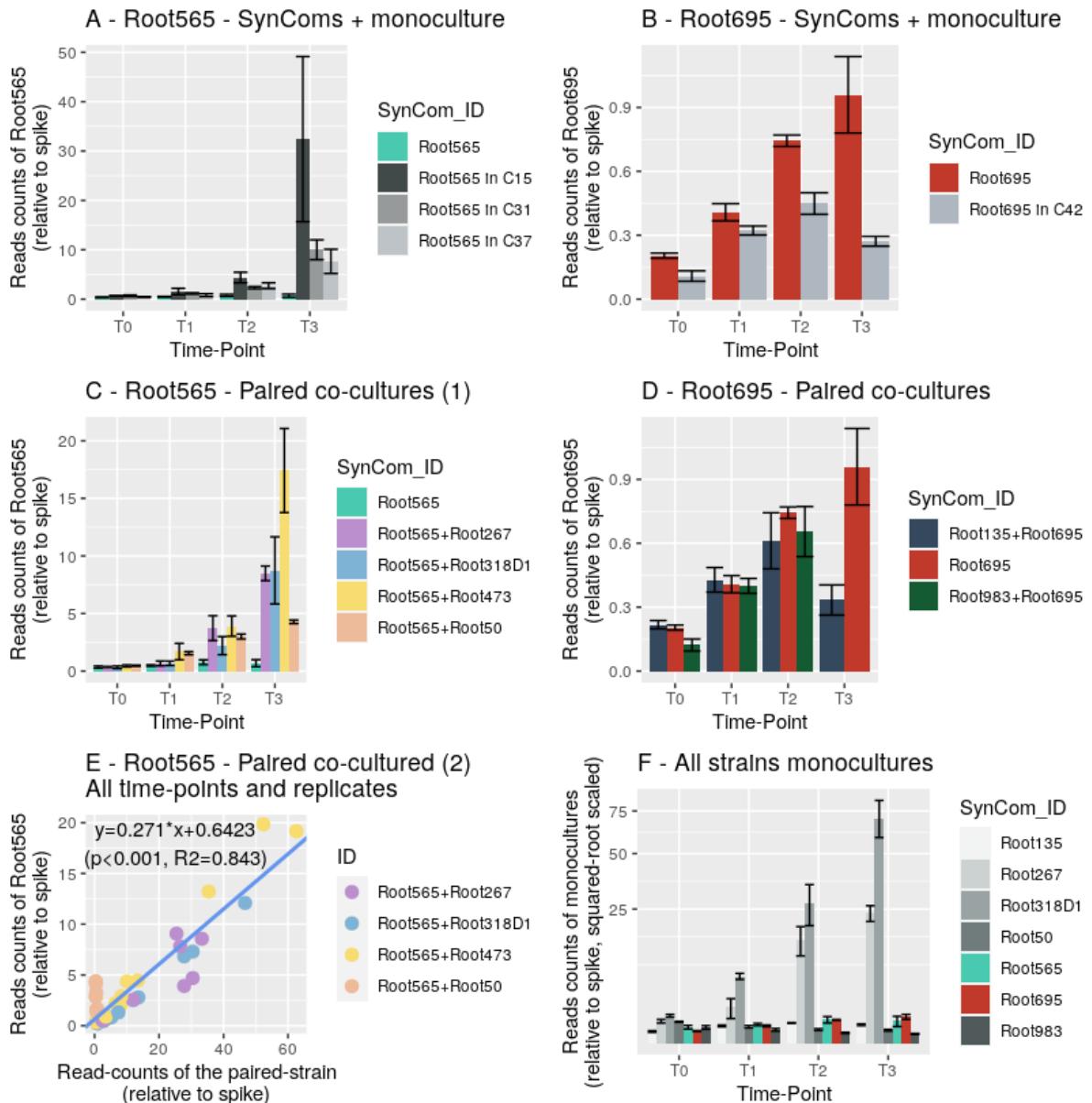


Figure 5: Abundances (in read counts relative to spike) of strains in the second set of cultures of SynComs C15, C31, C37, and C41 along time-points. A) Abundances of Root565 in SynComs, compared to its monoculture. B) Abundances of Root695 in SynComs, compared to its monoculture. C) Abundances of Root565 in all possible pairs of strains of SynComs C15, C31, C37 (except Root695). D) Abundances of Root695 in all possible pairs of strains of SynCom C42. E) Correlation between the abundance of Root565 (same as in C) with the abundance of its paired strain in co-cultures (all time-points and replicates). F) Abundances of each strain's monocultures (Root29 and Root473 were discarded because of contaminations).

In the split-system experiment, Root565 grew slightly better when associated with C31 and C37 compared to the control (figure 6A), but only the first time point (24h) was significant

(supplementary table 3). However, the insignificance of later time-points were likely due to an early saturation of the SynComs. C15 was dropped because of contaminations.

In fine, these results did demonstrate that Root565 is able to feed and grow by uptaking other strains' secretions, at least in homogeneous intermixing conditions. The growth improvement depended on the provider's growth rate. Root565 was also able to uptake nutrients secreted by the provider by a non-contact process, i.e. by a nutrient flow only.

3.4 The constant strain is a slow grower and a weak competitor that could benefit of non-contact secretions

The strain Root695 (*Mesorhizobium sp*) was added as a fourth member of each of the cultured SynComs, in order to discriminate strains with a different impact on its growth (enhancer or inhibitor). Root695's monocultures unfortunately got contaminated, strongly limiting this analysis. In most of the SynComs, Root695 exhibited very low abundances compared to the other strains, no growth or suppression (supplementary figure II and figure S2), suggesting a weak competition capacity combined to a weak growth capacity on M9+ARE media, potentially leading to competitive exclusion. The SynCom C42 was the only one where Root695 exhibited a clear growth pattern (supplementary file S6 and figure S2).

The SynCom C42 was cultivated again, as well as all possible pairs of Root695 and the SynCom's members. In this second set of co-cultures, Root695 still grew in the whole SynCom, but significantly less than its monoculture (TP1: $p=0.03$, TP2: $p=0.000876$, TP3: $p=0.00275$), with a decrease at the third time-point (figure 5B). When looking at pairwise co-cultures, only Root135 displayed a significant growth-inhibiting effect on Root695 at the third time-point ($p=0.00493$, other measurements were insignificant, supplementary table 4). Root29 was discarded because of no reads detected, and the last time point of the co-culture Root695-Root983 was contaminated) (figure 5D). In the split-system protocol, Root695 seemed to harbor an improved growth when isolated from the rest of the SynCom, benefiting only from the flow of their secretions (figure 6B). We however could not obtain a statistical difference (TP1: $p=0.374$, TP2: $p=0.205$, TP3: $p=0.141$) so these results remain observations to confirm.

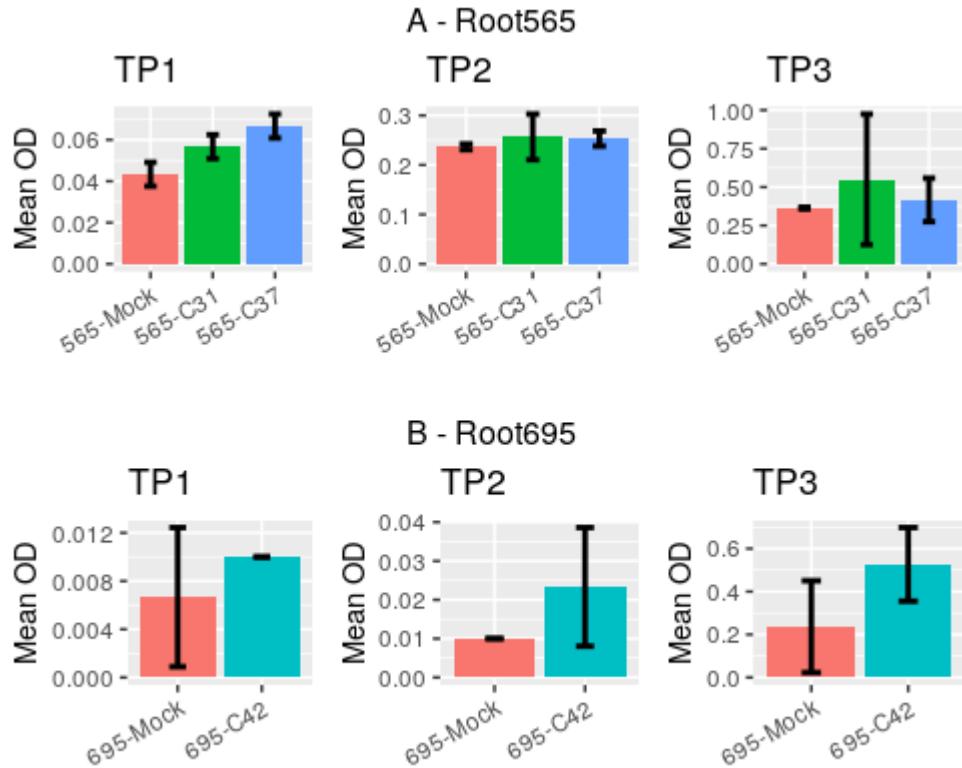


Figure 6 : Optical density in the split-system experiment, according to time-points and SynComs. (A) Cultures involving Root695 (Mock+SynCom C42). (B) Cultures involving Root 565 (Mock+SynComs C31, C37. C15 was discarded because of contaminants.

Hence, we observed that Root695 is a slow-grower and a weak competitor. We also observed some hints stating that Root695 could benefit from cross-feeding and avoid competition only if away enough from its competitors, by relying on the metabolites flow, but a replication of the experiments with more statistical power is needed. If these observations appeared to be correct, Root695 could benefit from other strains' secretion, but the benefit might have been erased by other factors, likely spatial or contact-dependent effects.

4 Discussion

Adopting a microbial system ecology point of view, this study aimed to confirm in silico predictions, we tested whether antagonistic or cooperative behaviors shaped the composition of 41 4-strains SynComs built with 83 strains derived from *A. thaliana*'s root microbiota. SynComs were built along a gradient of PPM, a metric often assimilated to cross-feeding opportunities and less competition, and the abundance of each strain in every SynComs was compared to its abundance when cultivated alone ("monoculture"). Cross-feeding and competition were investigated with

comparisons of strains' abundances in SynComs and monocultures. In addition, SynComs' total abundances served as a milestone for the strength of competition along the gradient.

4.1 The prevalence of antagonistic interactions

In most co-cultures, one strain appeared to dominate the others in terms of abundance. Abundances of strains in SynComs were most often lower than in their corresponding monoculture. In accordance with previous studies (Hibbing et al., 2010; Foster and Bell, 2012; Romdhane et al., 2021), such results highlight a high prevalence of antagonistic interactions between SynCom's members. Regular cases of competitive exclusion in the SynComs suggest that competition is mostly responsible. However, since half of monocultures were missing, we might miss many interesting patterns.

However, the higher total abundances of strains in SynComs with higher PPM (figure 4) can be interpreted as a reduced competition due to metabolic dissimilarity along with phylogenetic distance, as regularly mentioned in the literature, in accordance with our prediction (Zelezniak et al., 2015; Russel et al., 2017; Hester et al., 2019; Lam et al., 2020). In such SynComs, niche differentiation might be at play (Jacoby et al., 2018). However, we noticed that some studies take the counterpart of this statement. For example, (Simonet and McNally, 2021) stated that genomic relatedness is correlated with cooperative genes content in gut-microbiome genomes. In (Griffin et al., 2004), a cooperative trait (the production of siderophores), was demonstrated to be enhanced by relatedness between bacteria.

4.2 Some strong cross-feeding patterns for *Achromobacter sp*

In our study, we detected some strains with an improved growth in SynComs rather than in monocultures, notably for the strain referred to as "Root565" (*Achromobacter sp*) (figure 3A&C). According to the results of additional paired co-cultures, this strain was able to uptake the secretions of all of its co-cultured strains. The growth enhancement of Root565 was correlated to the growth capacity of its pair (figure 3E), but did not permit to know if Root565 is particularly adapted to *Variovorax*, *Acidovorax*, and *Sphingomonas* species or harboring a generalist, non-species specific cross-feeding. The involved compounds remain to elucidate, and might be either diverse or specific. Similarly, some bacteria from *A. thaliana* root microbiota were characterized to uptake efficiently root exudates, and could consequently secrete used as supplemental substrate by other strains (Jacoby et al., 2018). *Achromobacter* species are frequent and dominant in plant microbiomes (Yadav et al., 2017; Lata et al., 2019), and some species are known to be involved into their host's

drought stress tolerance (Hussain et al., 2018; Jha, 2020; Khoshru et al., 2020). Regarding the growth of Root565 when feeding on others' secretions, further research should focus on providers' growth rates, providers' diversity (i.e. specific or generalist cross-feeding), and secretomes to build models explaining precisely this case of cross-feeding, for instance with fluxes balance analysis (Orth et al., 2010; Douglas, 2020) and exometabolomics. A research goal would be to identify what are the shared and common compounds excreted by the different providers and how Root565 can uptake them, particularly waste metabolites (Goldford et al., 2018).

4.3 Cross-feeding could exist in many cases but might be masked by antagonistic interactions

The metric we used, PPM, is voluntarily simple, but might not be suited to investigate cross-feeding efficiently. Indeed, a complex community's interaction profile might not be well-represented by such a simple metric. Moreover, the apparent high prevalence of antagonistic interactions in our SynComs does not disprove the existence of cross-feeding interactions. Indeed, competition for other nutrients can still occur between microorganisms exchanging compounds (Hillesland and Stahl, 2010), or cross-feeding partners could still be competing for the shared nutrients or on other aspects (McCullly et al., 2017b). Cross-feeding could be then an underlying process, useful to soften competition or to maintain a low-level metabolism of dormant organisms, which could explain the very low abundances of many microorganisms species in microbiomes, such as the constant strain Root695 (Nemergut et al., 2013; McCullly et al., 2017b, 2017a; Goldford et al., 2018; Fritts et al., 2021). Such a case would also explain the difficulty to detect cross-feeding without appropriate methods. Nevertheless, the capacity of a bacteria to feed on organisms' secretions have been repeatedly observed experimentally. For example, Goldford et al., (2018) found that all isolates of a relatively large community were able to grow on a media composed of the secretions of other isolates. They discovered that such cross-feeding stabilized competition, in a collective interactions network rather than with pairwise interactions.

4.4 The importance of spatial configuration and metabolites flow

Connex to the aforementioned ability to feed on others' secretions is the spatial configuration of organisms and the flux of secreted metabolites in the environment. Indeed, It was demonstrated that spatial segregation between symbionts (i.e. via compartmentalization provoked by the host, biofilms ...) is a way to reduce competition over a shared resource such as public goods (Keymer et al., 2008; Chomicki et al., 2020). Our main experiment mixed strains in liquid cultures, without any

heterogeneity such as spatial structure. Consequently, such protocol might explain the high prevalence of competitive behaviors, for example because of contact-dependent inhibition of growth (Ikryannikova et al., 2020), in addition to resource competition.

We attempted to determine if Root565 and Root695 could benefit from the flow of other strains' secretions, without contact-dependent mechanisms, with the use of a split-system device. Root565 seemed to perform less, while Root695 seemed to perform better than in liquid, mixed co-cultures. However, the lack of statistical power could not highlight a significant difference, so this new hypothesis remains to be confirmed. If confirmed, such results could strongly highlight the capacity of some bacteria to feed on others' secretion, with or without contact-dependent mechanisms, according to the bacteria's strategy. Such a contact-independent ability would be considered as a way to profit from the provider's secretions while avoiding contact-dependent competition for other environmental variables that can bypass the benefit of cross-feeding.

5 Perspectives

Competition and cooperation might be intricate interactions, and the observed experimental outcomes might strongly depend on how traits, metrics, spatial scale, and experimental protocols are combined, making it difficult to return general patterns. We identified several ways to strengthen or extend our results. Syncoms combinations from the strains previously listed as potentially involved in cross-feeding could be tested again, possibly more extensively, as what has been done for Root565. The modelling and experimental validation of a steady state reached by competition or cross-feeding could be promising. The importance of spatial-scale and nutrient flow in cross-feeding interactions should be further explored, notably to determine what and how metabolites can flow in the environmental matrix, or if many species can avoid contact-dependent competition and still profit from public goods. Refined *in silico* models could be investigated, for instance with Flux Balance Analysis could provide better predictions or testable hypotheses. We could also have crossed PPM with previous results (Mataigne et al., in prep) based on combinatory analytics with Answer-Set-Programming, which returned putative combinations of cooperative strains. However, such combinations were extremely numerous and oriented towards the production of particular targeted compounds, not on strains abundances, which would have imply a more sophisticated experimental design. In addition, more robust predictions could be obtained with approaches using functional-trait rather than phylogeny (Boon et al., 2014). However such approaches rely on a better knowledge of organisms' ecology, data we do not dispose of yet. Finally, a way to improve

experimental set-ups for finding cross-feeding relationships would be to prioritize exometabolomics studies, as well as bacteria's uptake capacity of nutrients.

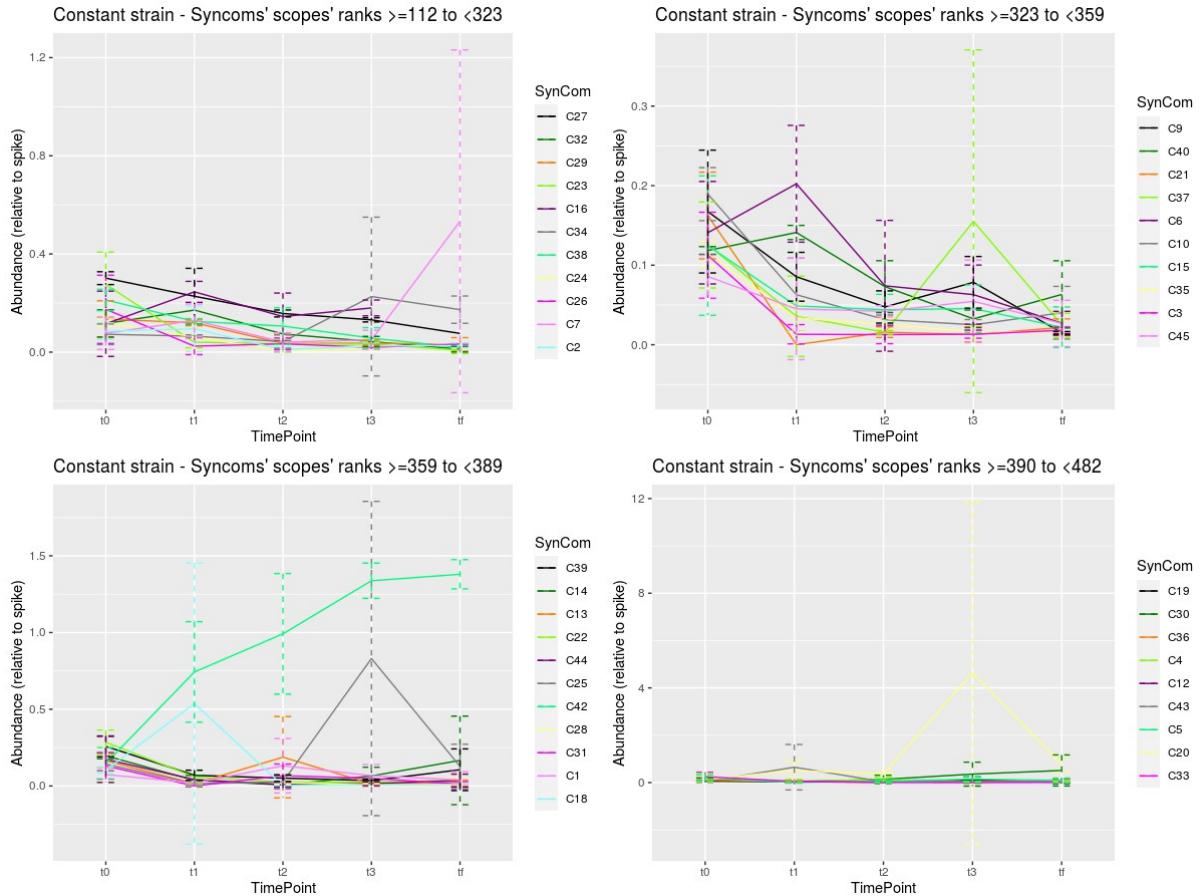
Acknowledgments and fundings

This work was supported by a grant from the French Ministry for Research and Innovation, by a grant from the CNRS (EC2CO), and also by a starting grant to S. Hacquard from a European Research Council (MICRORULES 758003), the Max Planck Institute, the Cluster of Excellence on Plant Sciences (CEPLAS), and the 'Priority Programme: Deconstruction and Reconstruction of the Plant Microbiota (SPP DECRyPT 2125)', both funded by the Deutsche Forschungsgemeinschaft.

Competing interests

The authors declare no competing interests.

Supplements



Supplementary figure I : A closer look at the constant strain's abundances (Root695) in all Syncoms. All SyncComs' curves are split in four panels (following the ranks of PPM) for visibility purposes. The SyncCom C42 was the only one where Root695 exhibited a clear growth pattern.

Supplementary table 1: t-tests for testing the difference of Root565 abundances in monocultures and SyncComs (liquid cultures) C15, C31, C37.

TP1			
<i>group1</i>	<i>group2</i>	<i>p</i>	<i>p.signif</i>
Root565	Root565_in_C15	0.00643	**
Root565	Root565_in_C31	0.525	ns
Root565	Root565_in_C37	0.303	ns
TP2			
<i>group1</i>	<i>group2</i>	<i>p</i>	<i>p.signif</i>
Root565	Root565_in_C15	0.000114	***

<i>Root565</i>	<i>Root565_in_C31</i>	0.0169	*
<i>Root565</i>	<i>Root565_in_C37</i>	0.00404	**
TP3			
<i>group1</i>	<i>group2</i>	<i>p</i>	<i>p.signif</i>
<i>Root565</i>	<i>Root565_in_C15</i>	0.00184	**
<i>Root565</i>	<i>Root565_in_C31</i>	0.217	<i>ns</i>
<i>Root565</i>	<i>Root565_in_C37</i>	0.346	<i>ns</i>

Supplementary table 2: pairwise t-tests for testing the difference of *Root565* abundances in monocultures and paired co-cultures from SynComs' members (liquid cultures) C15, C31, C37.

TP1			
<i>group1</i>	<i>group2</i>	<i>p</i>	<i>p.signif</i>
<i>Root565</i>	<i>Root565+Root267</i>	0.552	<i>ns</i>
<i>Root565</i>	<i>Root565+Root318D1</i>	0.519	<i>ns</i>
<i>Root565</i>	<i>Root565+Root473</i>	0.00162	**
<i>Root565</i>	<i>Root565+Root50</i>	0.00372	**
TP2			
<i>group1</i>	<i>group2</i>	<i>p</i>	<i>p.signif</i>
<i>Root565</i>	<i>Root565+Root267</i>	0.000549	***
<i>Root565</i>	<i>Root565+Root318D1</i>	0.0357	*
<i>Root565</i>	<i>Root565+Root473</i>	0.000356	***
<i>Root565</i>	<i>Root565+Root50</i>	0.00369	**
TP3			
<i>group1</i>	<i>group2</i>	<i>p</i>	<i>p.signif</i>
<i>Root565</i>	<i>Root565+Root267</i>	0.00111	**
<i>Root565</i>	<i>Root565+Root318D1</i>	0.000882	**
<i>Root565</i>	<i>Root565+Root473</i>	0.00000211	*****
<i>Root565</i>	<i>Root565+Root50</i>	0.064	<i>ns</i>

Supplementary table 3: pairwise t-tests (with Holm correction) for testing the difference of Root565 abundances in the split-system device.

TP1			
<i>group1</i>	<i>group2</i>	<i>p</i>	<i>p.signif</i>
565-C31	565-Mock	0.03	*
565-C37	565-Mock	0.00258	**
TP2			
<i>group1</i>	<i>group2</i>	<i>p</i>	<i>p.signif</i>
565-C31	565-Mock	0.42	ns
565-C37	565-Mock	0.498	ns
TP3			
<i>group1</i>	<i>group2</i>	<i>p</i>	<i>p.signif</i>
565-C31	565-Mock	0.41	ns
565-C37	565-Mock	0.809	ns

Supplementary table 4: pairwise t-tests for testing the difference of Root695 abundances in monocultures and paired co-cultures from SynComs' members (liquid cultures) C42.

TP1			
<i>group1</i>	<i>group2</i>	<i>p</i>	<i>p.signif</i>
Root695	Root135+Root695	0.597	ns
Root695	Root983+Root695	0.839	ns
TP2			
<i>group1</i>	<i>group2</i>	<i>p</i>	<i>p.signif</i>
Root695	Root135+Root695	0.168	ns
Root695	Root983+Root695	0.332	ns
TP3			
<i>group1</i>	<i>group2</i>	<i>p</i>	<i>p.signif</i>
Root695	Root135+Root695	0.00493	**

Supplementary figure(s) II : abundances plots per SynCom

Supplementary file F1 : sbml configuration file modeling the M9+ARE media

```

<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level2" level="2" version="1">
<model>
  <listOfSpecies>
    <!-- M9 -->
    <species id="M_NA_43_c" name="Na+" compartment="c"/>
    <species id="M_Pi_c" name="phosphate" compartment="c"/>
    <species id="M_CL_45_c" name="Cl-" compartment="c"/>
    <species id="M_K_43_c" name="K+" compartment="c"/>
    <species id="M_MG_43_2_c" name="Mg2+" compartment="c"/>
    <species id="M_SULFATE_c" name="sulfate" compartment="c"/>
    <species id="M_AMMONIUM_c" name="ammonium" compartment="c"/>

    <!-- ARE -->
    <species id="M_Glucopyranose_c" name="D-glucopyranose" compartment="c"/> <!-- glucose 1 -->
    <species id="M_ALPHA_45_GLUCOSE_c" name="α-D-glucopyranose" compartment="c"/> <!-- glucose 2 -->
    <species id="M_GLC_c" name="β-D-glucopyranose" compartment="c"/> <!-- glucose 3 -->
    <species id="M_BETA_45_D_45_FRUCTOSE_c" name="β-D-fructofuranose" compartment="c"/> <!-- fructose 1 -->
    <species id="M_CPD_45_15382_c" name="keto-D-fructose" compartment="c"/> <!-- fructose 2 -->
    <species id="M_SUCROSE_c" name="sucrose" compartment="c"/>
    <species id="M_CIT_c" name="citrate" compartment="c"/>
    <species id="M_Lactate_c" name="lactate" compartment="c"/> <!-- lactate 1 -->
    <species id="M_D_45_LACTATE_c" name="(R)-lactate" compartment="c"/> <!-- lactate 2 -->
    <species id="M_L_45_LACTATE_c" name="(S)-lactate" compartment="c"/> <!-- lactate 3 -->
    <species id="M_SUC_c" name="succinate" compartment="c"/>
    <species id="M_L_45_ALPHA_45_ALANINE_c" name="L-alanine" compartment="c"/>
    <species id="M_SER_c" name="L-serine" compartment="c"/>
    <species id="M_GLT_c" name="L-glutamate" compartment="c"/>

    <!-- Cofactors -->
    <species id="M_NAD_c" name="NAD" compartment="c" boundaryCondition="false"/>
    <species id="M_NADP_c" name="NADP" compartment="c" boundaryCondition="false"/>
    <species id="M_ADP_c" name="ADP" compartment="c" boundaryCondition="false"/>
    <species id="M_WATER_c" name="WATER" compartment="c"/>
    <species id="M_ATP_c" name="WATER" compartment="c"/>
    <species id="M_NADPH_c" name="NADPH" compartment="c"/>
    <species id="M_PROTON_c" name="PROTON" compartment="c"/>
    <species id="M_OXYGEN_45_MOLECULE_c" name="OXYGEN" compartment="c"/>
    <species id="M_Pi_c" name="Pi" compartment="c"/>
    <species id="M_PPi_c" name="PPi" compartment="c"/>
    <species id="M_CARBON_45_DIOXIDE_c" name="CARBON-DIOXYDE" compartment="c"/>
    <species id="M_CO_45_A_c" name="CO-A" compartment="c"/>
    <species id="M_UDP_c" name="UDP" compartment="c"/>
    <species id="M_NADH_c" name="NADH" compartment="c"/>
    <species id="M_AMP_c" name="AMP" compartment="c"/>
    <species id="M_AMMONIA_c" name="AMMONIA" compartment="c"/>
    <species id="M_HYDROGEN_45_PEROXIDE_c" name="HYDROGEN-PEROXYDE" compartment="c"/>
    <species id="M_Acceptor_c" name="Acceptor" compartment="c"/>
    <species id="M_Donor_45_H2_c" name="Donor" compartment="c"/>
    <species id="M_3_45_5_45_ADP_c" name="3-5-ADP" compartment="c"/>
    <species id="M_GDP_c" name="GDP" compartment="c"/>
    <species id="M_CARBON_45_MONOXIDE_c" name="CARBON-MONOXYDE" compartment="c"/>
    <species id="M_GTP_c" name="GTP" compartment="c"/>
    <species id="M_FAD_c" name="FAD" compartment="c"/>
  </listOfSpecies>
</model>
</sbml>

```

Supplementary file F2 : Summary of the 45 SynComs (C11, C17, C8, C41 were excluded of the analysis because of contaminants)

ID	ScopeRank	Rank	Isolates	Mean_Phylodist	PPM
C27	112.0-238.0	Min	Root1203-Root172-Root672-Root695	0,5898333333	235
C32	112.0-238.0	Median	Root181-Root553-Root695-Root918	0,9884216667	235
C29	112.0-238.0	Max	Root1464-Root180-Root402-Root695	1,3413116667	237
C23	238.0-291.0	Min	Root274-Root278-Root31-Root695	0,3307816667	247
C16	238.0-291.0	Median	Root1455-Root65-Root670-Root695	0,8891883333	289
C34	238.0-291.0	Max	Root1497-Root280D1-Root627-Root695	1,296585	262
C38	291.0-308.0	Min	Root258-Root274-Root672-Root695	0,6271783333	292
C17	291.0-308.0	Median	Root190-Root343-Root483D1-Root695	1,0000333333	306
C24	291.0-308.0	Max	Root61-Root695-Root77-Root935	2,0351766667	306
C26	308.0-322.0	Min	Root274-Root483D2-Root672-Root695	0,6511283333	308
C7	308.0-322.0	Median	Root405-Root561-Root695-Root74	1,0144233333	321
C2	308.0-322.0	Max	Root166-Root280D1-Root695-Root935	2,0404483333	317
C9	322.0-333.0	Min	Root1212-Root31-Root434-Root695	0,808955	323
C40	322.0-333.0	Median	Root335-Root423-Root630-Root695	1,0258033333	329
C21	322.0-333.0	Max	Root423-Root52-Root553-Root695	1,8696466667	331
C37	333.0-342.0	Min	Root267-Root318D1-Root565-Root695	0,8757983333	336
C6	333.0-342.0	Median	Root151-Root456-Root558-Root695	1,05027	341
C10	333.0-342.0	Max	Root1464-Root186-Root672-Root695	2,00653	336
C15	342.0-350.0	Min	Root29-Root473-Root565-Root695	0,874535	342
C35	342.0-350.0	Median	Root1204-Root431-Root695-Root983	1,1447216667	346
C3	342.0-350.0	Max	Root420-Root52-Root670-Root695	2,036045	347
C41	350.0-358.0	Min	Root1212-Root157-Root695-Root71	0,8616583333	351
C8	350.0-358.0	Median	Root1221-Root224-Root55-Root695	1,1899383333	352
C45	350.0-358.0	Max	Root186-Root190-Root627-Root695	1,9300533333	357
C39	358.0-365.0	Min	Root1221-Root695-Root720-Root74	0,9611016667	358
C14	358.0-365.0	Median	Root343-Root472D3-Root695-Root70	1,2059916667	358
C13	358.0-365.0	Max	Root1293-Root420-Root695-Root70	2,0885366667	361
C22	365.0-372.0	Min	Root1312-Root267-Root695-Root710	0,9841766667	369
C11	365.0-372.0	Median	Root122-Root1485-Root342-Root695	1,1942366667	368
C44	365.0-372.0	Max	Root1293-Root267-Root695-Root935	2,0899216667	371
C25	372.0-380.0	Min	Root491-Root569-Root695-Root954	0,9030433333	377
C42	372.0-380.0	Median	Root135-Root29-Root695-Root983	1,2178116667	373
C28	372.0-380.0	Max	Root1221-Root137-Root695-Root901	1,9812516667	379
C31	380.0-389.0	Min	Root318D1-Root50-Root565-Root695	1,0296566667	380
C1	380.0-389.0	Median	Root105-Root265-Root568-Root695	1,2194883333	388
C18	380.0-389.0	Max	Root137-Root275-Root420-Root695	1,9928816667	387
C19	389.0-401.0	Min	Root569-Root672-Root695-Root96	1,1047783333	393
C30	389.0-401.0	Median	Root180-Root690-Root695-Root9	1,3616933333	397
C36	389.0-401.0	Max	Root147-Root186-Root236-Root695	2,176505	397
C4	401.0-424.0	Min	Root105-Root275-Root63-Root695	1,1942033333	402
C12	401.0-424.0	Median	Root147-Root170-Root278-Root695	1,5856233333	405
C43	401.0-424.0	Max	Root131-Root236-Root695-Root901	2,211535	402
C5	424.0-481.0	Min	Root135-Root329-Root695-Root708	1,2638266667	438
C20	424.0-481.0	Median	Root147-Root217-Root672-Root695	1,6471066667	429
C33	424.0-481.0	Max	Root1295-Root52-Root68-Root695	1,8202733333	428

Supplementary file F3 : taxonomy and PPM of the 83 strains used in SynComs

Isolate	Regnum	Phylum	Class	Order	Family	Genus	PPM
Root105	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae	Devosia	91
Root1203	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Rhizobium/Agrobacterium	161
Root1204	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Rhizobium/Agrobacterium	184
Root1212	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Rhizobium/Agrobacterium	178
Root1221	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Rhizobacter		103
Root1293	Bacteria	Actinobacteria	Actinobacteria	Micrococcales	Microbacteriaceae	Leifsonia	92
Root1295	Bacteria	Actinobacteria	Actinobacteria	Streptomycetales	Streptomycetaceae	Streptomyces	154
Root131	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	169
Root1312	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Sinorhizobium/Ensifer	183
Root135	Bacteria	Actinobacteria	Actinobacteria	Corynebacteriales	Mycobacteriaceae	Mycobacterium	88
Root137	Bacteria	Actinobacteria	Actinobacteria	Micrococcales	Cellulomonadaceae	Cellulomonas	130
Root1455	Bacteria	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	Caulobacter	138
Root1464	Bacteria	Actinobacteria	Actinobacteria	Micrococcales	Microbacteriaceae	Agromyces	119
Root147	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	158
Root1497	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingopyxis	95
Root151	Bacteria	Actinobacteria	Actinobacteria	Propionibacterales	Nocardioidaceae	Nocardioides	100
Root166	Bacteria	Actinobacteria	Actinobacteria	Micrococcales	Microbacteriaceae	Microbacterium	114
Root170	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	Achromobacter	82
Root172	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Phyllobacteriaceae	Mesorhizobium	176
Root180	Bacteria	Actinobacteria	Actinobacteria	Micrococcales	Microbacteriaceae	Microbacterium	115
Root181	Bacteria	Actinobacteria	Actinobacteria	Micrococcales	Intrasporangiaceae	Terrabacter	160
Root186	Bacteria	Bacteroidetes	Flavobacteria	Flavobacterales	Flavobacteriaceae	Flavobacterium	60
Root190	Bacteria	Actinobacteria	Actinobacteria	Propionibacterales	Nocardioidaceae	Nocardioides	86
Root217	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Acidovorax	80
Root224	Bacteria	Actinobacteria	Actinobacteria	Propionibacterales	Nocardioidaceae	Nocardioides	137
Root236	Bacteria	Actinobacteria	Actinobacteria	Propionibacterales	Nocardioidaceae	Aeromicrobium	68
Root258	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Sinorhizobium/Ensifer	183
Root265	Bacteria	Actinobacteria	Actinobacteria	Corynebacteriales	Mycobacteriaceae	Mycobacterium	82
Root267	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Acidovorax	74
Root274	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Rhizobium/Agrobacterium	127
Root275	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Acidovorax	75
Root278	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Sinorhizobium/Ensifer	180
Root280D1	Bacteria	Actinobacteria	Actinobacteria	Micrococcales	Microbacteriaceae	Microbacterium	63
Root29	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Rhizobacter		156
Root31	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Sinorhizobium/Ensifer	173
Root318D1	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Variovorax	169
Root329	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	219
Root335	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Massilia	118
Root343	Bacteria	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	Caulobacter	102
Root402	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Acidovorax	63
Root405	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Pelomonas	89
Root420	Bacteria	Bacteroidetes	Flavobacteria	Flavobacterales	Flavobacteriaceae	Flavobacterium	55
Root423	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Sinorhizobium/Ensifer	166
Root431	Bacteria	Actinobacteria	Actinobacteria	Streptomycetales	Streptomycetaceae	Streptomyces	132
Root434	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Variovorax	126
Root456	Bacteria	Actinobacteria	Actinobacteria	Kineosporiales	Kineosporiaceae		110
Root472D3	Bacteria	Actinobacteria	Actinobacteria	Propionibacterales	Nocardioidaceae	Aeromicrobium	70
Root473	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Variovorax	121
Root483D1	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	Bosea	147
Root483D2	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Rhizobium/Agrobacterium	171
Root491	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Rhizobium/Agrobacterium	128
Root50	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas	152
Root52	Bacteria	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Paenibacillus	118
Root55	Bacteria	Actinobacteria	Actinobacteria	Streptomycetales	Streptomycetaceae	Streptomyces	146
Root553	Bacteria	Actinobacteria	Actinobacteria	Micrococcales	Microbacteriaceae	Microbacterium	57
Root558	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Sinorhizobium/Ensifer	181
Root561	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Rhodanobacteraceae	Rhodanobacter	98

Isolate	Regnum	Phylum	Class	Order	Family	Genus	PPM
Root565	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	Achromobacter	88
Root568	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Acidovorax	65
Root569	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	194
Root61	Bacteria	Actinobacteria	Actinobacteria	Micrococcales	Microbacteriaceae	Microbacterium	112
Root627	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Rhodanobacteraceae	Rhodanobacter	98
Root63	Bacteria	Actinobacteria	Actinobacteria	Streptomycetales	Streptomycetaceae	Streptomyces	155
Root630	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Pseudoxanthomonas	75
Root65	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Pseudoxanthomonas	73
Root670	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	Bosea	142
Root672	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Erythrobacteraceae	Altererythrobacter	123
Root68	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	194
Root690	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Lysobacter	73
Root695	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Phyllobacteriaceae	Mesorhizobium	166
Root70	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Acidovorax	60
Root708	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Rhizobium/Agrobacterium	183
Root710	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas	152
Root720	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas	154
Root74	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Sinorhizobium/Ensifer	183
Root77	Bacteria	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	Phenylobacterium	51
Root9	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	192
Root901	Bacteria	Bacteroidetes	Flavobacterii	Flavobacterales	Flavobacteriaceae	Flavobacterium	56
Root918	Bacteria	Actinobacteria	Actinobacteria	Micrococcales	Cellulomonadaceae	Oerskovia	120
Root935	Bacteria	Bacteroidetes	Flavobacterii	Flavobacterales	Flavobacteriaceae	Flavobacterium	53
Root954	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Sinorhizobium/Ensifer	183
Root96	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Lysobacter	72
Root983	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Lysobacter	73

Supplementary data X : Composition of the liquid media M9+ARE

M9 salt solution 10X

			Final amount for 1L
Na2HPO4 x 2H2O	0.422M	75.1118g.l-1	75.1118g
KH2PO4	0.22M	29.9398g.l-1	29.9398g
NaCl	0.0855M	4.9966g.l-1	4.9966g
NH4Cl	0.0935M	5.0013g.l-1	5.0013g

Dissolve the salts in 800 mL. Adjust pH to 7.2 with NaOH. mq water up to final volume of 1 L. Autoclave 15 min

MgSO4 1M 1000X

			Final amount for 0.1L
MgSO4 x 7H2O	1M	246.47g.l-1	24.647g

mq water up to final volume of 0.1L. Autoclave 15min 121°C

CaCl2 1M 3333X

			Final amount for 0.1L
CaCl2 x 2H2O	1M	147.01g.l-1	14.701g

mq water up to final volume of 0.1L. Autoclave 15min 121°C

Microelements/trace elements 100X

	Molar concentration	100X mg/L
FeSO4 x 7H2O	5.00e-5	1390.050
MnSO4 xH2O	1.00e-6	16.902
ZnCl2	1.00e-5	136.300
CuSO4 x 5H2O	1.00e-6	24.969
CoCl2 x 6H2O	1.00e-6	23.793
BO3H3	1.00e-5	61.830
EDTA	5.00e-5	1461.200
HCl	1.00e-3	10mL HCl 10M
Na2MoO4 x 2H2O	1.00e-6	24.195
NiCl2 x 6H2O	2.00e-7	4.754

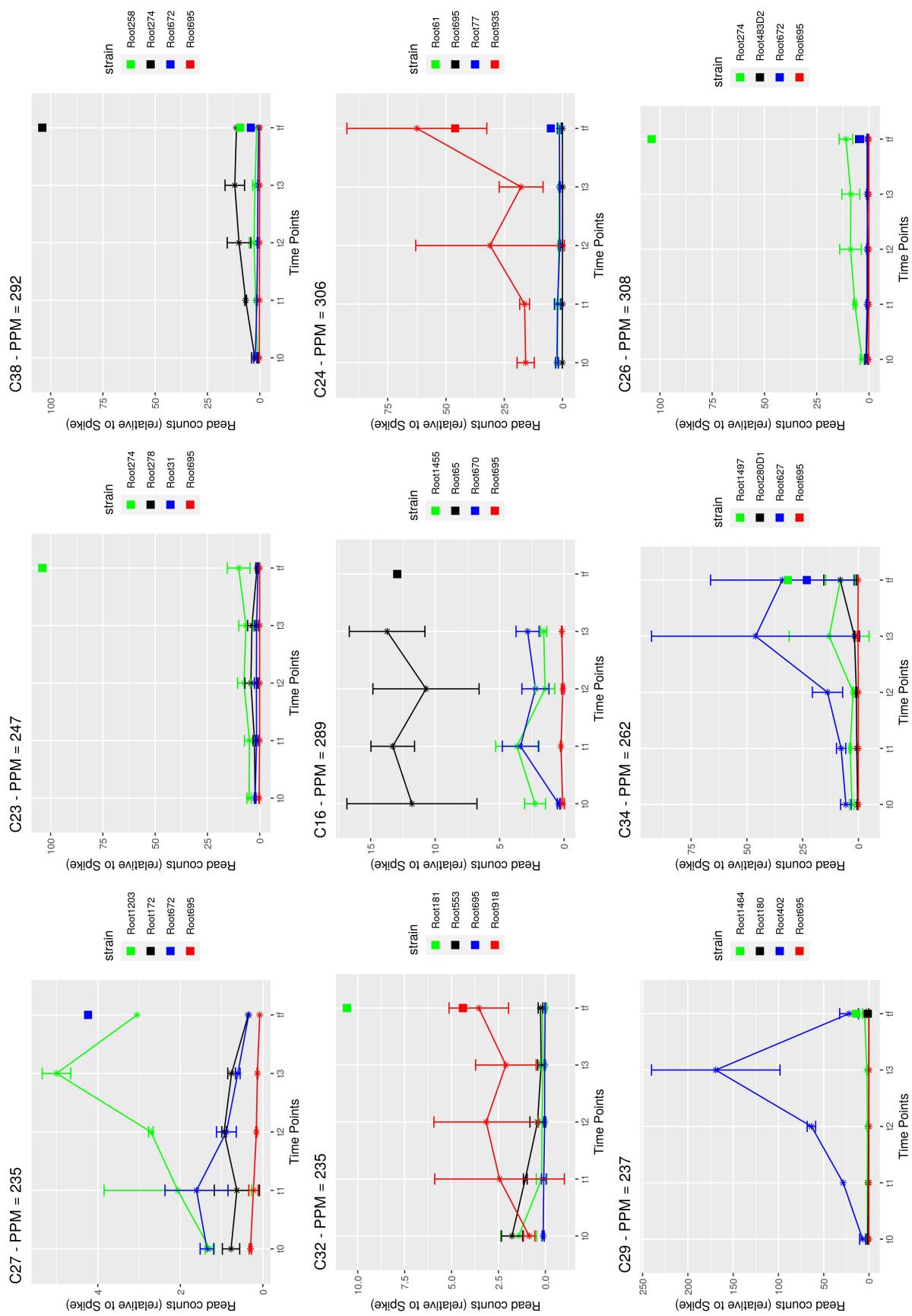
Filtrate

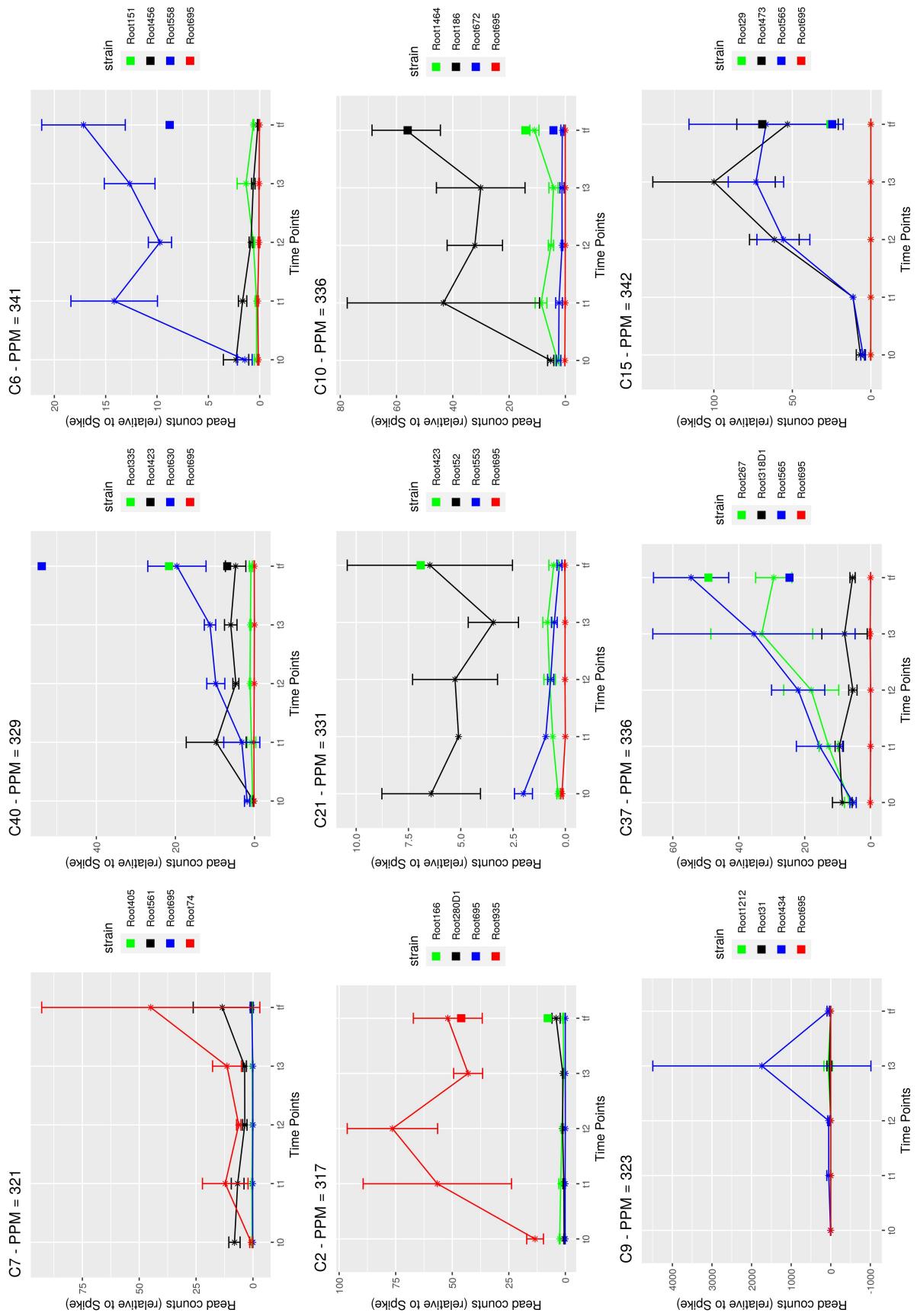
Artificial Root Exudate

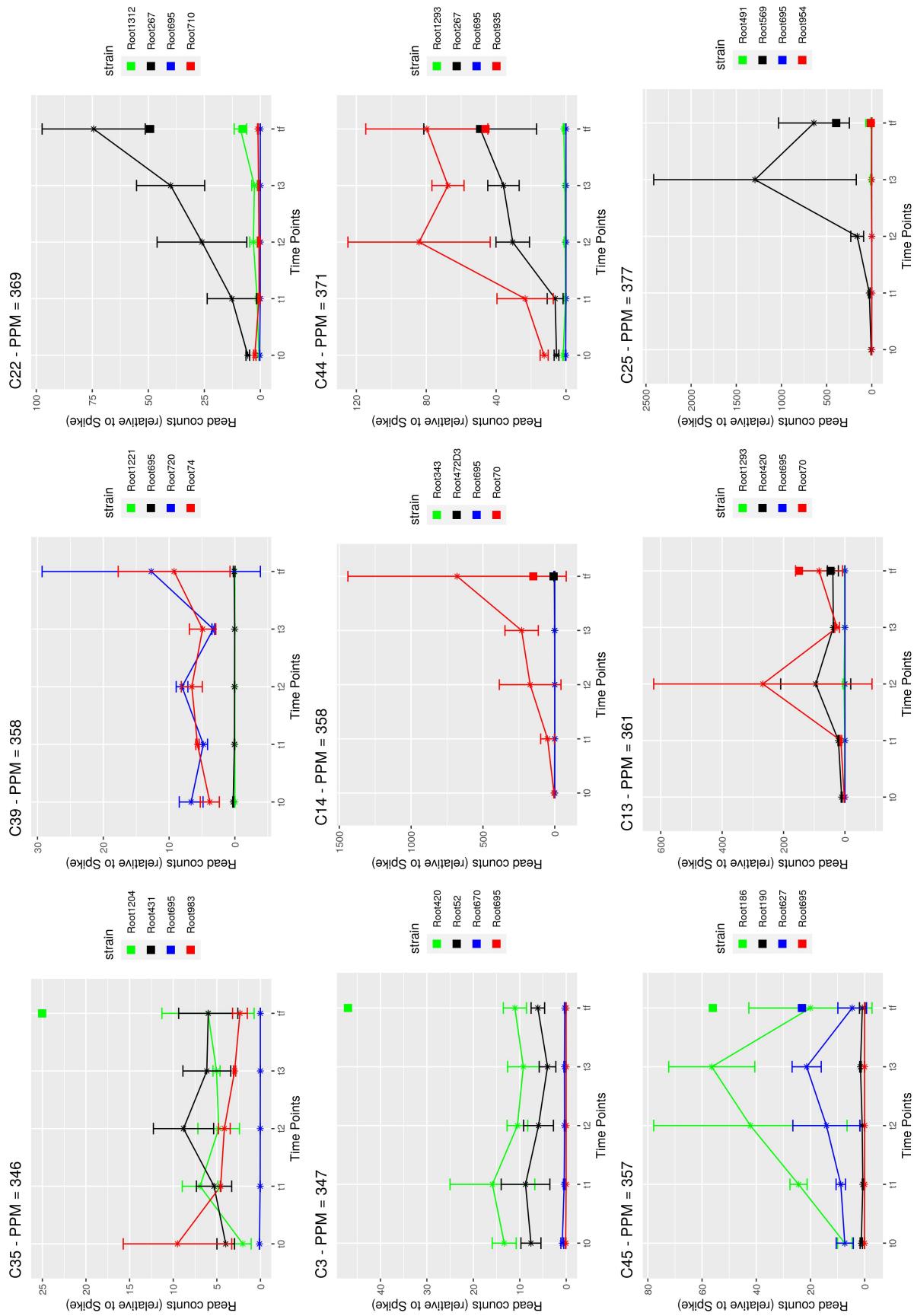
	5X Final volume 500mL
Glucose	Amount
Fructose	4.1g
Saccharose	4.1g
Citric acid	2.1g
Lactic acid	1.6g (1.75g C6H8O7 x H2O)
Succinic acid	1.6g (4.03mL 50 % w/v C3H5NaO3)
Alanine	2.3g (5.35g C4H4NaO4 x 6H2O)
Serine	2g
Glutamic acid	2.4g
Filtrate	2g

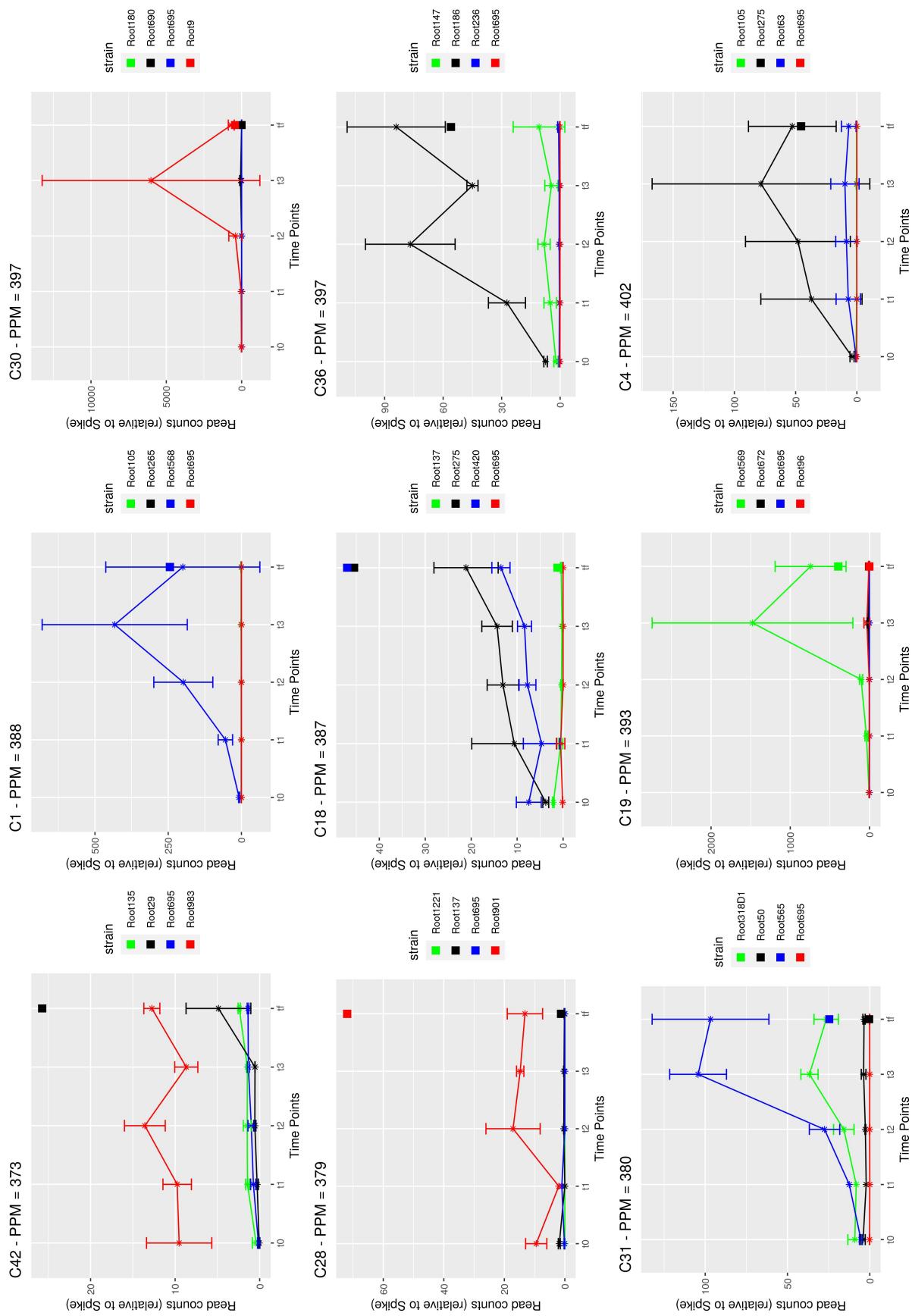
Mix

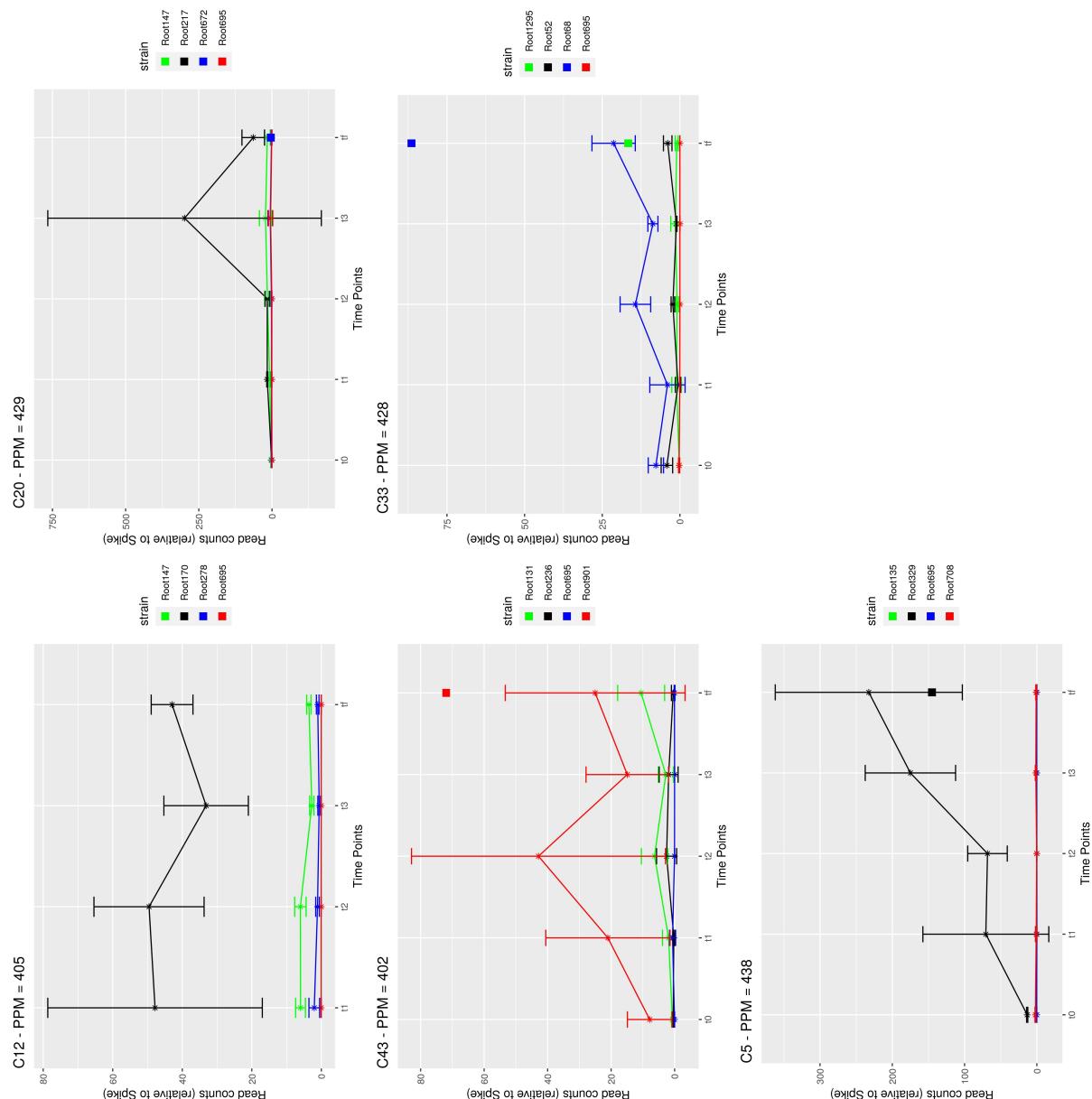
Stock solution	Concentration	Final volume 50mL	100mL	500mL
M9 salt solution	10X	5mL	10mL	50µL
MgSO4 1M	1000X	50µL	100µL	500µL
CaCl2 1M	3333X	15µL	30µL	150µL
Microelements	100X	500µL	1mL	5mL
ARE	5X	10mL	20mL	100mL
mq		33.935mL	67.87mL	339.35mL











Chapter 4. General discussion

General Discussion

1 Metabolic cross-feeding, coexistence, and community assembly : a summary

In its early years, despite the great benefits taken from the adaptation of macro-ecology theories and methods, microbial ecology suffered of a lack of precise theoretical background, as well as lack of understanding between micro and macro-scales (Balser et al., 2006; Prosser et al., 2007; Widder et al., 2016). One key padlock in microbial ecology is the understanding of a species and what is called a species. The species concept has progressively been modified with the acceptance of a phylogenetic species concept (Mishler and Brandon, 1987) allowing to describe any microorganisms of a microbial community, known or unknown, by the use of DNA fragments. More recently, the intensive use of -omics data and modelling permitted highlighting the importance of cross-feeding and cooperative interactions into microbiota's structure, but the lack of integration of positive interactions in the niche theory framework limited a deep formalized understanding of microbial communities. This issue is being progressively solved with accumulating modeling and experimental evidences (Koffel et al., 2021).

1.1 The prevalence and importance of cross-feeding

Various approaches permit the prediction of cross-feeding interactions. The most straightforward method is the use of reconstructed metabolic networks, which simulate an organism's metabolism, thus its potential interactions with others. Other approaches can for example involve the simulation of evolution within a community, and show that cross-feeding can emerge as a result of gene loss or compounds leakage. Most methods usually predict a strong prevalence of cross-feeding interactions, thus cooperative behaviors within microbial communities (San Roman and Wagner, 2018). Hence, cooperation via the secretion of metabolites could be the answer to species coexistence within microbial communities, whose species richness and diversity are greater than expected by classical theoretical ecology (Vandenkoornhuyse et al., 2015; Mas et al., 2016). Some argue that cross-feeding could be an answer to an evolutionary dilemma : resource investment vs. function trade-off, in other words the amount of energy spent compared to the efficiency of the performed function (Carlson, 2007; Flamholz et al., 2013; Carlson et al., 2018).

Considering the high frequencies of auxotrophies and their benefit to organisms' fitness, as well as a likely unavoidable leaking of public goods during bacterial growth (Morris et al., 2012b; Morris, 2015), microbial communities are shaping complex, highly connected intercellular networks (D'Souza et al., 2014). Hence, microbial communities are fundamentally decentralized systems (Gralka et al., 2020), despite the probable existence of core, key species. The results presented in chapter 2 shed light on such statements, since a number of cross-feeding possibilities (under well-defined environmental constraints) have been hypothesized, but with some highly redundant bacterial strains and a relatively high degree of functional redundancy among members of the community.

The importance of cross-feeding can also be highlighted by its potential for niche construction. Niche construction is a process where organisms modify their environment, in a manner that affect both their and others' dynamics, such as biofilms construction, antibiotics secretion, emission of a detoxifying enzyme, byproducts secretion (Callahan et al., 2014; Loudon et al., 2016; San Roman and Wagner, 2018). *In silico* predictions of interactions with metabolic models often returns dramatically high amounts of possible cross-feeding interactions between microorganisms on various carbon sources, suggesting high possibilities of niche construction (San Roman and Wagner, 2018), since metabolites excretion create conditions that allow the coexistence of many different species, sometimes even on a single carbon source (Gralka et al., 2020).

However, despite its promotion of species coexistence, cross-feeding is inherently an unstable situation, and mathematically challenging to solve. Different ways are explored, such as competition between cross-feeding mutualists and cheaters, the parameters favoring the emergence and invasion of a cheater, nutrient specificity, biosynthesis rates (Sun et al., 2019).

1.2 Cross-feeding selective drivers

Related to the evolutionary stability of cross-feeding is its evolutionary origin. Many bacteria are often characterized by genome streamlining, a genome-reduction process occurring in nutrient-rich and/or constant environments, leading to the loss of metabolic functions. By this process, a bacteria reduces the metabolic cost of a function that is not mandatory under these permissive environmental conditions, and gains a selective advantage over other bacteria that still pay these costs (Giovannoni et al., 2005; D'Souza et al., 2014). Hence, the auxotrophic bacteria depends on available external sources of the involved nutrient that can be supplied by prototrophic bacteria, creating a cross-feeding interaction. Accordingly to some studies, auxotrophy is common in the bacterial domain of life. Notably, auxotrophies for amino-acids, nucleosides, and vitamins are very frequent (D'Souza et

al., 2014; Mee et al., 2014; Liu et al., 2018). Far from being a load, it can provides a fitness gain (verified with experimental assays) if the considered auxotrophy is compensated by the molecule availability in the environment (D’Souza et al., 2014). If gene loss and auxotrophy confer a selective advantage under appropriate environmental conditions, by energy saving, then the microbial community complexity should be explained with auxotrophs relying on each other and forming a complex inextricably woven network of interactions. This selective advantage of being auxotroph has been experimentally demonstrated by the enhanced root colonization observed of a genetically diminished bacteria for amino-acids production (Cole et al., 2017). In addition, there can also be positive selection for the excretion or transport of these valuable compounds (Douglas, 2020; Fritts et al., 2021). It was also demonstrated that cross-feeding at high fluxes return a better functional return on investment than monocultures (Pande et al., 2014; Carlson et al., 2018).

1.3 Integration of cross-feeding with competition and environmental factors : from *in silico* predictions to experimental validation

Due to the complexity of microbiomes, synthetic communities emerged as a powerful and suited tool to decipher microbial interactions (Jacoby and Kopriva, 2019). Metabolic modelling of such synthetic consortia recognized cross-feeding as an ubiquitous process in microbial communities (Widder et al., 2016; Pande and Kost, 2017). Moreover, experimental approaches validated the capacity of numerous bacteria to feed on others’ exudates (Goldford et al., 2018). However, experimental approaches still struggle to confirm *in silico* predictions (Fritts et al., 2021), suggesting a predominance of competition (Foster and Bell, 2012), even in synthetic communities designed for cooperation (Hillesland and Stahl, 2010). Nonetheless, it does not necessarily mean that cross-feeding does not exist. Its benefits might be most often bypassed by competition, thus undetected when using simple experimental protocols, as discussed in chapter 3. Making the part between cooperative and competitive outcomes is then not an easy task.

1.3.1 The difficulty to find simple and appropriate metrics

Several studies attempted to predict cross-feeding with simple metrics obtained from reconstructed metabolic networks and phylogeny, as we did in chapter 2, with similar results. However, we could not confirm these *in silico* predictions with experimental observations with classical co-cultures and regular metrics such as SynComs evenness (chapter 3). Such mismatch between modelling and experiments are regularly discussed, notably regarding phylogeny being a weak predictor of metabolic functionality and cooperation (Young, 2016; Simonet and McNally,

2021). In addition, in Lam et al., (2020), it is argued that metabolic completion scores must be normalized according to phylogenetic distance, which is treated as a bias.

1.3.2 Cross-feeding as a mostly underlying process compared to competition ?

In chapter 3, we found that competitive exclusion shaped most of the tested SynComs, in opposition with our hypothesis stating that phylogenetic distance is favorable to cross-feeding. However, as discussed in the same chapter, there are limits mostly identified as the interference of other interactions, possibly due to the oversimplification of natural habitats in experimental protocol, that could have masked the effect of cross-feeding.

Competition for other nutrients can still occur between microorganisms exchanging compounds, sometimes bypassing cross-feeding (Hillesland and Stahl, 2010). In some cases, the cross-fed nutrients are valuable both for the receiver and producer, leading to competition between cross-feeding partners. In such cases, the receiver must have a competitive advantage towards the acquisition of the public good in order to maintain mutualism (McCully et al., 2017a). Cross-feeding could be an underlying process stabilizing co-existence of competitive guilds in a dense network of high-order interactions (Goldford et al., 2018). Also, competitors between receivers can benefit to the stability of cross-feeding, as demonstrated by (Celiker and Gore, 2012), where *Escherichia coli* competes with *Saccharomyces cerevisiae* cheaters for extracellular sucrose. *E. coli* limited yeast population and public good availability at durable levels.

Cross-feeding might be relevant in harsh conditions, to prevent starvation of several species at once (McCully et al., 2017b; Fritts et al., 2021), as highlighted in chapter 2 and further discussed in chapter 3. Such situations make the existence of cross-feeding interactions not necessarily correlated with organisms' abundances (which is already discussed in chapter 3), but suggests that positive interactions are a mechanism of maintenance of coexistence and are consistent with the idea that microbial communities are decentralized, highly interconnected networks. Indeed, the high functional redundancy and predicted possibilities of combinations capable of cross-feeding highlighted in chapter 2 suggested such intricate and diverse connections.

Nevertheless, we found that cross-feeding can occasionally be the main driver of a small bacteria consortium. Indeed, the growth patterns of an *Achromobacter sp* did show that dependencies to other strains' exudates dramatically improved its growth (chapter 3). The observed growth enhancement did not seem specific to the providers' taxa, and apparently depend on their growth rate, thus their capacity to excrete wastes (or to set up overflow metabolism).

1.3.3 A few competitive behaviors at the origin of many cross-feeding opportunities ?

Metabolic leakage is however not necessarily a metabolic deficiency (i.e. the loss of compounds that could have been further metabolized). For instance, cells may get rid of elements that are in excess. Various mechanisms can explain such excretion as an optimal strategy. An example, identified in *E. coli* (Basan et al., 2015) is referred to as overflow metabolism and is a good example to explain that metabolic cross-feeding is not necessarily dissociated from competition. Overflow metabolism is the secretion of large quantities of organic byproducts as a result of competitive strategies adopted by some microbial species. It does not mean waste metabolism, but it rather possibly reflects an optimum resource usage. Hence, those byproducts constitute public goods that can initiate and stabilize cross-feeding interactions in microbial communities. (Carlson et al., 2018). Hence, such a mechanism would mean that a few highly competitive species can benefit to multiple cooperating ones.

1.3.4 Environmental constraints

Finally, deciphering ecological interactions cannot be done without considering the environmental context, and the way to model it. Notably, as seen in chapter 2 and the literature, many interactions outcomes depend on nutrient availability and utilization (Coyte and Rakoff-Nahoum, 2019). Cooperative behavior are thought to be favored by nutrient depletion, and competitive behavior are prominent under weak nutritional constraints (Magnúsdóttir et al., 2017; Coyte and Rakoff-Nahoum, 2019). In addition, time, and space can modify how microorganisms interact together.

In brief, depending on numerous conditions, species can either be competing or enter an obligate mutualism, which was highlighted both by experimental and modeling methods (Mahowald et al., 2009; Hoek et al., 2016), and causes challenges to experimental validation of predicted cross-feeding interactions based on reconstructed metabolic networks. Consequently, experimental validations of cross-feeding should attempt to be closer to the modelling parameters (Widder et al., 2016; Zhalnina et al., 2018b) or reciprocally modelling should be more realistic of the environmental reality and/or experimental protocols..

2 Metabolic cross-feeding is not only about GEMs and various padlocks have to be considered

If the data mining of GEMs constitutes the basis of *in silico* predictions of cross-feeding between organisms, it was highlighted that numerous other factors, detailed hereafter, have to be considered.

2.1 Metabolites availability, secretion, transport, and uptake

We repeatedly discussed that microbes metabolism and metabolic dependencies are dependent on available nutrients, which are key environmental constraints. Consequently, a connected and as much important area of research is the screening of nutrients composition of the surrounding environment, notably the secretome of microbes and hosts, also called the exometabolome (Douglas, 2020).

Most of the exometabolome is composed of compounds derived from living-cells and not cell lysis, and is believed to contain mostly by-products, whose intracellular accumulation would cause damage such as redox imbalance (Pinu et al., 2018; Douglas, 2020). However, there is growing evidence of bigger variety of exometabolites, such as enzymes, sugars, organic acids, amino-acids, vitamins (Fritts et al., 2021). Advances in exometabolomics assays could produce giant steps into a mechanistic understanding of microbiomes, notably regarding metabolic interactions or substrates preferences, both with applications in *in vitro* culture and microbiome engineering (Lubbe et al., 2017). Such data would indeed determine with a higher precision which compounds a microbe is capable to secrete and uptake (Jacoby and Kopriva, 2019), opening the gate to discover all resultant multi-way cross-feedings (Douglas, 2020). As examples, two recent studies have applied exometabolomics workflows on the rhizosphere of *Avena barbata* and *Arabidopsis thaliana*, providing information about which compounds are uptaken by bacteria when supplied with root exudates only (Jacoby et al., 2018; Zhelnina et al., 2018a).

Close to metabolites secretion is their transport from providers to receivers organisms. Transport can be passive or active, contact-dependent or not (D’Souza et al., 2018). Identifying genes coding for transporters of cross-fed nutrients’ is an active part of research (Douglas, 2020). Finally, the capacity of various microbes to uptake various public goods must be taken into account and investigated. Indeed, we only have relatively few knowledge about which compounds are consumed or to which extent microbes differ in their substrate preferences (Jacoby et al., 2018). Species’ specificity is to consider first, but specific structures such as biofilms since they constitute a strategy

to assimilate particulate compounds and could favor cooperation rather than cheating because of particular spatial arrangements (Sivadon et al., 2019).

2.2 Root exudates composition and their effect on microbiota

Specific exudates components can be used to recruit or deter certain microbial strains, mediated via defined biochemical mechanisms, for instance by secreting scopoleptin, an anti-microbial compound (Stringlis et al., 2018). A well-known mechanism is the mutualistic exchange of carbon and nitrogen between the plants (carbon donor) and its microbiota (nitrogen donor). However, since they can serve as nutrients for microbes, plant root exudates can also mediate metabolic interactions, from competition to metabolic cross-feeding, depending on their composition and on the composition of the microbial pool. It was also demonstrated that plant root exudates vary according to the plant genotype and development stage; for example, exudates of *Avena barbata*'s seedlings are enriched in sucrose, while senescent individuals secrete more quaternary amines (Zhalnina et al., 2018a). In *Arabidopsis thaliana*, natural root exudates variations were discovered by comparing several accessions (Monchgesang et al., 2016). To go further, a way to embrace host-microbes interactions is to consider cross-kingdom models gathering components of all partners, i.e. metanetworks dealing for example co-transcriptome networks (Zhang et al., 2019), co-occurrences networks of bacteria and fungi (Wassermann et al., 2019), or interactions between GEM of bacteria, fungi, and host plant.

2.3 Spatial scale

Ideally, spatial scale should be considered in any case of cross-feeding, associated to public goods transport mechanisms. Indeed, spatial patterns can have different effects which are not fully understood yet. For example, Ebrahimi et al., (2019) highlighted that strains of *Vibrio splendidus* form large aggregates in different ways, depending on the strength of enzymatic activity. Weak enzyme producers rearranged spatially to promote cooperative behavior instead of displaying competitive behaviors. In contrast, cross-feeding partners for electrons were found to display a cellular activity independent to their distance and intermixing (McGlynn et al., 2015). Finally, spatially structured environments (such as biofilms) provoke a diffusion imbalance, hence concentrations gradients of nutrients. These gradients influence competition and cooperation. They tend, notably, to drive the segregation of competing organisms and to promote the mixing of cooperative ones (Carlson et al., 2018). In summary, most natural environments are far from being homogeneous and are characterized by various microscale gradients (Cordero and Datta, 2016), to

which cross-feeding partners answer in different ways. Such an issue needs to be answered by learning to how to scale observed patterns and models (Gralka et al., 2020).

2.4 Genomes annotation

The prediction of metabolic interactions with reconstructed metabolic networks relies on the annotation of genomes. As mentioned in the general introduction (part II), a metabolic network quality and reliability strongly depends on how well the organism's genome can be annotated (i.e. how many biological sequences can be correctly identified). Up to now, a major part of meta-omics data cannot be annotated (50-80% in 2018) because of a lack of knowledge about involved genes, genome assembly and gene prediction accuracy. Such an amount of missing data obviously leads to an incomplete picture of the studied systems, hence research is somehow limited to well-known organisms, processes or pathways, reliability decreasing as annotation's quality decreases (Raina et al., 2018). This issue is frequently encountered in microbiology and particularly in symbiotic systems, since they imply many poorly known organisms, that probably contain specific, unknown genes (Porter et al., 2017).

2.5 Models complexity or simplicity ?

Current research frequently aims to improve models with a better generalization and an increased complexity (i.e less specific to a model organism or particular conditions, sometimes with more species, with the consideration of both cooperative and competition) (Wade et al., 2016).

However, all approaches face challenging issues. On the first hand, models of a high complexity cannot be correctly parameterized, notably without supporting experimental data (Karr et al., 2012). For example, flux balance analysis requires flux measurements. On the other hand, too simple models (for example Lotka-Volterra models or consumer-resource models) cannot enlighten cellular mechanisms, organisms being considered as “black boxes” (Momeni et al., 2017). For instance, in synthetic biology and microbiome engineering, metabolic networks are often too complex and cannot be handled efficiently for the development of microbial factories. Hence, some frameworks attempted to cope with a compromise of both sides, by using simplified metabolic networks. Roughly, these methods operate a reduction of metabolic networks in order to keep the core parts significantly involved in the studied process. However, the part of the networks to be conserved is decided by the user, and is far from being an easy task (Richelle et al., 2020).

Such issues are being partially compensated with the combination of top-down and bottom-up approaches, working together iteratively in a design-build-test-learn process (Lawson et al.,

2019) (as detailed in the general introduction, part II). Notably, statistical modelling approaches are developed in order to predict coexistence and species abundances of numerous assemblages of species by learning on the results of limited experimental assays (Maynard et al., 2020). Finally, it was demonstrated that models at an intermediate scale can accurately (i.e. close to experimental data) quantify metabolic processes such as cross-feeding and competitive exclusion within reasonably complex microbial consortia (up to 14 amino-acids-auxotrophic bacteria) (Liao et al., 2020).

3 Conclusion and perspectives

The present work confirmed the patterns generally observed when investigating microbial cross-feeding in host-associated microbiota, i.e. numerous predicted interactions, that turned out to be difficult to validate with standard experimental protocols.

In terms of direct perspectives, a particular attention should be given to the sole observed and highly effective experimental pattern of cross-feeding with the strain *Achromobacter sp*. First, pairwise models of this strain coupled with different providers could be computed and compared, notably with flux balance analysis between organisms (Budinich et al., 2017), under the condition to find a suitable growth function. Second, constraints on what compounds can be excreted or not could be incorporated, if possible. Third, models studying time and space dynamics of these combinations could follow. Finally, the stability of such pairwise (or more) associations could be modeled, notably according to the flow rate of metabolites between strains. Indeed, high rates are supposed to allow cheater to appear more quickly (Sun et al., 2019).

However, at a larger study-scale with many microorganisms, pairwise interactions might not be suitable. Indeed, many predictions from pairwise assays might not occur in the community, because of high-order interactions (i.e. when a pairwise interactions is altered by the presence of other species), or because of a more heterogeneous environment (De Vos et al., 2017; Gralka et al., 2020). Alternatively, other methods such as leave-one-out protocols can be employed, in which the community is built by alternatively adding or removing members (Maynard et al., 2020). In any case, any mechanistic model, even complex in terms of number of species, environmental factors or any other parameter, should stick on simple to intermediate scales, for example two-dimensional niches (San Roman and Wagner, 2018; Koffel et al., 2021).

In regards to the aforementioned padlocks, microbial system ecology proposes several conceptual and technical attempts. Computational and mathematical modelling are always

incorporated into microbiome engineering frameworks, and put great efforts to explain and predict coexistence, notably by integrating as much as possible (i) ecological processes such as interactions, (ii) fluctuating environments, (iii) evolutionary aspects (Escalante et al., 2015; Wade et al., 2016), and (iv) integrated -omics data, i.e. the cross-talk between species abundances, metabolites fluxes, gene expression (Volkova et al., 2020). For example, metatranscriptomics combined with GEM analysis can be helpful to decipher the effects of environmental parameters on genes' expression and the metabolic pathways they are involved in. Another possibility is the recent development of single cell technologies such as microfluidics also offer promising tool to study the complexity of microbial ecology (Richards et al., 2019; Mauger et al., 2021). Notably, they provide new ways to understand how microbial communities can exchanges metabolites through cross-feeding, since they give access to levels of definition never captured before (Richelle et al., 2020; Mauger et al., 2021). For instance, microfluidics demonstrated that cells generate metabolite gradients on the micrometer scale, resulting in glucose-acetate cross-feeding interactions among sub-populations of cells allowing antibiotic resistance (Dal Co et al., 2019).

Conceptually, there is a need to see beyond lists of species, genes, and metabolites. For instance, energy flow must be integrated from nutrients (biomass precursor) to effective growth (biomass production) (Gralka et al., 2020). Another way to apprehend microbial communities would be to think in terms of functional units instead of single species, each unit representing a key metabolic or ecological role. Echoing our results about functional redundancy detailed in chapter 2, such approach is convenient to identify functionally redundant taxa that can replace each other, without threatening the stability of the metabolic process they assure (Bordenstein and Theis, 2015; Gralka et al., 2020). All the presented approaches are the building blocks of synthetic ecology (Zomorrodi and Segre, 2016), i.e. artificial microbial communities designed to perform a task while maintaining the resilience and complexity close to native microbial communities (Bosi et al., 2017). Microbial system ecology has many applications, from medicine, environment science, sustainable agriculture, and biotechnology. These applications require a deep understanding of microbiotas and how microbes interact with their hosts and their environments. The present work falls into this ambition and highlights several key points that need to be addressed, for example phylogenetic signal, functional redundancy, mechanisms of coexistence, environmental parameters and integration to the niche theory.

Supplementary chapter 5 Phylogenetic and functional clustering in a representative bacterial consortia of the *Arabidopsis thaliana*'s root microbiota

(Side project : preliminary results)

Phylogenetic and functional clustering in a representative bacterial consortia of the *Arabidopsis thaliana*'s root microbiota

Abstract

Disentangling the structure of microbial communities is challenging in multiple ways, notably because of the intermixing of effects from environmental variables and from biotic interactions occurring between microorganisms. There are different approaches that permit studying both aspects, without however being self-sufficient and requiring integration with other methods. Co-occurrences networks are among these approaches, and are regularly used to study the effect of environmental shifts on the structure of microbial communities, and to open a door upon the acquisition of the underlying interaction network. In this study, we inferred the co-occurrences network of the root microbiota of 7-weeks old *Arabidopsis thaliana* plants, whose germ-free seeds had been colonized with a representative consortia of root and soil derived bacteria. The network was relatively poorly dense, likely due to the reduced size of the inoculum compared to natural communities, but followed a scale-free topology, as regularly observed in many biological networks. We identified some keystone nodes, connecting nodes among and within hubs, or being central to most paths in the networks. Most of these keystones contained strains involved in relevant interactions with the host plant, notably nutrition. Finally, we observed phylogenetic clustering within the different modules and connected nodes compared to unconnected ones, as well as functional similarity and redundancy between nodes of the same modules. Such clustering highlighted habitat filtering, i.e. selection of the best suited traits to colonize this environment, leading to a higher similarity between micro-organisms, that could possibly be an early-colonization effect.

1 Introduction

Networks are widely used in biology and ecology, due to their usefulness to represent and analyze how biological components are connected and possibly interact together. Several biological entities can be represented as networks, such as genes regulatory networks, metabolic networks, or pollination networks (Olesen et al., 2007). However, inferring the complete or exact interactions between species is challenging due to the inherently intricate and complex nature of biotic

interactions, particularly when it comes to study microbial communities. Indeed, such communities harbor thousands of coexisting species, and deciphering their functioning harbors additional challenges compared to macro-ecology, for example species' definition, quite different dispersal abilities, dormancy, complex spatial and temporal structure (Rosselló-Mora and Amann, 2001; Martiny et al., 2006).

Co-occurrences networks have been used to efficiently analyze variations and resilience of microbial communities structures in response to environmental variations, such as stresses in agriculture soils (i.e. effects of soil compaction, chemical inputs, etc.) (Karimi et al., 2017; Ramirez et al., 2018; Schlatter et al., 2018; Price et al., 2021). Co-occurrences networks are suited to decipher community dynamics by highlighting shifts in their structure (Layeghifard et al., 2017). However, co-occurrences networks are not interaction networks. Connected species are only statistically significant co-occurrences or co-exclusion (Röttjers and Faust, 2019), , and identifying the ecological interaction between them is much more challenging. Interactions between organisms are diverse, classified according to the outcome for both organisms : negative for both (competition), beneficial for both (mutualism), negative for one and positive for the other (predation and parasitism). In between, there are also interactions with no effect for one the organism (commensalism and ammensalism) (Faust and Raes, 2012; Pacheco and Segrè, 2019). Hence, abundances-based occurrences between two species within a network can translate an interaction outcome, but without many clues about which interaction is taking place. Moreover, species abundances extracted from meta-omics data are compositional, thus require additional statistical treatments other than standard correlation coefficients to avoid false correlations (Aitchison, 1982; Friedman and Alm, 2012; Gloor et al., 2017). Finally, networks being intricate associations of many organisms, there is a risk for false associations. Indeed, an organism A might seem to directly affect the presence of an organism B, while it actually affects an intermediate organism C, which has a direct effect on B. Some methods can actually fix this issue by using conditional independence (i.e. A and B are conditionally independent if $P(A | B,C) = P(A | C)$) (Kurtz et al., 2015).

There are two main ecological principles for interpreting co-occurrences networks. First, the principle of competitive exclusion states that two species with similar niches exclude each other. Similar species would then not coexist, hence would not co-occur (Faust and Raes, 2012). Second, the principle of habitat filtering states the opposite, arguing that the environment filters the species accordingly to their capacity to survive within. Hence, species with similar niches would be more likely to coexist (Zhang et al., 2017). On this basis, many hypothesis can be made. For example two

competing species could be negatively correlated in the network, while cooperating ones would be positively correlated. However, such statements are not fully validated. Consequently, co-occurrences network are best suited to study the structure and dynamics of a microbial community with a global point of view. Deciphering the underlying interaction networks is much more challenging and subject to many errors, and studies require complementary data to go further. The use of metrics on the constructed networks like centrality degree allow to go further and to build hypotheses and predictions on why microorganisms co-occur (Berry and Widder, 2014a).

In this exploratory analysis, we used a co-occurrence network to analyze the structure of a representative bacterial consortia of *Arabidopsis thaliana*'s root microbiota. The network was inferred based on OTUs abundances obtained from a previous study (Bai et al., 2015). We aimed to identify putative keystone species with metrics dedicated to networks analyses, as well to link the observed structure to known ecological mechanisms such as habitat filtering. We hypothesized that a few OTUs would harbor high centrality measures, but without making assumptions on their ecological role (the formulation of hypotheses about underlying interactions between connected OTUs is however briefly discussed). In accordance with previous studies that highlighted phylogenetic clustering in co-occurrences networks (Horner-Devine and Bohannan, 2006), we also hypothesized that the closer two OTUs were in the networks (i.e. short paths), the smaller was their phylogenetic distance.

2 Material and methods

2.1 Data acquisition

Raw data were taken from a previous study (Bai et al., 2015). In this study, two synthetic bacterial consortia were used, one derived from soil and roots, the other from leaves of *A. thaliana*. After being isolated and gathered in a culture collection, they were inoculated to germ-free *A. thaliana* plants by different methods. We took interest in root+soil derived inoculum in particular (referred as ‘R+S’), composed of 193 bacterial strains of *Arabidopsis thaliana* root microbiota and 30 bacterial isolates from soil was inoculated in sterilized calcined clay directly before the sowing of surface-sterilized germ-free seeds of *A. thaliana* (this inoculation is referred as ‘clay’ throughout the text). This inoculum was chosen because we disposed of the annotated genomes of all the 193 root-derived bacterial strains, obtained in the same study (Bai et al., 2015). However, the 30 soil-derived bacterial strains did not have their genomes sequenced at the time of this study, which causes a few gaps in the data analysis. In Bai et al., (2015), bacterial abundances were measured by

16S RNA sequencing after seven weeks. The resulting abundances table displayed the abundances of OTUs rather than strains because some strains have identical 16S RNA sequences and cannot be distinguished by sequencing when present in the same sample.

2.2 Data filtering

The co-occurrences table gathered many samples from various inoculum and inoculation methods. Because co-occurrences networks methods need homogeneous sample sets to build reliable associations, the table was filtered to keep only samples of the same plant compartment (roots) from replicates that had been inoculated with the same inoculum ('R+S') and the same substrate ('clay'). Then, in order to have the most homogeneous table as possible, OTUs were filtered according to their prevalence. We applied a stringent threshold : kept OTUs had to be present in at least 80% of the samples.

2.3 Co-occurrences network

The co-occurrences network was inferred with the R package SpiecEasi, that assumes that the network is sparse and has a built-in statistical treatment for compositional data (Kurtz et al., 2015). The network's modules were computed with the edge-betweenness centrality algorithm (available in the R package igraph (Csardi and Tamas, 2006)). Key module members and the associated Z-P plot were detected according to the algorithm described in (Deng et al., 2012).

2.4 Bacterial strains metrics, and scale to the OTU level

A phylogenetic tree had been previously computed (Bai et al., 2015) and was used to compute phylogenetic distances between all pairs of strains with the Python package ete3 (Huerta-Cepas et al., 2016). Phylogenetic distance between two OTUs were obtained by computing the phylogenetic distances of all possible strains pairs among them, without considering OTU's belonging, then computing the mean (i.e merging the 2 OTUs together and taking all the pairs of strains). Genome-scale metabolic models (GEM) were reconstructed from annotated genomes with an automated version of Pathway Tools (Karp et al., 2002, 2011, 2020; Belcour et al., 2020) and AuReMe (Aite et al., 2018). The list of chemical reactions of each metabolic network was then available in AuReme reports. The list of chemical reactions of each GEM (one per strain) was then available in AuReme reports. An OTUs' reactome (i.e. the whole set of encoded reactions) were computed by merging the reactomes of all its strains. Phylogenetic distances between connected and unconnected OTUs

We tested the relation between phylogenetic distances and connected OTUs in two manners. First, by comparing all the phylogenetic distances for all directly connected OTUs (82 OTUs pairs) with all the phylogenetic distances of all non-existing edges in the network (i.e. all not directly connected OTUs), which returns many OTUs pairs (2546). Even if rank tests are suited for uneven samples sizes, the extremely big difference between connected OTUs ($n=82$) and unconnected OTUs ($n=2546$) led to a supplemental analysis with a random sub-sampling procedure : (I) 1000 sets of 82 unconnected OTUs pairs were sampled; (ii) for each set, the phylogenetic distance mean was recorded and (iii) we computed the distribution of these 1000 means as well as the probability to observe the phylogenetic distance mean of the observed connected edges under this distribution. The difference of distributions of phylogenetic distances between connected and unconnected OTUs were tested in R with a one-sided Wilcoxon rank sum-test.

2.5 Modules' phylogenetic distances

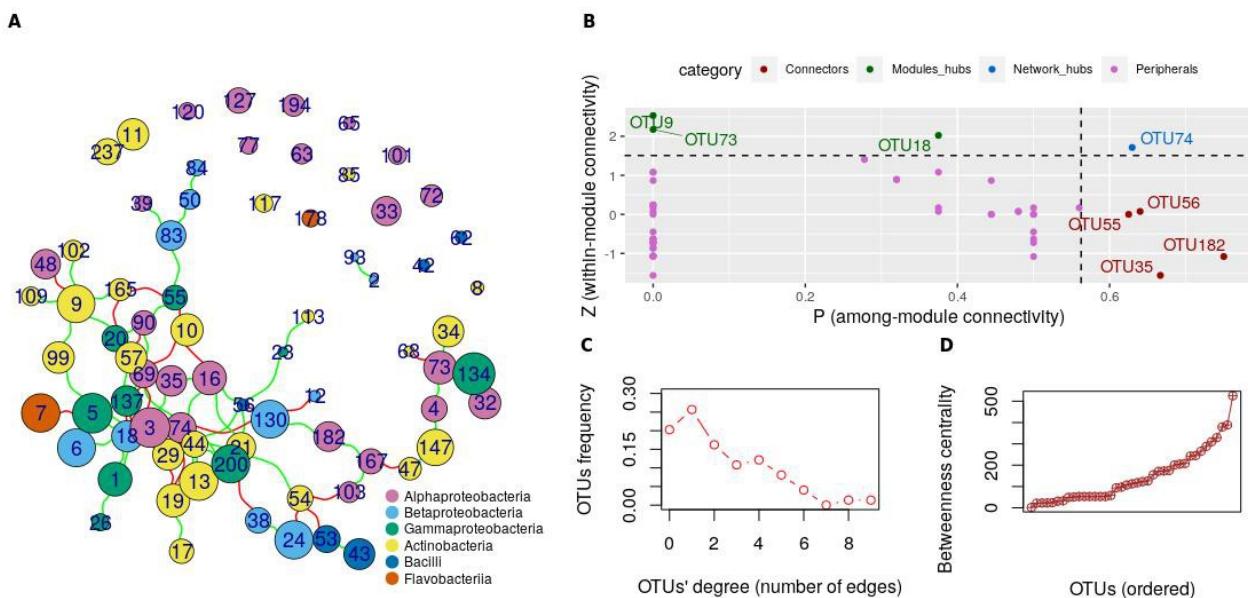
Phylogenetic clustering within the network modules was computed as follows. For each module of the network, the phylogenetic distances between all the members (every possible pair) were computed. All intra-modules distances were then gathered into a single dataset ($n=208$). Then, for each pair of modules, the phylogenetic distances between every member of the first module and every member of the second module were computed. All these distances were also gathered into a single dataset ($n=1067$). The difference of distributions of phylogenetic distances between intra and inter-modules OTUs were tested in R with a one-sided Wilcoxon rank sum-test.

2.6 Accumulative null models

Functional redundancy within the network's modules was tested against null models, one for the reactome (all chemical reactions found in the strains' GEM), and one for the core reactome (all chemical reactions common to every strain's GEM), both along the number of bacterial strains. For an increasing number of GEMs (i.e strains) going from 1 to 35, 500 random picks were done, picked from the set of 193 GEM. For each pick, the reactome and the core reactome were recorded from the GEM. Thus, for each number of strains, a null distribution of reactomes and core reactomes were set. For each module, its reactome and core reactome were computed, by adding one-by-one (in a random order) the reactome of each strain of the module.

3 Results

3.1 The co-occurrences network follows a scale-free topology



*Figure 1: Co-occurrences network's properties. A) Co-occurrences network of OTUs mostly composed of root-derived bacterial strains of a representative consortia of *A. thaliana* root microbiota. Red and green edges represent negative and positive associations. Node sizes match the OTU abundance in all samples. Node colors match their taxonomy. B) ZP-plot of the network's nodes. P is the connectivity of a node to the members of other modules, while Z is the connectivity of a node to the members of its own module. C) Nodes' degrees plot of the network: following the scale-free network topology, most nodes have few edges while a few have many edges. D) Centrality of the network nodes. Connex to the nodes' degrees, some nodes are central (many paths go through them) while others are peripherals.*

The basis of the study was to build a co-occurrences networks of bacterial strains in root samples of *A. thaliana* microbiota (germ-free younglings inoculated with a synthetic consortia). As expected with the SpiecEasi method, the network was sparse (figure 1A) which can also be explained by a limited microbial diversity contained in the initial inoculum (Bai et al., 2015). Most connections between OTUs were positive, only a few OTUs were negatively associated. When identifying key module members with their within and among-module connectivity (Z-P values) (Deng et al., 2012), 3 OTUs (9 -*Arthrobacter*-, 18 -*Variovorax*-, 73 -*Rhizobium*-) were assigned as modules hubs, 4 OTUs (35 -*Afipia*-, 55 -*Rhizobacter*-, 56 -*Brevibacillus*, 182 -*Ensifer*-) as modules connectors, one OTU (74 -*Mesorhizobium+Aminobacter*-) as network hubs (figure 1B). It appeared that the network had a scale-free topology, that is to say only a few nodes with a high degree (i.e. number of connections) and many nodes with a weak degree (figure 1C). All the OTUs found to be relevant in their among and within-module connectivity were included in the OTUs (i.e. nodes) with highest centrality measures. Accordingly, some nodes displayed a high centrality (i.e. were part of many

paths of the network) while others were peripherals (figure 1D). 19 OTUs were isolated from the network, and among them 15 did not have any edge.

3.2 Phylogenetic clustering and functional redundancy in modules

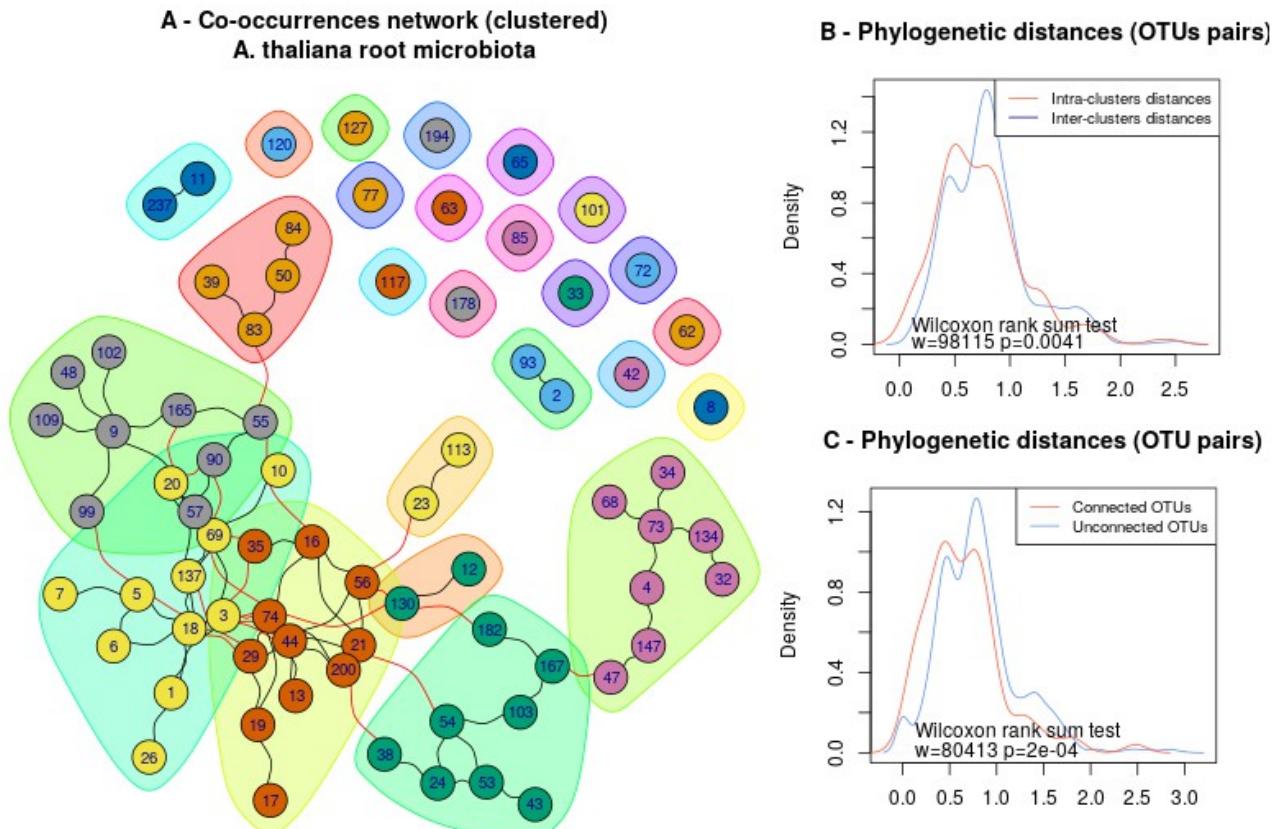


Figure 2: Co-occurrences network's modules and phylogenetic clustering. A) Modules of the co-occurrences networks displayed on figure 1A. B) Distribution (density curves) of phylogenetic distances between OTUs pairs for OTUs in the same module (red) and OTUs pairs in different modules (blue). C) Distribution (density curves) of phylogenetic distances between directly connected OTUs (red) and not directly connected OTUs (blue).

The co-occurrence network was clustered into several modules with the betweenness centrality algorithm. In the biggest connex part of the network (i.e. without all the unconnected nodes), 8 modules were returned, containing 2 to 11 OTUs (1 to 27 strains). It appeared that phylogenetic distances between members (strains in OTUs) of the same module were slightly but significantly smaller than that of members of different modules (Figure 2B, Wilcoxon, Mann & Whitney $p = 0.0041$). This result remained true when considering the difference in phylogenetic distance between connected and unconnected OTUs (figure 2C, Wilcoxon, Mann & Whitney $p = 2e-4$). In addition to phylogenetic clustering, there was also a functional redundancy within modules. Indeed, most modules displayed less diverse chemical reactions (figure 3A) and more shared chemical reactions (figure 3B) between strains compared to combinations of strains expected from null models.

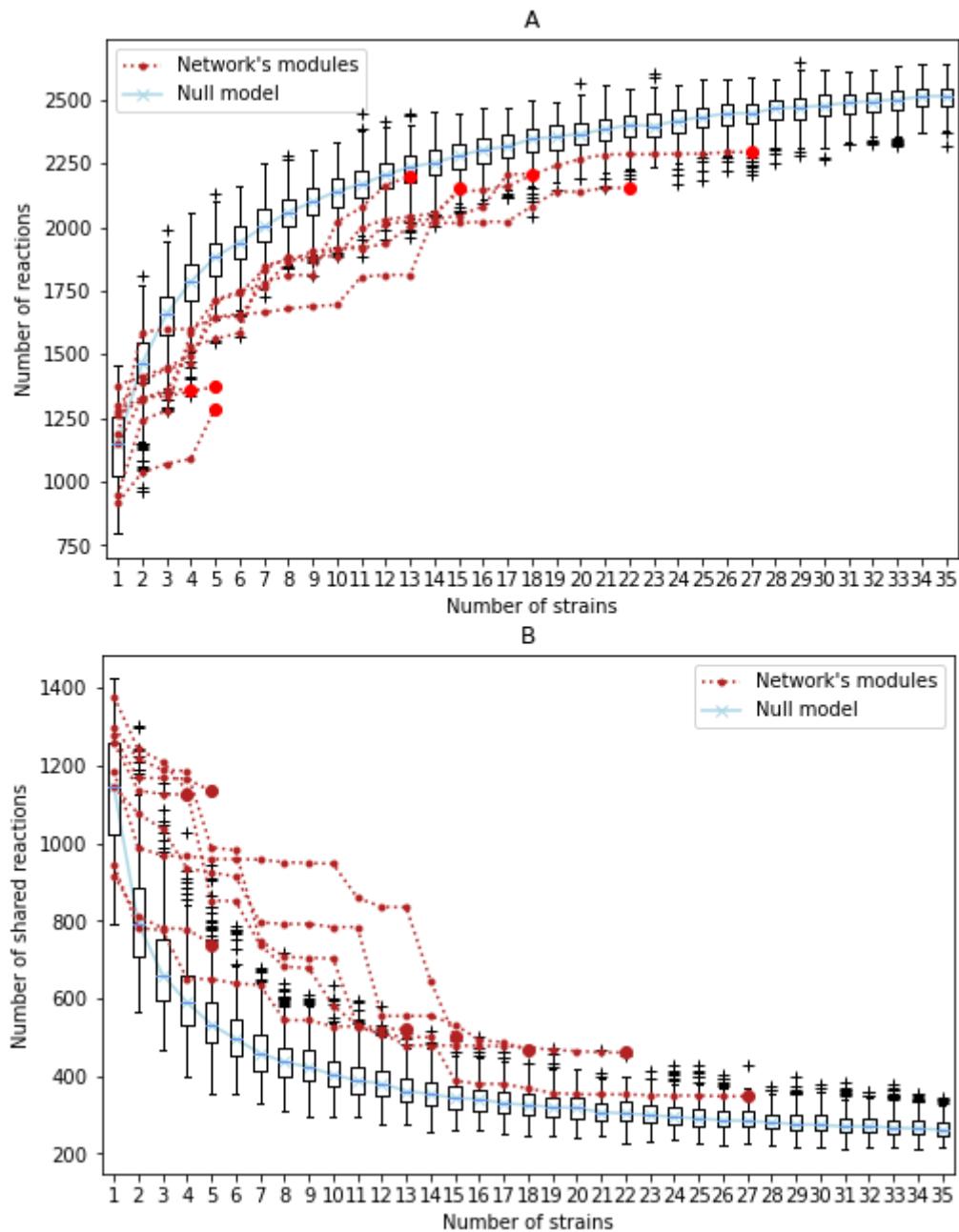


Figure 3: Modules' metabolic properties compared to null models. A) Metabolic diversity (in terms of number of different chemical reactions) of modules compared to a null model. B) metabolic redundancy, i.e. shared chemical reactions between strains of the same module, compared to a null model. Nulls models were computed with random picks of the bacterial strains pool

4 Discussion

Many studies involving co-occurrences networks attempt to describe the structure of microbial communities. For instance, they are useful to analyze the impact of fertilizers on soil microbial communities (Ishimoto et al., 2021), to compare microbiota of spatially segregated populations of the same species (Riera and Baldo, 2020), to detect response to herbivory (Malacrinò et al., 2021), or to compare diets (Read et al., 2019). In our study, the network we built is oriented towards the structure of the community after its establishment and colonization of the roots of young host-plants, reflecting the early life of the system.

4.1 The network is a basis to hypothesize interactions, notably between keystone species and the host

Centrality measures are commonly used to find important nodes in a network, i.e. keystone OTUs/species with a relevant ecological importance in a microbial community (Layeghifard et al., 2017). Keystone species are species that are, regardless of their abundance, considered as central to the structure and function of their community (Agler et al., 2016). Our network followed a scale-free topology, that describes networks with a few nodes highly connected with a high degree and a high betweenness (i.e. a high centrality), and many nodes displaying the opposite (figure 1C&D) (Layeghifard et al., 2017). Such a topology is common in microbiology and other fields, thus in accordance with the literature, and the alteration of central nodes is thought to be potentially highly detrimental (Ma et al., 2020). Such keystone nodes (species) suggest the idea of a limited set of core microbes among microbial communities. However, our network is inferred from a reduced inoculum of bacteria, which might have overly simplified the result. Indeed, networks inferred from full natural samples are usually much more dense because of the high number of species they contain, creating numerous paths even if they still follow a scale-free topology. Hence, in nature, core microbiota are bigger and denser, and likely allow a sufficient connection between members and maintain the system resilience when faced with perturbations.

Thus, our first hypothesis was validated. In addition to these centrality measures, the computation of among and within-modules connectivity highlighted 3 module hubs, 4 module connectors, and one network hub (figure 1B). Most of the strains contained in these OTUs belonged to genus that are known to have symbiotic or beneficial relation with the host-plant, such as nitrogen fixation (Rogel et al., 2001; Pini et al., 2011; Satola et al., 2013; De Meyer et al., 2019). Hence, species highly interacting with the host plant are likely to be relevant into the microbiota

structure. For example, we suggest that such keystone species could be the origin of niche engineering, due to unique secretions that would be the result of their interaction with the host. New niches would then be colonized by other species, likely specialists, because generalists are usually identified as peripheral nodes (Deng et al., 2012), that might be less involved in relations with the host and could profit from the engineered environment by keystone species. Co-occurrence networks are known to recapitulate interaction networks in some conditions, but they present some issues. First, spurious correlations (thus interactions) can arise in the neighborhood of highly connected nodes (likely keystones). Second, according to some studies co-occurrences networks could be more suited to decipher competitive interactions, but not to predation or parasitism (Hirano and Takemoto, 2019). Third, co-occurrences networks lose their interpretability when habitat filtering has a significant effect (Berry and Widder, 2014b), which was detected in our network (figure 2 and 3).

4.2 Phylogenetic and functional clustering reveal habitat filtering

Patterns of co-occurrences in microorganisms are known to be driven by deterministic processes. (Horner-Devine et al., 2007). Among those, it appears that biotic and abiotic environmental factors play a great influence in selecting the species that can colonize an environment, a process referred to as habitat filtering. Habitat filtering is a process which selects for the best trait values that give tolerance to an environment, leading to a functional convergence among species (Zhang et al., 2017). Our co-occurrences network displayed significant patterns of phylogenetic and functional clustering. Directly connected OTUs were slightly, but significantly more phylogenetically related than unconnected ones, as well as intra-modules OTUs and inter-modules OTUs (figure 2B). OTUs within a module had a more similar metabolism (in terms of chemical reactions diversity and redundancy) than expected by a null model. This result validated our second hypothesis that assumed a correlation between phylogenetic relatedness and closeness in the network, and also highlighted that the functional clustering was correlated with the phylogenetic clustering (figure 3).

Phylogenetic clustering has been frequently observed in natural communities (Horner-Devine and Bohannan, 2006; Violle et al., 2011; Mondav et al., 2017), and is recognized to be a proxy of habitat filtering (Gerhold et al., 2015). Niche analysis revealed strong niche overlaps between co-occurring bacteria, highlighting shared ecological properties between members of the community (Michalska-Smith et al., 2021). Indeed, phylogenetically close species are often functionally similar (Martiny et al., 2015), a redundancy which also reflects habitat filtering. However, phylogenetic clustering can also arise from biotic interactions, in cases where dissimilar species have been

excluded due to weak competition abilities, leaving relatively similar species (Goberna et al., 2014; Zhang et al., 2017).

Finally, our co-occurrences network harbored mostly positive edges and relatively few negative ones. Negative edges are supposed to translate co-exclusion mechanisms, including differential niche adaptation (Faust et al., 2012). Hence, low proportions of negative edges suggest a prevalence of overlapping niches between the members of the community (Ma et al., 2020), strengthening the possibility of habitat filtering in this environment. One could think about conflicting statements here. Indeed, overlapping niches and phylogenetic relatedness are likely synonyms of competition, however the weak prevalence of negative edges suggests otherwise. However, we mentioned above the fact that co-occurrences networks struggle to infer correct interactions when habitat filtering is important, a statement that reconciles these statements.

5 Conclusion and prospects

This co-occurrences network analysis on *A. thaliana* root microbiota extended many previous studies highlighting both the scale-free topology and the phylogenetic and functional clustering of microbial communities. The first constitutes a basis for inferring deeper interaction networks, and the latter shed light on environmental filtering, most likely by the host plant. A perspective to this work would be to proceed to a similar analysis on *A. thaliana*'s leaf microbiota, both with a leaf-derived inoculum and the same root-derived inoculum. Such approaches would both highlight the structure of the leaf microbiota, and the microbes that are able to establish in both compartments, or transition from roots to leaves. Another possibility would be to investigate how generalist and specialist species are distributed in the networks. For example, according to (Barberán, Bates, Casamayor, & Fierer (2012), generalist species are species encountered in most samples in relatively high to high abundances, while specialists are encountered in few samples, in high abundances.

Some limits must however be mentioned, particularly the method of identifying keystone species with co-occurrences networks, that has not yet arrived at a consensus. The validity of this approach should be subject to experimental testing of removal/addition of keystone taxa in microbial communities, which is challenging to do (Röttjers and Faust, 2019). Co-occurrence networks approaches are not interaction networks, yet enable exploring biotic interactions, but do not permit to go much further since they rather highlight statistically significant co-occurrences or co-exclusion (Röttjers and Faust, 2019). Hence, they need to be combined with more functional-oriented data and integrated to other approaches, for example with Lotka-Volterra models (Berry

and Widder, 2014b), metabolic networks to study putative competition or cooperation for metabolites (Muller et al., 2018), or experimental validation of co-occurrences and co-exclusions.. They are then more useful to describe a community structure with many unknowns about how connected species interact, and generate hypotheses to be tested (Hartman et al., 2018; Ramirez et al., 2018).

Acknowledgments and fundings

This work was supported by a grant from the French Ministry for Research and Innovation, by a grant from the CNRS (EC2CO), and also by a starting grant to S. Hacquard from a European Research Council (MICRORULES 758003), the Max Planck Institute, the Cluster of Excellence on Plant Sciences (CEPLAS), and the ‘Priority Programme: Deconstruction and Reconstruction of the Plant Microbiota (SPP DECRyPT 2125)’, both funded by the Deutsche Forschungsgemeinschaf.

Competing interests

The authors declare no competing interests.

General Bibliography

- Agler, M. T., Ruhe, J., Kroll, S., Morhenn, C., Kim, S. T., Weigel, D., et al. (2016). Microbial Hub Taxa Link Host and Abiotic Factors to Plant Microbiome Variation. *PLoS Biol.* 14, 1–31. doi:10.1371/journal.pbio.1002352.
- Aguiar-Pulido, V., Huang, W., Suarez-Ulloa, V., Cickovski, T., Mathee, K., and Narasimhan, G. (2016). Metagenomics, metatranscriptomics, and metabolomics approaches for microbiome analysis. *Evol. Bioinforma.* 12, 5–16. doi:10.4137/EBO.S36436.
- Aguirre De Cácer, D. (2019). A conceptual framework for the phylogenetically constrained assembly of microbial communities. *Microbiome* 7, 1–11. doi:10.1186/s40168-019-0754-y.
- Aitchison, J. (1982). The Statistical Analysis of Compositional Data . Series B (Methodological), Vol . 44 , No . 2 Published by : Wiley for the Royal Statistical Society Stable URL : <https://www.j. J. R. Stat. Soc. 44, 139–177>.
- Aite, M., Chevallier, M., Frioux, C., Trottier, C., Got, J., Cortés, M. P., et al. (2018). Traceability, reproducibility and wiki-exploration for “à-la-carte” reconstructions of genome-scale metabolic models. *PLoS Comput. Biol.* 14, 1–25. doi:10.1371/journal.pcbi.1006146.
- Alneberg, J., Bennke, C., Beier, S., Bunse, C., Quince, C., Ininbergs, K., et al. (2020). Ecosystem-wide metagenomic binning enables prediction of ecological niches from genomes. *Commun. Biol.* 3, 1–10. doi:10.1038/s42003-020-0856-x.
- Alon, U. (2006). *An Introduction to Systems Biology : Design Principles of Biological Circuits.* , ed. T. & F. Group NW: Chapman & Hall/CRC.
- Amor, D. R., and Bello, M. D. (2019). Bottom-up approaches to synthetic cooperation in microbial communities. *Life* 9. doi:10.3390/life9010022.
- Andreote, F. D., Gumiere, T., and Durrer, A. (2014). Exploring interactions of plant microbiomes. *Sci. Agric.* 71, 528–539. doi:10.1590/0103-9016-2014-0195.
- Angert, A. L., Huxman, T. E., Chesson, P., and Venable, D. L. (2009). Functional tradeoffs determine species coexistence via the storage effect. *Proc. Natl. Acad. Sci. U. S. A.* 106, 11641–11645. doi:10.1073/pnas.0904512106.
- Badri, D. V., and Vivanco, J. M. (2009). Regulation and function of root exudates. *Plant, Cell Environ.* 32, 666–681. doi:10.1111/j.1365-3040.2009.01926.x.
- Bai, Y., Müller, D. B., Srinivas, G., Garrido-oter, R., Potthoff, E., Rott, M., et al. (2015). Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature* 528, 364–369. doi:10.1038/nature16192.
- Bakker, M. G., Schlatter, D. C., Otto-Hanson, L., and Kinkel, L. L. (2014). Diffuse symbioses: Roles of plant-plant, plant-microbe and microbe-microbe interactions in structuring the soil microbiome. *Mol. Ecol.* 23, 1571–1583. doi:10.1111/mec.12571.
- Balser, T. C., McMahon, K. D., Bart, D., Bronson, D., Coyle, D. R., Craig, N., et al. (2006). Bridging the gap between micro - and macro-scale perspectives on the role of microbial communities in global change ecology. *Plant Soil* 289, 59–70. doi:10.1007/s11104-006-9104-5.
- Baran, R., Brodie, E. L., Mayberry-Lewis, J., Hummel, E., Da Rocha, U. N., Chakraborty, R., et al. (2015). Exometabolite niche partitioning among sympatric soil bacteria. *Nat. Commun.* 6. doi:10.1038/ncomms9289.
- Barberán, A., Bates, S. T., Casamayor, E. O., and Fierer, N. (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* 6, 343–351. doi:10.1038/ismej.2011.119.

General bibliography

- Barberán, A., Ramirez, K. S., Leff, J. W., Bradford, M. A., Wall, D. H., and Fierer, N. (2014). Why are some microbes more ubiquitous than others? Predicting the habitat breadth of soil bacteria. *Ecol. Lett.* 17, 794–802. doi:10.1111/ele.12282.
- Basan, M., Hui, S., Okano, H., Zhang, Z., Shen, Y., Williamson, J. R., et al. (2015). Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nature* 528, 99–104. doi:10.1038/nature15765.
- Bashan, A., Gibson, T. E., Friedman, J., Carey, V. J., Weiss, S. T., Hohmann, E. L., et al. (2016). Universality of human microbial dynamics. *Nature* 534, 259–262. doi:10.1038/nature18301.
- Baudoin, E., Benizri, E., and Guckert, A. (2003). Impact of artificial root exudates on the bacterial community structure in bulk soil and maize rhizosphere. *Soil Biol. Biochem.* 35, 1183–1192. doi:10.1016/S0038-0717(03)00179-2.
- Bednarek, P., Kwon, C., and Schulze-Lefert, P. (2010). Not a peripheral issue: Secretion in plant-microbe interactions. *Curr. Opin. Plant Biol.* 13, 378–387. doi:10.1016/j.pbi.2010.05.002.
- Belcour, A., Frioux, C., Aite, M., Hildebrand, F., and Siegel, A. (2020). Metage2Metabo, microbiota-scale metabolic complementarity for the identification of key species. *eLife* 9, 1–33.
- Belzer, C., Chia, L. W., Aalvink, S., Chamlagain, B., Piironen, V., Knol, J., et al. (2017). Microbial metabolic networks at the mucus layer lead to diet-independent butyrate and vitamin B12 production by intestinal symbionts. *MBio* 8. doi:10.1128/mBio.00770-17.
- Bérard, A., Sassi, M. Ben, Kaisermann, A., and Renault, P. (2015). Soil microbial community responses to heat wave components: Drought and high temperature. *Clim. Res.* 66, 243–264. doi:10.3354/cr01343.
- Berendsen, R. L., Pieterse, C. M. J., and Bakker, P. A. H. M. (2012). The rhizosphere microbiome and plant health. *Trends Plant Sci.* 17, 478–486. doi:10.1016/j.tplants.2012.04.001.
- Bernstein, D. B., Dewhurst, F. E., and Segrè, D. (2019). Metabolic network percolation quantifies biosynthetic capabilities across the human oral microbiome. *eLife* 8, 1–33. doi:10.7554/eLife.39733.001.
- Berry, D., and Widder, S. (2014a). Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.* 5. doi:10.3389/fmicb.2014.00219.
- Berry, D., and Widder, S. (2014b). Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.* 5, 1–14. doi:10.3389/fmicb.2014.00219.
- Biggs, M. B., Medlock, G. L., Moutinho, T. J., Lees, H. J., Swann, J. R., Kolling, G. L., et al. (2017). Systems-level metabolism of the altered Schaedler flora, a complete gut microbiota. *ISME J.* 11, 426–438. doi:10.1038/ismej.2016.130.
- Blasche, S., Kim, Y., Mars, R. A. T., Machado, D., Maansson, M., Kafkia, E., et al. (2021). Metabolic cooperation and spatiotemporal niche partitioning in a kefir microbial community. *Nat. Microbiol.* 6, 196–208. doi:10.1038/s41564-020-00816-5.
- Bodenhausen, N., Bortfeld-Miller, M., Ackermann, M., and Vorholt, J. A. (2014). A Synthetic Community Approach Reveals Plant Genotypes Affecting the Phyllosphere Microbiota. *PLoS Genet.* 10. doi:10.1371/journal.pgen.1004283.
- Boon, E., Meehan, C. J., Whidden, C., Wong, D. H. J., Langille, M. G. I., and Beiko, R. G. (2014). Interactions in the microbiome: Communities of organisms and communities of genes. *FEMS Microbiol. Rev.* 38, 90–118. doi:10.1111/1574-6976.12035.
- Bordbar, A., Monk, J. M., King, Z. A., and Palsson, B. O. (2014). Constraint-based models predict metabolic and associated cellular functions. *Nat. Rev. Genet.* 15, 107–120. doi:10.1038/nrg3643.

General bibliography

- Bordenstein, S. R., and Theis, K. R. (2015). Host biology in light of the microbiome: Ten principles of holobionts and hologenomes. *PLoS Biol.* 13, 1–23. doi:10.1371/journal.pntd.0005749.
- Bordron, P., Latorre, M., Cortés, M. P., González, M., Thiele, S., Siegel, A., et al. (2016). Putative bacterial interactions from metagenomic knowledge with an integrative systems ecology approach. *Microbiologyopen* 5, 106–117. doi:10.1002/mbo3.315.
- Borenstein, E., Kupiec, M., Feldman, M. W., and Ruppin, E. (2008). Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proc. Natl. Acad. Sci. U. S. A.* 105, 14482–14487. doi:10.1073/pnas.0806162105.
- Borer, B., Ciccarese, D., Johnson, D., and Or, D. (2020). Spatial organization in microbial range expansion emerges from trophic dependencies and successful lineages. *Commun. Biol.* 3, 1–10. doi:10.1038/s42003-020-01409-y.
- Borer, B., Tecon, R., and Or, D. (2018). Spatial organization of bacterial populations in response to oxygen and carbon counter-gradients in pore networks. *Nat. Commun.* 9. doi:10.1038/s41467-018-03187-y.
- Bosi, E., Bacci, G., Mengoni, A., and Fondi, M. (2017). Perspectives and challenges in microbial communities metabolic modeling. *Front. Genet.* 8, 1–9. doi:10.3389/fgene.2017.00088.
- Brochet, S., Quinn, A., Mars, R. A., Neuschwander, N., Sauer, U., and Engel, P. (2021). Niche partitioning facilitates coexistence of closely related honey bee gut bacteria. *Elife* 10. doi:10.7554/elife.68583.
- Bruggeman, F. J., and Westerhoff, H. V. (2007). The nature of systems biology. *Trends Microbiol.* 15, 45–50. doi:10.1016/j.tim.2006.11.003.
- Budinich, M., Bourdon, J., Larhlimi, A., and Eveillard, D. (2017). A multi-objective constraint-based approach for modeling genome-scale microbial ecosystems. *PLoS One* 12, 1–22. doi:10.1371/journal.pone.0171744.
- Bui, T. P. N., Schols, H. A., Jonathan, M., Stams, A. J. M., de Vos, W. M., and Plugge, C. M. (2019). Mutual Metabolic Interactions in Co-cultures of the Intestinal Anaerostipes rhamnosivorans With an Acetogen, Methanogen, or Pectin-Degrader Affecting Butyrate Production. *Front. Microbiol.* 10, 1–12. doi:10.3389/fmicb.2019.02449.
- Bulgarelli, D., Schlaeppi, K., Spaepen, S., Van Themaat, E. V. L., and Schulze-Lefert, P. (2013). Structure and functions of the bacterial microbiota of plants. *Annu. Rev. Plant Biol.* 64, 807–838. doi:10.1146/annurev-arplant-050312-120106.
- Burson, A., Stomp, M., Mekkes, L., and Huisman, J. (2019). Stable coexistence of equivalent nutrient competitors through niche differentiation in the light spectrum. *Ecology* 100, 1–12. doi:10.1002/ecy.2873.
- Çakir, T., Panagiotou, G., Uddin, R., and Durmuş, S. (2020). Novel Approaches for Systems Biology of Metabolism-Oriented Pathogen-Human Interactions: A Mini-Review. *Front. Cell. Infect. Microbiol.* 10. doi:10.3389/fcimb.2020.00052.
- Caldwell, J. P., and Vitt, L. J. (1999). Dietary Asymmetry in Leaf Litter Frogs and Lizards in a Transitional Northern Amazonian Rain Forest Published by : Wiley on behalf of Nordic Society Oikos Stable URL : <https://www.jstor.org/stable/3546419> REFERENCES Linked references are available on JSTOR. 84, 383–397.
- Callahan, B. J., Fukami, T., and Fisher, D. S. (2014). Rapid evolution of adaptive niche construction in experimental microbial populations. *Evolution (N. Y.)* 68, 3307–3316. doi:10.1111/evo.12512.
- Caporaso, J. G., Paszkiewicz, K., Field, D., Knight, R., and Gilbert, J. A. (2012). The Western English Channel contains a persistent microbial seed bank. *ISME J.* 6, 1089–1093. doi:10.1038/ismej.2011.162.
- Carlson, R. P. (2007). Metabolic systems cost-benefit analysis for interpreting network structure and regulation. *Bioinformatics* 23, 1258–1264. doi:10.1093/bioinformatics/btm082.

General bibliography

- Carlson, R. P., Beck, A. E., Phalak, P., Fields, M. W., Gedeon, T., Hanley, L., et al. (2018). Competitive resource allocation to metabolic pathways contributes to overflow metabolisms and emergent properties in cross-feeding microbial consortia. *Biochem. Soc. Trans.* 46, 269–284. doi:10.1042/BST20170242.
- Carrero-Colón, M., Nakatsu, C. H., and Konopka, A. (2006). Effect of nutrient periodicity on microbial community dynamics. *Appl. Environ. Microbiol.* 72, 3175–3183. doi:10.1128/AEM.72.5.3175-3183.2006.
- Caspi, R., Foerster, H., Fulcher, C. A., Kaipa, P., Krummenacker, M., Latendresse, M., et al. (2008). The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 36, 623–631. doi:10.1093/nar/gkm900.
- Cavaliere, M., Feng, S., Soyer, O. S., and Jiménez, J. I. (2017). Cooperation in microbial communities and their biotechnological applications. *Environ. Microbiol.* 19, 2949–2963. doi:10.1111/1462-2920.13767.
- Celiker, H., and Gore, J. (2012). Competition between species can stabilize public-goods cooperation within a species. *Mol. Syst. Biol.* 8, 1–9. doi:10.1038/msb.2012.54.
- Chai, Q., Li, Y., Li, X., Wu, W., Peng, H., Jia, R., et al. (2019). Assessment of variation in paddy microbial communities under different storage temperatures and relative humidity by Illumina sequencing analysis. *Food Res. Int.* 126, 108581. doi:10.1016/j.foodres.2019.108581.
- Chase, J. M. (2003). Community assembly: When should history matter? *Oecologia* 136, 489–498. doi:10.1007/s00442-003-1311-7.
- Chomicki, G., Werner, G. D. A., West, S. A., and Kiers, E. T. (2020). Compartmentalization drives the evolution of symbiotic cooperation: Compartmentalisation drives symbiosis. *Philos. Trans. R. Soc. B Biol. Sci.* 375. doi:10.1098/rstb.2019.0602rstb20190602.
- Chubukov, V., Mukhopadhyay, A., Petzold, C. J., Keasling, J. D., and Martín, H. G. (2018). Synthetic and systems biology for microbial production of commodity chemicals. *npj Syst. Biol. Appl.* 2, 1–11. doi:10.1038/npgsba.2016.9.
- Clark, J. S. (2009). Beyond neutral science. *Trends Ecol. Evol.* 24, 8–15. doi:10.1016/j.tree.2008.09.004.
- Cole, B. J., Feltcher, M. E., Waters, R. J., Wetmore, K. M., Mucyn, T. S., Ryan, E. M., et al. (2017). Genome-wide identification of bacterial plant colonization genes. *PLoS Biol.* 15, 1–24. doi:10.1371/journal.pbio.2002860.
- Coleman-Derr, D., Desgarennes, D., Fonseca-Garcia, C., Gross, S., Clingenpeel, S., Woyke, T., et al. (2016). Plant compartment and biogeography affect microbiome composition in cultivated and native Agave species. *New Phytol.* 209, 798–811. doi:10.1111/nph.13697.
- Collet, G., Eveillard, D., Gebser, M., and Prigent, S. Extending the Metabolic Network of Ectocarpus Siliculosus Using Answer Set Programming. 245–256.
- Compan, S., Samad, A., Faist, H., and Sessitsch, A. (2019). A review on the plant microbiome: Ecology, functions, and emerging trends in microbial application. *J. Adv. Res.* 19, 29–37. doi:10.1016/j.jare.2019.03.004.
- Cordero, O. X., and Datta, M. S. (2016). Microbial interactions and community assembly at microscales. *Curr. Opin. Microbiol.* 31, 227–234. doi:10.1016/j.mib.2016.03.015.
- Coyte, K. Z., and Rakoff-Nahoum, S. (2019). Understanding Competition and Cooperation within the Mammalian Gut Microbiome. *Curr. Biol.* 29, R538–R544. doi:10.1016/j.cub.2019.04.017.
- Coyte, K. Z., Schluter, J., and Foster, K. R. (2015). The ecology of the microbiome: Networks, competition, and stability. *Science (80-.).* 350, 663–666. doi:10.1126/science.aad2602.
- Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J., and Smith, A. G. (2005). Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* 438, 90–93. doi:10.1038/nature04056.

General bibliography

- Csardi, G., and Tamas, N. (2006). The igraph software package for complex network research. *InterJournal Complex Sy*, 2191–2193. doi:10.3724/sp.j.1087.2009.02191.
- Cui, J., Yuan, X., Zhang, Q., Zhou, J., Lin, K., Xu, J., et al. (2021). Nutrient availability is a dominant predictor of soil bacterial and fungal community composition after nitrogen addition in subtropical acidic forests. *PLoS One* 16, 1–20. doi:10.1371/journal.pone.0246263.
- Curiel Yuste, J., Fernandez-Gonzalez, A. J., Fernandez-Lopez, M., Ogaya, R., Penuelas, J., and Lloret, F. (2014). Functional diversification within bacterial lineages promotes wide functional overlapping between taxonomic groups in a Mediterranean forest soil. *FEMS Microbiol. Ecol.* 90, 54–67. doi:10.1111/1574-6941.12373.
- D’Souza, G., and Kost, C. (2016). Experimental Evolution of Metabolic Dependency in Bacteria. *PLoS Genet.* 12, 1–27. doi:10.1371/journal.pgen.1006364.
- D’Souza, G., Shitut, S., Preussger, D., Yousif, G., Waschina, S., and Kost, C. (2018). Ecology and evolution of metabolic cross-feeding interactions in bacteria. *Nat. Prod. Rep.* 35, 455–488. doi:10.1039/c8np00009c.
- D’Souza, G., Waschina, S., Pande, S., Bohl, K., Kaleta, C., and Kost, C. (2014). Less is more: Selective advantages can explain the prevalent loss of biosynthetic genes in bacteria. *Evolution (N. Y.)* 68, 2559–2570. doi:10.1111/evo.12468.
- Dal Co, A., Van Vliet, S., and Ackermann, M. (2019). Emergent microscale gradients give rise to metabolic cross-feeding and antibiotic tolerance in clonal bacterial populations. *Philos. Trans. R. Soc. B Biol. Sci.* 374. doi:10.1098/rstb.2019.0080.
- De Meyer, S. E., Andrews, M., James, E. K., and Willems, A. (2019). Mesorhizobium carmichaelinearum sp. nov., isolated from Carmichaeliae spp. root nodules. *Int. J. Syst. Evol. Microbiol.* 69, 146–152. doi:10.1099/ijsem.0.003120.
- De Vos, M. G. J., Zagorski, M., McNally, A., and Bollenbach, T. (2017). Interaction networks, ecological stability, and collective antibiotic tolerance in polymicrobial infections. *Proc. Natl. Acad. Sci. U. S. A.* 114, 10666–10671. doi:10.1073/pnas.1713372114.
- Degnan, P. H., Taga, M. E., and Goodman, A. L. (2014). Vitamin B12 as a modulator of gut microbial ecology. *Cell Metab.* 20, 769–778. doi:10.1016/j.cmet.2014.10.002.
- Deng, Y., Jiang, Y. H., Yang, Y., He, Z., Luo, F., and Zhou, J. (2012). Molecular ecological network analyses. *BMC Bioinformatics* 13. doi:10.1186/1471-2105-13-113.
- Dombrowski, N., Seitz, K. W., Teske, A. P., and Baker, B. J. (2017). Genomic insights into potential interdependencies in microbial hydrocarbon and nutrient cycling in hydrothermal sediments. *Microbiome* 5, 106. doi:10.1186/s40168-017-0322-2.
- Dong, X., and Fisher, S. G. (2019). Ecosystem spatial self-organization: Free order for nothing? *Ecol. Complex.* 38, 24–30. doi:10.1016/j.ecocom.2019.01.002.
- Douglas, A. E. (2020). The microbial exometabolome: Ecological resource and architect of microbial communities. *Philos. Trans. R. Soc. B Biol. Sci.* 375. doi:10.1098/rstb.2019.0250.
- Drenovsky, R. E., Steenwerth, K. L., Jackson, L. E., and Scow, K. M. (2010). Land use and climatic factors structure regional patterns in soil microbial communities. *Glob. Ecol. Biogeogr.* 19, 27–39. doi:10.1111/j.1466-8238.2009.00486.x.
- Drew, G. C., Stevens, E. J., and King, K. C. (2021). Microbial evolution and transitions along the parasite–mutualist continuum. *Nat. Rev. Microbiol.* 0123456789. doi:10.1038/s41579-021-00550-7.

General bibliography

- Duhamel, M., and Vandenkoornhuyse, P. (2013). Sustainable agriculture: Possible trajectories from mutualistic symbiosis and plant neodomestication. *Trends Plant Sci.* 18, 597–600. doi:10.1016/j.tplants.2013.08.010.
- Easson, C. G., and Thacker, R. W. (2014). Phylogenetic signal in the community structure of host-specific microbiomes of tropical marine sponges. *Front. Microbiol.* 5, 1–11. doi:10.3389/fmicb.2014.00532.
- Ebrahimi, A., Schwartzman, J., and Cordero, O. X. (2019). Multicellular behaviour enables cooperation in microbial cell aggregates. *Philos. Trans. R. Soc. B Biol. Sci.* 374. doi:10.1098/rstb.2019.0077.
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi:10.1093/bioinformatics/btq461.
- Erlandson, S., Wei, X., Savage, J., Cavender-Bares, J., and Peay, K. (2018). Soil abiotic variables are more important than Salicaceae phylogeny or habitat specialization in determining soil microbial community structure. *Mol. Ecol.* 27, 2007–2024. doi:10.1111/mec.14576.
- Escalante, A. E., Rebollo-Gómez, M., Benítez, M., and Travisano, M. (2015). Ecological perspectives on synthetic biology: Insights from microbial population biology. *Front. Microbiol.* 6, 1–10. doi:10.3389/fmicb.2015.00143.
- Ewald, P. W. (1987). Transmission Modes and Evolution of the Parasitism-Mutualism Continuum. *Ann. N. Y. Acad. Sci.* 503, 295–306. doi:10.1111/j.1749-6632.1987.tb40616.x.
- Fahimipour, A. K., and Gross, T. (2020). Mapping the bacterial metabolic niche space. *Nat. Commun.* 11, 1–8. doi:10.1038/s41467-020-18695-z.
- Faust, K., Bauchinger, F., Laroche, B., de Buyl, S., Lahti, L., Washburne, A. D., et al. (2018). Signatures of ecological processes in microbial community time series. *Microbiome* 6, 1–13. doi:10.1186/s40168-018-0496-2.
- Faust, K., and Raes, J. (2012). Microbial interactions: From networks to models. *Nat. Rev. Microbiol.* 10, 538–550. doi:10.1038/nrmicro2832.
- Faust, K., Sathirapongsasuti, J. F., Izard, J., Segata, N., Gevers, D., Raes, J., et al. (2012). Microbial Co-occurrence Relationships in the Human Microbiome. *PLoS Comput. Biol.* 8, 410–414. doi:10.1371/journal.pcbi.1002606.
- Feist, A. M., Herrgård, M. J., Thiele, I., Reed, J. L., and Palsson, B. (2009). Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.* 7, 129–143. doi:10.1038/nrmicro1949.
- Ferriere, R., Bronstein, J. L., Rinaldi, S., Law, R., and Gauduchon, M. (2002). Cheating and the evolutionary stability of mutualisms. *Proc. R. Soc. B Biol. Sci.* 269, 773–780. doi:10.1098/rspb.2001.1900.
- Fierer, N. (2017). Embracing the unknown: Disentangling the complexities of the soil microbiome. *Nat. Rev. Microbiol.* 15, 579–590. doi:10.1038/nrmicro.2017.87.
- Flamholz, A., Noor, E., Bar-Even, A., Liebermeister, W., and Milo, R. (2013). Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc. Natl. Acad. Sci. U. S. A.* 110, 10039–10044. doi:10.1073/pnas.1215283110.
- Foo, J. L., Ling, H., Lee, Y. S., and Chang, M. W. (2017). Microbiome engineering: Current applications and its future. *Biotechnol. J.* 12. doi:10.1002/biot.201600099.
- Foster, K. R., and Bell, T. (2012). Competition, not cooperation, dominates interactions among culturable microbial species. *Curr. Biol.* 22, 1845–1850. doi:10.1016/j.cub.2012.08.005.
- Franzosa, E. A., Hsu, T., Sirota-Madi, A., Shafquat, A., Abu-Ali, G., Morgan, X. C., et al. (2015). Sequencing and beyond: Integrating molecular “omics” for microbial community profiling. *Nat. Rev. Microbiol.* 13, 360–372. doi:10.1038/nrmicro3451.

General bibliography

- Frederickson, J. K., Balkwill, D. L., Zahara, J. M., Li, S. M. W., Brockman, F. J., and Simmons, M. A. (1991). Physiological diversity and distributions of heterotrophic bacteria in deep cretaceous sediments of the Atlantic coastal plain. *Appl. Environ. Microbiol.* 57, 402–411. doi:10.1128/aem.57.2.402-411.1991.
- Friedman, J., and Alm, E. J. (2012). Inferring Correlation Networks from Genomic Survey Data. *PLoS Comput. Biol.* 8, 1–11. doi:10.1371/journal.pcbi.1002687.
- Friedman, J., Higgins, L. M., and Gore, J. (2017). Community structure follows simple assembly rules in microbial microcosms. *Nat. Ecol. Evol.* 1, 1–7. doi:10.1038/s41559-017-0109.
- Frioux, C., Dittami, S. M., and Siegel, A. (2020). Using automated reasoning to explore the metabolism of unconventional organisms: A first step to explore host–microbial interactions. *Biochem. Soc. Trans.* 48, 901–913. doi:10.1042/BST20190667.
- Frioux, C., Fremy, E., Trottier, C., and Siegel, A. (2018). Scalable and exhaustive screening of metabolic functions carried out by microbial consortia. *Bioinformatics* 34, i934–i943. doi:10.1093/bioinformatics/bty588.
- Fritts, R. K., McCully, A. L., and McKinlay, J. B. (2021). Extracellular Metabolism Sets the Table for Microbial Cross-Feeding. *Microbiol. Mol. Biol. Rev.* 85, 1–29. doi:10.1128/mmbr.00135-20.
- Fuhrman, J. A. (2009). Microbial community structure and its functional implications. *Nature* 459, 193–199. doi:10.1038/nature08058.
- Fukami, T., Dickie, I. A., Paula Wilkie, J., Paulus, B. C., Park, D., Roberts, A., et al. (2010). Assembly history dictates ecosystem functioning: Evidence from wood decomposer communities. *Ecol. Lett.* 13, 675–684. doi:10.1111/j.1461-0248.2010.01465.x.
- Gentile, C. L., and Weir, T. L. (2018). The gut microbiota at the intersection of diet and human health. *Science (80-.).* 362, 776–780. doi:10.1126/science.aau5812.
- Gerber, G. K. (2014). The dynamic microbiome. *FEBS Lett.* 588, 4131–4139. doi:10.1016/j.febslet.2014.02.037.
- Gerhold, P., Cahill, J. F., Winter, M., Bartish, I. V., and Prinzing, A. (2015). Phylogenetic patterns are not proxies of community assembly mechanisms (they are far better). *Funct. Ecol.* 29, 600–614. doi:10.1111/1365-2435.12425.
- Gil, R., and Peretó, J. (2015). Small genomes and the difficulty to define minimal translation and metabolic machineries. *Front. Ecol. Evol.* 3, 1–12. doi:10.3389/fevo.2015.00123.
- Giovannoni, S. J., Cameron Thrash, J., and Temperton, B. (2014). Implications of streamlining theory for microbial ecology. *ISME J.* 8, 1553–1565. doi:10.1038/ismej.2014.60.
- Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., et al. (2005). Genetics: Genome streamlining in a cosmopolitan oceanic bacterium. *Science (80-.).* 309, 1242–1245. doi:10.1126/science.1114057.
- Gloor, G. B., Macklaim, J. M., Pawlowsky-glahn, V., and Egoscue, J. J. (2017). Microbiome Datasets Are Compositional : And This Is Not Optional. 8, 1–6. doi:10.3389/fmicb.2017.02224.
- Goberna, M., García, C., and Verdú, M. (2014). A role for biotic filtering in driving phylogenetic clustering in soil bacterial communities. *Glob. Ecol. Biogeogr.* 23, 1346–1355. doi:10.1111/geb.12227.
- Goberna, M., and Verdú, M. (2016). Predicting microbial traits with phylogenies. *ISME J.* 10, 959–967. doi:10.1038/ismej.2015.171.
- Goldford, J. E., Lu, N., Bajić, D., Estrela, S., Tikhonov, M., Sanchez-Gorostiaga, A., et al. (2018). Emergent simplicity in microbial community assembly. *Science (80-.).* 361, 469–474. doi:10.1126/science.aat1168.
- Gore, J., Youk, H., and Van Oudenaarden, A. (2009). Snowdrift game dynamics and facultative cheating in yeast. *Nature* 459, 253–256. doi:10.1038/nature07921.

General bibliography

- Götzenberger, L., Bello, F. De, Brathen, K. A., Davison, J., Dubuis, A., Guisan, A., et al. (2012). Ecological assembly rules in plant communities — approaches , patterns and prospects. *Biol. Rev.* 87, 111–127. doi:10.1111/j.1469-185X.2011.00187.x.
- Goyal, A., Wang, T., Dubinkina, V., and Maslov, S. (2021). Ecology-guided prediction of cross-feeding interactions in the human gut microbiome. *Nat. Commun.* 12, 1–10. doi:10.1038/s41467-021-21586-6.
- Gralka, M., Szabo, R., Stocker, R., and Cordero, O. X. (2020). Trophic Interactions and the Drivers of Microbial Community Assembly. *Curr. Biol.* 30, R1176–R1188. doi:10.1016/j.cub.2020.08.007.
- Green, J. L., Holmes, A. J., Westoby, M., Oliver, I., Briscoe, D., Dangerfield, M., et al. (2004). Spatial scaling of microbial eukaryote diversity. *Nature* 432, 747–750. doi:10.1038/nature03034.
- Griffin, A. S., West, S. A., and Buckling, A. (2004). Cooperation and competition in pathogenic bacteria. *Nature* 430, 1021–1024. doi:10.1038/nature02802.1.
- Griggs, R. G., Steenwerth, K. L., Mills, D. A., and Bokulich, N. A. (2021). Sources and Assembly of Microbial Communities in Vineyards as a Functional Component of Winegrowing. 12. doi:10.3389/fmicb.2021.673810.
- Grover, J. P. (1994). Assembly Rules for Communities of Nutrient-Limited Plants and Specialist Herbivores Author (s): James P . Grover Source : The American Naturalist , Vol . 143 , No . 2 (Feb ., 1994), pp . 258-282 Published by : The University of Chicago Press for The A. *Am. Nat.* 143, 258–282.
- Guo, J., Ling, N., Chen, Z., Xue, C., Li, L., Liu, L., et al. (2020a). Soil fungal assemblage complexity is dependent on soil fertility and dominated by deterministic processes. *New Phytol.* 226, 232–243. doi:10.1111/nph.16345.
- Guo, X., Zhang, X., Qin, Y., Liu, Y. X., Zhang, J., Zhang, N., et al. (2020b). Host-Associated Quantitative Abundance Profiling Reveals the Microbial Load Variation of Root Microbiome. *Plant Commun.* 1, 1–17. doi:10.1016/j.xplc.2019.100003.
- Gusareva, E. S., Gaultier, N. P. E., Premkrishnan, B. N. V., Kee, C., Lim, S. B. Y., Heinle, C. E., et al. (2020). Taxonomic composition and seasonal dynamics of the air microbiome in West Siberia. *Sci. Rep.* 10, 1–7. doi:10.1038/s41598-020-78604-8.
- Gutiérrez, N., and Garrido, D. (2019). Species Deletions from Microbiome Consortia Reveal Key Metabolic Interactions between Gut Microbes. *mSystems* 4. doi:10.1128/msystems.00185-19.
- Hacquard, S. (2016). Disentangling the factors shaping microbiota composition across the plant holobiont. *New Phytol.* 209, 454–457. doi:10.1111/nph.13760.
- Hacquard, S., Garrido-Oter, R., González, A., Spaepen, S., Ackermann, G., Lebeis, S., et al. (2015). Microbiota and host nutrition across plant and animal kingdoms. *Cell Host Microbe* 17, 603–616. doi:10.1016/j.chom.2015.04.009.
- Haichar, F. E. Z., Marol, C., Berge, O., Rangel-Castro, J. I., Prosser, J. I., Balesdent, J., et al. (2008). Plant host habitat and root exudates shape soil bacterial community structure. *ISME J.* 2, 1221–1230. doi:10.1038/ismej.2008.80.
- Hamilton, J. J., Garcia, S. L., Brown, B. S., Oyserman, B. O., Moya-Flores, F., Bertilsson, S., et al. (2017). Metabolic Network Analysis and Metatranscriptomics Reveal Auxotrophies and Nutrient Sources of the Cosmopolitan Freshwater Microbial Lineage acI. *mSystems* 2, 1–13. doi:10.1128/msystems.00091-17.
- Hansen, B. L., Pessotti, R. de C., Fischer, M. S., Collins, A., El-Hifnawi, L., Liu, M. D., et al. (2020). Cooperation, competition, and specialized metabolism in a simplified root nodule microbiome. *MBio* 11, 1–20. doi:10.1128/mBio.01917-20.
- Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C., and Martiny, J. B. H. (2012). Beyond biogeographic patterns: Processes shaping the microbial landscape. *Nat. Rev. Microbiol.* 10, 497–506. doi:10.1038/nrmicro2795.

General bibliography

- Harodim, P. R., van Overbeek, L. S., Berg, G., Pirttilä, A. M., Compant, S., Campisano, A., et al. (2015). The Hidden World within Plants: Ecological and Evolutionary Considerations for Defining Functioning of Microbial Endophytes. *Microbiol. Mol. Biol. Rev.* 79, 293–320. doi:10.1128/MMBR.00050-14.
- Hartman, K., van der Heijden, M. G. A., Wittwer, R. A., Banerjee, S., Walser, J. C., and Schlaeppi, K. (2018). Cropping practices manipulate abundance patterns of root and soil microbiome members paving the way to smart farming. *Microbiome* 6, 1–14. doi:10.1186/s40168-017-0389-9.
- Hartmann, A., Schmid, M., van Tuinen, D., and Berg, G. (2009). Plant-driven selection of microbes. *Plant Soil* 321, 235–257. doi:10.1007/s11104-008-9814-y.
- Hassani, M. A., Durán, P., and Hacquard, S. (2018). Microbial interactions within the plant holobiont. *Encycl. Syst. Biol.*, 902–902. doi:10.1007/978-1-4419-9863-7_100630.
- Heather, J. M., and Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics* 107, 1–8. doi:10.1016/j.ygeno.2015.11.003.
- Heinken, A., and Thiele, I. (2015). Anoxic conditions promote species-specific mutualism between gut microbes In Silico. *Appl. Environ. Microbiol.* 81, 4049–4061. doi:10.1128/AEM.00101-15.
- Heintz-Buschart, A., May, P., Laczny, C. C., Lebrun, L. A., Bellora, C., Krishna, A., et al. (2016). Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. *Nat. Microbiol.* 2. doi:10.1038/nmicrobiol.2016.180.
- Henriques, S. F., Dhakan, D. B., Serra, L., Francisco, A. P., Carvalho-Santos, Z., Baltazar, C., et al. (2020). Metabolic cross-feeding in imbalanced diets allows gut microbes to improve reproduction and alter host behaviour. *Nat. Commun.* 11, 1–15. doi:10.1038/s41467-020-18049-9.
- Henry, C. S., Dejongh, M., Best, A. A., Frybarger, P. M., Lindsay, B., and Stevens, R. L. (2010). High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotechnol.* 28, 977–982. doi:10.1038/nbt.1672.
- Hester, E. R., Jetten, M. S. M., Welte, C. U., and Lücker, S. (2019). Metabolic overlap in environmentally diverse microbial communities. *Front. Genet.* 10, 1–10. doi:10.3389/fgene.2019.00989.
- Hibbing, M. E., Fuqua, C., Parsek, M. R., and Peterson, S. B. (2010). Bacterial competition: surviving and thriving in the microbial jungle. *Nat Rev Microbiol* 123, 75–88. doi:10.1353/nsm.2015.0002.
- Hillesland, K. L. (2018). Evolution on the bright side of life: microorganisms and the evolution of mutualism. *Ann. N. Y. Acad. Sci.* 1422, 88–103. doi:10.1111/nyas.13515.
- Hillesland, K. L., and Stahl, D. A. (2010). Rapid evolution of stability and productivity at the origin of a microbial mutualism. *Proc. Natl. Acad. Sci. U. S. A.* 107, 2124–2129. doi:10.1073/pnas.0908456107.
- Hirano, H., and Takemoto, K. (2019). Difficulty in inferring microbial community structure based on co-occurrence network approaches. *BMC Bioinformatics* 20. doi:10.1186/s12859-019-2915-1.
- Hoek, T. A., Axelrod, K., Biancalani, T., Yurtsev, E. A., Liu, J., and Gore, J. (2016). Resource Availability Modulates the Cooperative and Competitive Nature of a Microbial Cross-Feeding Mutualism. *PLoS Biol.* 14, 1–17. doi:10.1371/journal.pbio.1002540.
- Horner-Devine, M. C., and Bohannan, B. J. M. (2006). Phylogenetic clustering and overdispersion in bacterial communities. *Ecology* 87, 100–108. doi:10.1890/0012-9658(2006)87[100:pcaoib]2.0.co;2.
- Horner-Devine, M. C., Silver, J. M., Leibold, M. A., Bohannan, B. J. M., Colwell, R. K., Fuhrman, J. A., et al. (2007). A comparison of taxon co-occurrence patterns for macro- and microorganisms. *Ecology* 88, 1345–1353. doi:10.1890/06-0286.

General bibliography

- Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol. Biol. Evol.* 33, 1635–1638. doi:10.1093/molbev/msw046.
- Hussain, S. S., Mehnaz, S., and Siddique, K. H. M. (2018). “Harnessing the Plant Microbiome for Improved Abiotic Stress Tolerance,” in *Plant Microbiome: Stress Response*, eds. D. Egamberdieva and P. Ahmad (Singapore: Springer Singapore), 21–43. doi:10.1007/978-981-10-5514-0_2.
- Ikryannikova, L. N., Kurbatov, L. K., Gorokhovets, N. V., and Zamyatnin, A. A. (2020). Contact-dependent growth inhibition in bacteria: Do not get too close! *Int. J. Mol. Sci.* 21, 1–16. doi:10.3390/ijms21217990.
- Inda, M. E., Broset, E., Lu, T. K., and de la Fuente-Nunez, C. (2019). Emerging Frontiers in Microbiome Engineering. *Trends Immunol.* 40, 952–973. doi:10.1016/j.it.2019.08.007.
- Ishimoto, C. K., Aono, A. H., Nagai, J. S., Sousa, H., Miranda, A. R. L., Melo, V. M. M., et al. (2021). Microbial co-occurrence network and its key microorganisms in soil with permanent application of composted tannery sludge. *Sci. Total Environ.* 789. doi:10.1016/j.scitotenv.2021.147945.
- Jacoby, R. P., and Kopriva, S. (2019). Metabolic niches in the rhizosphere microbiome: New tools and approaches to analyse metabolic mechanisms of plant-microbe nutrient exchange. *J. Exp. Bot.* 70, 1087–1094. doi:10.1093/jxb/ery438.
- Jacoby, R. P., Koprivova, A., and Kopriva, S. (2021). Pinpointing secondary metabolites that shape the composition and function of the plant microbiome. *J. Exp. Bot.* 72, 57–69. doi:10.1093/jxb/eraa424.
- Jacoby, R. P., Martyn, A., and Kopriva, S. (2018). Exometabolomic profiling of bacterial strains as cultivated using arabidopsis root extract as the sole carbon source. *Mol. Plant-Microbe Interact.* 31, 803–813. doi:10.1094/MPMI-10-17-0253-R.
- Jaffe, A. L., Castelle, C. J., Matheus Carnevali, P. B., Gribaldo, S., and Banfield, J. F. (2020). The rise of diversity in metabolic platforms across the Candidate Phyla Radiation. *BMC Biol.* 18, 1–15. doi:10.1186/s12915-020-00804-5.
- Jansma, J., and Aidy, S. El (2020). Understanding the host-microbe interactions using metabolic modeling. *bioRxiv*, 1–14. doi:10.1101/2020.06.12.147918.
- Jha, Y. (2020). “Plant Microbiomes with Phytohormones: Attribute for Plant Growth and Adaptation Under the Stress Conditions,” in *Advances in Plant Microbiome and Sustainable Agriculture: Functional Annotation and Future Challenges*, eds. A. N. Yadav, A. A. Rastegari, N. Yadav, and D. Kour (Singapore: Springer Singapore), 85–103. doi:10.1007/978-981-15-3204-7_5.
- Jiang, T., Li, C., Teng, Y., Zhang, R., and Yan, Y. (2020). Recent advances in improving metabolic robustness of microbial cell factories. *Curr. Opin. Biotechnol.* 66, 69–77. doi:10.1016/j.copbio.2020.06.006.
- Johnson, D. R., Goldschmidt, F., Lilja, E. E., and Ackermann, M. (2012). Metabolic specialization and the assembly of microbial communities. *ISME J.* 6, 1985–1991. doi:10.1038/ismej.2012.46.
- Joyce, A. R., and Palsson, B. (2006). The model organism as a system: Integrating “omics” data sets. *Nat. Rev. Mol. Cell Biol.* 7, 198–210. doi:10.1038/nrm1857.
- Kang, S., Niu, J., Zhang, Q., Zhang, X., Han, G., and Zhao, M. (2020). Niche differentiation is the underlying mechanism maintaining the relationship between community diversity and stability under grazing pressure. *Glob. Ecol. Conserv.* 24, e01246. doi:10.1016/j.gecco.2020.e01246.
- Karimi, B., Maron, P. A., Chemidlin-Prevost Boure, N., Bernard, N., Gilbert, D., and Ranjard, L. (2017). Microbial diversity and ecological networks as indicators of environmental quality. *Environ. Chem. Lett.* 15, 265–281. doi:10.1007/s10311-017-0614-6.

General bibliography

- Karp, P. D., Latendresse, M., and Caspi, R. (2011). The Pathway Tools Pathway Prediction Algorithm. 424–429. doi:10.4056/sigs.1794338.
- Karp, P. D., Paley, S. M., Midford, P. E., Krummenacker, M., Billington, R., Kothari, A., et al. (2020). Pathway Tools version 24.0: Integrated Software for Pathway/Genome Informatics and Systems Biology. *arXiv*. doi:10.1109/bib/bbp043.
- Karp, P. D., Paley, S., and Romero, P. (2002). The Pathway Tools software Supplementary. *Bioinformatics* 18, 1–8.
- Karr, J. R., Sanghvi, J. C., MacKlin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B., et al. (2012). A whole-cell computational model predicts phenotype from genotype. *Cell* 150, 389–401. doi:10.1016/j.cell.2012.05.044.
- Kelsic, E. D., Zhao, J., Vetsigian, K., and Kishony, R. (2015). Counteraction of antibiotic production and degradation stabilizes microbial communities. *Nature* 521, 516–519. doi:10.1038/nature14485.
- Keymer, J. E., Galajda, P., Lambert, G., Liao, D., and Austin, R. H. (2008). Computation of mutual fitness by competing bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 105, 20269–20273. doi:10.1073/pnas.0810792105.
- Khandelwal, R. A., Olivier, B. G., Röling, W. F. M., Teusink, B., and Bruggeman, F. J. (2013). Community Flux Balance Analysis for Microbial Consortia at Balanced Growth. *PLoS One* 8. doi:10.1371/journal.pone.0064567.
- Khoshru, B., Moharramnejad, S., Gharajeh, N. H., Asgari Lajayer, B., and Ghorbanpour, M. (2020). “Plant Microbiome and Its Important in Stressful Agriculture,” in *Plant Microbiome Paradigm*, eds. A. Varma, S. Tripathi, and R. Prasad (Cham: Springer International Publishing), 13–48. doi:10.1007/978-3-030-50395-6_2.
- Kiers, E. T., Duhamel, M., Beesetty, Y., Mensah, J. A., Franken, O., Verbruggen, E., et al. (2011). Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science* (80-). 333, 880–882. doi:10.1126/science.1208473.
- Kim, T., Dreher, K., Nilo-Poyanco, R., Lee, I., Fiehn, O., Lange, B. M., et al. (2015). Patterns of metabolite changes identified from large-scale gene perturbations in arabidopsis using a genome-scale metabolic network. *Plant Physiol.* 167, 1685–1698. doi:10.1104/pp.114.252361.
- Kitano, H. (2002a). Computational Systems Biology. *Nature* 420. doi:10.1016/b978-0-12-809633-8.20287-2.
- Kitano, H. (2002b). Systems biology: A brief overview. *Science* (80-). 295, 1662–1664. doi:10.1126/science.1069492.
- Klitgord, N., and Segré, D. (2010). Environments that Induce Synthetic Microbial Ecosystems. *PloS CB* 6, 1–10. doi:10.1371/Citation.
- Knight, R., Navas, J., Quinn, R. A., Sanders, J. G., and Zhu, Q. (2018). Best practices for analysing microbiomes. *Nat. Rev. Microbiol.* 16, 410–422. doi:10.1038/s41579-018-0029-9.
- Koffel, T., Daufresne, T., and Klausmeier, C. A. (2021). From competition to facilitation and mutualism: a general theory of the niche. *Ecol. Monogr.* 91, 1–31. doi:10.1002/ecm.1458.
- Kong, W., Meldgin, D. R., Collins, J. J., and Lu, T. (2018). Designing microbial consortia with defined social interactions. *Nat. Chem. Biol.* 14, 821–829. doi:10.1038/s41589-018-0091-7.
- Konopka, A. (2009). What is microbial community ecology. *ISME J.* 3, 1223–1230. doi:10.1038/ismej.2009.88.
- Kumar, M., Ji, B., Babaei, P., Das, P., Lappa, D., Ramakrishnan, G., et al. (2018). Gut microbiota dysbiosis is associated with malnutrition and reduced plasma amino acid levels: Lessons from genome-scale metabolic modeling. *Metab. Eng.* 49, 128–142. doi:10.1016/j.ymben.2018.07.018.
- Kumar, M., Ji, B., Zengler, K., and Nielsen, J. (2019). Modelling approaches for studying the microbiome. *Nat. Microbiol.* 4, 1253–1267. doi:10.1038/s41564-019-0491-9.

General bibliography

- Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., and Bonneau, R. A. (2015). Sparse and Compositionally Robust Inference of Microbial Ecological Networks. *PLoS Comput. Biol.* 11, 1–25. doi:10.1371/journal.pcbi.1004226.
- Kuypers, M. M. M., Marchant, H. K., and Kartal, B. (2018). The microbial nitrogen-cycling network. *Nat. Rev. Microbiol.* 16, 263–276. doi:10.1038/nrmicro.2018.9.
- Lam, T. J., Stamboulian, M., Han, W., and Ye, Y. (2020). Model-based and phylogenetically adjusted quantification of metabolic interaction between microbial species. *PLoS Comput. Biol.* 16, 1–17. doi:10.1371/journal.pcbi.1007951.
- Laniau, J., Frioux, C., Nicolas, J., Baroukh, C., Cortes, M., Got, J., et al. (2017). Combining graph and flux-based structures to decipher phenotypic essential metabolites within metabolic networks. 1–29. doi:10.7717/peerj.3860.
- Larhlimi, A., Blachon, S., Selbig, J., and Nikoloski, Z. (2011). Robustness of metabolic networks: A review of existing definitions. *BioSystems* 106, 1–8. doi:10.1016/j.biosystems.2011.06.002.
- Larimer, A. L., Bever, J. D., and Clay, K. (2010). The interactive effects of plant microbial symbionts: A review and meta-analysis. *Symbiosis* 51, 139–148. doi:10.1007/s13199-010-0083-1.
- Lata, R. K., Divjot, K., and Nath, Y. A. (2019). Endophytic Microbiomes: Biodiversity, Ecological Significance and Biotechnological Applications. *Res. J. Biotechnol.* 14, 142–162.
- Lawson, C. E., Harcombe, W. R., Hatzenpichler, R., Lindemann, S. R., Löffler, F. E., O’Malley, M. A., et al. (2019). Common principles and best practices for engineering microbiomes. *Nat. Rev. Microbiol.* 17, 725–741. doi:10.1038/s41579-019-0255-9.
- Lawson, C. E., Wu, S., Bhattacharjee, A. S., Hamilton, J. J., McMahon, K. D., Goel, R., et al. (2017). Metabolic network analysis reveals microbial community interactions in anammox granules. *Nat. Commun.* 8, 1–12. doi:10.1038/ncomms15416.
- Layeghifard, M., Hwang, D. M., and Guttman, D. S. (2017). Disentangling Interactions in the Microbiome: A Network Perspective. *Trends Microbiol.* 25, 217–228. doi:10.1016/j.tim.2016.11.008.
- Lebeis, S. L., Paredes, S. H., Lundberg, D. S., Breakfield, N., Gehring, J., McDonald, M., et al. (2015). Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science (80-.).* 349, 860–864. doi:10.1126/science.aaa8764.
- Lennon, J. T., and Jones, S. E. (2011). Microbial seed banks: The ecological and evolutionary implications of dormancy. *Nat. Rev. Microbiol.* 9, 119–130. doi:10.1038/nrmicro2504.
- Levy, R., and Borenstein, E. (2013). Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proc. Natl. Acad. Sci. U. S. A.* 110, 12804–12809. doi:10.1073/pnas.1300926110.
- Lewin-epstein, O., and Hadany, L. (2020). Host – microbiome coevolution can promote cooperation in a rock – paper – scissors dynamics. *Proc. R. Soc. B* 287. doi:10.1098/rspb.2019.2754.
- Li, C., Lim, K. M. K., Chng, K. R., and Nagarajan, N. (2016). Predicting microbial interactions through computational approaches. *Methods* 102, 12–19. doi:10.1016/j.ymeth.2016.02.019.
- Li, T., Long, M., Li, H., Gatesoupe, F. J., Zhang, X., Zhang, Q., et al. (2017). Multi-omics analysis reveals a correlation between the host phylogeny, gut microbiota and metabolite profiles in cyprinid fishes. *Front. Microbiol.* 8, 1–11. doi:10.3389/fmicb.2017.00454.

General bibliography

- Li, Y., Shipley, B., Price, J. N., Dantas, V. de L., Tamme, R., Westoby, M., et al. (2018). Habitat filtering determines the functional niche occupancy of plant communities worldwide. *J. Ecol.* 106, 1001–1009. doi:10.1111/1365-2745.12802.
- Liao, C., Maslov, S., Wang, T., and Xavier, J. B. (2020). Modeling microbial cross-feeding at intermediate scale portrays community dynamics and species coexistence. *PLoS Comput. Biol.* 16, 1–23. doi:10.1371/journal.pcbi.1008135.
- Lim, Y. Y., Lee, Y. S., and Ooi, D. S. Q. (2020). Engineering the Gut Microbiome for Treatment of Obesity: A Review of Current Understanding and Progress. *Biotechnol. J.* 15. doi:10.1002/biot.202000013.
- Lin, Y., Cornell, W. C., Jo, J., Price-whelan, A., and Dietrich, L. E. P. (2018). The *Pseudomonas aeruginosa* Complement of Lactate Dehydrogenases Enables Use of D- and L-Lactate and Metabolic Cross-Feeding. *MBio* 9, 1–12.
- Liu, Y. F., Galzerani, D. D., Mbadinga, S. M., Zaramela, L. S., Gu, J. D., Mu, B. Z., et al. (2018). Metabolic capability and in situ activity of microorganisms in an oil reservoir. *Microbiome* 6, 1–12. doi:10.1186/s40168-017-0392-1.
- Liu, Y., and Nielsen, J. (2019). Recent trends in metabolic engineering of microbial chemical factories. *Curr. Opin. Biotechnol.* 60, 188–197. doi:10.1016/j.copbio.2019.05.010.
- Liu, Z., Huang, H., Qi, M., Wang, X., and Adebanjo, O. O. (2019). Metabolite Cross-Feeding between *Rhodococcus ruber* YYL and *Bacillus cereus* MLY1 in the Biodegradation of Tetrahydrofuran under pH Stress. *Appl. Environ. Microbiol.* 85, 1–12.
- Lloyd-Price, J., Arze, C., Ananthakrishnan, A. N., Schirmer, M., Avila-Pacheco, J., Poon, T. W., et al. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* 569, 655–662. doi:10.1038/s41586-019-1237-9.
- Logan, S. L., Thomas, J., Yan, J., Baker, R. P., Shields, D. S., Xavier, J. B., et al. (2018). The *Vibrio cholerae* type VI secretion system can modulate host intestinal mechanics to displace gut bacterial symbionts. *Proc. Natl. Acad. Sci. U. S. A.* 115, E3779–E3787. doi:10.1073/pnas.1720133115.
- Louca, S., Jacques, S. M. S., Pires, A. P. F., Leal, J. S., Srivastava, D. S., Parfrey, L. W., et al. (2016). High taxonomic variability despite stable functional structure across microbial communities. *Nat. Ecol. Evol.* 1, 1–12. doi:10.1038/s41559-016-0015.
- Loudon, C. M., Matthews, B., Sevilgen, D. S., and Ibelings, B. W. (2016). Experimental evidence that evolution by niche construction affects dissipative ecosystem dynamics. *Evol. Ecol.* 30, 221–234. doi:10.1007/s10682-015-9802-7.
- Lovley, D. R. (2017). Happy together: Microbial communities that hook up to swap electrons. *ISME J.* 11, 327–336. doi:10.1038/ismej.2016.136.
- Lubbe, A., Bowen, B. P., and Northen, T. (2017). Exometabolomic analysis of cross-feeding metabolites. *Metabolites* 7. doi:10.3390/metabo7040050.
- Ma, B., Wang, Y., Ye, S., Liu, S., Stirling, E., Gilbert, J. A., et al. (2020). Earth microbial co-occurrence network reveals interconnection pattern across microbiomes. *Microbiome* 8, 1–12. doi:10.1186/s40168-020-00857-2.
- Machado, D., Maistrenko, O. M., Andrejev, S., Kim, Y., Bork, P., Patil, K. R., et al. (2021). Polarization of microbial communities between competitive and cooperative metabolism. *Nat. Ecol. Evol.* 5, 195–203. doi:10.1038/s41559-020-01353-4.
- Madsen, E. L. (2011). Microorganisms and their roles in fundamental biogeochemical cycles. *Curr. Opin. Biotechnol.* 22, 456–464. doi:10.1016/j.copbio.2011.01.008.

General bibliography

- Magnúsdóttir, S., Heinken, A., Kutt, L., Ravcheev, D. A., Bauer, E., Noronha, A., et al. (2017). Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* 35, 81–89. doi:10.1038/nbt.3703.
- Mahowald, M. A., Rey, F. E., Seedorf, H., Turnbaugh, P. J., Fulton, R. S., Wollam, A., et al. (2009). Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. *Proc. Natl. Acad. Sci. U. S. A.* 106, 5859–5864. doi:10.1073/pnas.0901529106.
- Maignien, L., DeForce, E. A., Chafee, M. E., Murat Eren, A., and Simmons, S. L. (2014). Ecological succession and stochastic variation in the assembly of *Arabidopsis thaliana* phyllosphere communities. *MBio* 5. doi:10.1128/mBio.00682-13.
- Malacrino, A., Wang, M., Caul, S., Karley, A. J., and Bennett, A. E. (2021). Herbivory shapes the rhizosphere bacterial microbiota in potato plants. *Environ. Microbiol. Rep.* 00. doi:10.1111/1758-2229.12998.
- Marchesi, J. R., and Ravel, J. (2015). The vocabulary of microbiome research: a proposal. *Microbiome* 3, 1–3. doi:10.1186/s40168-015-0094-5.
- Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., et al. (2006). Microbial biogeography: Putting microorganisms on the map. *Nat. Rev. Microbiol.* 4, 102–112. doi:10.1038/nrmicro1341.
- Martiny, J. B. H., Jones, S. E., Lennon, J. T., and Martiny, A. C. (2015). Microbiomes in light of traits: A phylogenetic perspective. *Science (80-.).* 350. doi:10.1126/science.aac9323.
- Marx, J. (2004). The Roots of Plant-Microbe Collaborations. *Science (80-.).* 304.
- Mas, A., Jamshidi, S., Lagadeuc, Y., Eveillard, D., and Vandenkoornhuyse, P. (2016). Beyond the Black Queen Hypothesis. *ISME J.* 10, 2085–2091. doi:10.1038/ismej.2016.22.
- Mataigne, V., Nathan., V., Philippe, V., and Stéphane., H. Microbial System Ecology to understand metabolic dependencies in microbiomes : A mini-review. *Submitt. to Front. Microbiol.*
- Mataigne, V., Nathan., V., Philippe, V., and Stéphane., H. Multi-genomes metabolic modelling predicts functional inter-dependencies in the *Arabidopsis* root microbiome. *prep. (planned to be Submitt. to Microbiome).*
- Mauger, S., Monard, C., Thion, C., and Vandenkoornhuyse, P. (2021). Ecology & Evolution Contribution of single-cell omics to microbial ecology. *Trends Ecol. Evol.*, 1–12. doi:10.1016/j.tree.2021.09.002.
- Maynard, D. S., Miller, Z. R., and Allesina, S. (2020). Predicting coexistence in experimental ecological communities. *Nat. Ecol. Evol.* 4, 91–100. doi:10.1038/s41559-019-1059-z.
- McCully, A. L., Lasarre, B., and McKinlay, J. B. (2017a). Recipient-Biased Competition for an Intracellularly Generated Cross-Fed Nutrient Is Required for Coexistence of Microbial Mutualists. *MBio* 8, 1–13.
- McCully, A. L., LaSarre, B., and McKinlay, J. B. (2017b). Growth-independent cross-feeding modifies boundaries for coexistence in a bacterial mutualism. *Environ. Microbiol.* 19, 3538–3550. doi:10.1111/1462-2920.13847.
- McGlynn, S. E., Chadwick, G. L., Kempes, C. P., and Orphan, V. J. (2015). Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature* 526, 531–535. doi:10.1038/nature15512.
- McLaren, M. R., and Callahan, B. J. (2020). Pathogen resistance may be the principal evolutionary advantage provided by the microbiome: Pathogen resistance by the microbiome. *Philos. Trans. R. Soc. B Biol. Sci.* 375. doi:10.1098/rstb.2019.0592rstb20190592.
- McNally, L., and Brown, S. P. (2016). Microbiome: Ecology of stable gut communities. *Nat. Microbiol.* 1, 2015–2016. doi:10.1038/nm microbiol.2015.16.

General bibliography

- Medlock, G. L., Carey, M. A., McDuffie, D. G., Mundy, M. B., Giallourou, N., Swann, J. R., et al. (2018). Inferring Metabolic Mechanisms of Interaction within a Defined Gut Microbiota. *Cell Syst.* 7, 245–257.e7. doi:10.1016/j.cels.2018.08.003.
- Mee, M. T., Collins, J. J., Church, G. M., and Wang, H. H. (2014). Syntrophic exchange in synthetic microbial communities. 2149–2156. doi:10.1073/pnas.1405641111.
- Meijer, J., van Dijk, B., and Hogeweg, P. (2020). Contingent evolution of alternative metabolic network topologies determines whether cross-feeding evolves. *Commun. Biol.* 3, 1–13. doi:10.1038/s42003-020-1107-x.
- Mello, B. L., Alessi, A. M., McQueen-Mason, S., Bruce, N. C., and Polikarpov, I. (2016). Nutrient availability shapes the microbial community structure in sugarcane bagasse compost-derived consortia. *Sci. Rep.* 6, 1–8. doi:10.1038/srep38781.
- Mendoza, S. N., Olivier, B. G., Molenaar, D., and Teusink, B. (2019). A Systematic Assessment Of Current Genome-Scale Metabolic Reconstruction Tools. *Genome Biol.* 20, 1–20. doi:10.1101/558411.
- Meron, D., Rodolfo-Metalpa, R., Cunning, R., Baker, A. C., Fine, M., and Banin, E. (2012). Changes in coral microbial communities in response to a natural pH gradient. *ISME J.* 6, 1775–1785. doi:10.1038/ismej.2012.19.
- Michalska-Smith, M., Song, Z., Spawn-Lee, S. A., Hansen, Z. A., Johnson, M., May, G., et al. (2021). Network structure of resource use and niche overlap within the endophytic microbiome. *ISME J.*, 1–12. doi:10.1038/s41396-021-01080-z.
- Mishler, B. D., and Brandon, R. N. (1987). Individuality, pluralism, and the phylogenetic species concept. *Biol. Philos.* 2, 397–414. doi:10.1007/BF00127698.
- Miteva, V. I., Sheridan, P. P., and Brenchley, J. E. (2004). Phylogenetic and Physiological Diversity of Microorganisms Isolated from a Deep Greenland Glacier Ice Core. *Appl. Environ. Microbiol.* 70, 202–213. doi:10.1128/AEM.70.1.202-213.2004.
- Momeni, B., Xie, L., and Shou, W. (2017). Lotka-Volterra pairwise modeling fails to capture diverse pairwise microbial interactions. *Elife* 6, 1–34. doi:10.7554/elife.25051.
- Monchgesang, S., Strehmel, N., Schmidt, S., Westphal, L., Taruttis, F., Muller, E., et al. (2016). Natural variation of root exudates in *Arabidopsis thaliana*-linking metabolomic and genomic data. *Sci. Rep.* 6, 1–11. doi:10.1038/srep29033.
- Mondav, R., Mccalley, C. K., Hodgkins, S. B., Frolking, S., Saleska, S. R., Rich, V. I., et al. (2017). Microbial network, phylogenetic diversity and community membership in the active layer across a permafrost thaw gradient. *Env. Microbiol.* 19, 3201–3218. doi:10.1111/1462-2920.13809.
- Monk, J., Nogales, J., and Palsson, B. O. (2014). Optimizing genome-scale network reconstructions. *Nat. Biotechnol.* 32, 447–452. doi:10.1038/nbt.2870.
- Moran, M. A. (2015). The global ocean microbiome. *Science (80-).* 350. doi:10.1126/science.aac8455.
- Morris, B. E. L., Henneberger, R., Huber, H., and Moissl-Eichinger, C. (2013). Microbial syntrophy: Interaction for the common good. *FEMS Microbiol. Rev.* 37, 384–406. doi:10.1111/1574-6976.12019.
- Morris, J. J. (2015). Black Queen evolution: The role of leakiness in structuring microbial communities. *Trends Genet.* 31, 475–482. doi:10.1016/j.tig.2015.05.004.
- Morris, J. J., Lenski, R. E., and Zinser, E. R. (2012a). The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss. 3, 1–9. doi:10.1128/mBio.00036-12. Copyright.
- Morris, J. J., Lenski, R. E., and Zinser, E. R. (2012b). The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss. 3, 1–9. doi:10.1128/mBio.00036-12. Copyright.

General bibliography

- Morrison-Whittle, P., and Goddard, M. R. (2015). Quantifying the relative roles of selective and neutral processes in defining eukaryotic microbial communities. *ISME J.* 9, 2003–2011. doi:10.1038/ismej.2015.18.
- Muller, E. E. L., Faust, K., Widder, S., Herold, M., Martínez Arbas, S., and Wilmes, P. (2018). Using metabolic networks to resolve ecological properties of microbiomes. *Curr. Opin. Syst. Biol.* 8, 73–80. doi:10.1016/j.coisb.2017.12.004.
- Musilova, L., Ridl, J., Polivkova, M., Macek, T., and Uhlik, O. (2016). Effects of secondary plant metabolites on microbial populations: Changes in community structure and metabolic activity in contaminated environments. *Int. J. Mol. Sci.* 17. doi:10.3390/ijms17081205.
- Nelson, E. B. (2018). The seed microbiome: Origins, interactions, and impacts. *Plant Soil* 422, 7–34. doi:10.1007/s11104-017-3289-7.
- Nemergut, D. R., Costello, E. K., Hamady, M., Lozupone, C., Jiang, L., Schmidt, S. K., et al. (2011). Global patterns in the biogeography of bacterial taxa. *Environ. Microbiol.* 13, 135–144. doi:10.1111/j.1462-2920.2010.02315.x.
- Nemergut, D. R., Schmidt, S. K., Fukami, T., O'Neill, S. P., Bilinski, T. M., Stanish, L. F., et al. (2013). Patterns and Processes of Microbial Community Assembly. *Microbiol. Mol. Biol. Rev.* 77, 342–356. doi:10.1128/mmbr.00051-12.
- Nielsen, J. (2017). Systems biology of metabolism. *Annu. Rev. Biochem.* 86, 245–275. doi:10.1146/annurev-biochem-061516-044757.
- Ning, D., Yuan, M., Wu, L., Zhang, Y., Guo, X., Zhou, X., et al. (2020). A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. *Nat. Commun.* 11. doi:10.1038/s41467-020-18560-z.
- Niu, B., Paulson, N. J., Zheng, X., and Kolter, R. (2017). Simplified and representative bacterial community of maize roots. 24–26. doi:10.1073/pnas.1616148114.
- Noecker, C., McNally, C. P., Eng, A., and Borenstein, E. (2016). High-resolution characterization of the human microbiome. *Transl. Res.* 179, 7–23. doi:10.1016/j.trsl.2016.07.012.
- Oakley, B. B., Carbonero, F., Van Der Gast, C. J., Hawkins, R. J., and Purdy, K. J. (2010). Evolutionary divergence and biogeography of sympatric niche-differentiated bacterial populations. *ISME J.* 4, 488–497. doi:10.1038/ismej.2009.146.
- Oberhardt, M. A., Zarecki, R., Gronow, S., Lang, E., Klenk, H. P., Gophna, U., et al. (2015). Harnessing the landscape of microbial culture media to predict new organism-media pairings. *Nat. Commun.* 6. doi:10.1038/ncomms9493.
- Ochman, Howard, Lawrence, Jeffrey G., and Groisman, Eduardo A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299–304. Available at: <http://adsabs.harvard.edu/abs/2000Natur.405..299O>.
- Ogbonna, J. C., Yoshizawa, H., and Tanaka, H. (2000). Treatment of high strength organic wastewater by a mixed culture of photosynthetic microorganisms. *J. Appl. Phycol.* 12, 277–284. doi:10.1023/a:1008188311681.
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2020). vegan: Community Ecology Package. Available at: <https://cran.r-project.org/package=vegan>.
- Olesen, J. M., Bascompte, J., Dupont, Y. L., and Jordano, P. (2007). The modularity of pollination networks. *Proc. Natl. Acad. Sci.* 104.
- Oliveira, N. M., Niehus, R., and Foster, K. R. (2014). Evolutionary limits to cooperation in microbial communities. *Proc. Natl. Acad. Sci. U. S. A.* 111, 17941–17946. doi:10.1073/pnas.1412673111.
- Opal, S. M. (2009). “A Brief History of Microbiology and Immunology,” in *Vaccines: A Biography*, ed. A. W. Artenstein, 31–56. doi:10.1007/978-1-4419-1108-7_3.

General bibliography

- Orozco-Mosqueda, M. del C., Rocha-Granados, M. del C., Glick, B. R., and Santoyo, G. (2018). Microbiome engineering to improve biocontrol and plant growth-promoting mechanisms. *Microbiol. Res.* 208, 25–31. doi:10.1016/j.micres.2018.01.005.
- Orsi, W. D., Richards, T. A., and Francis, W. R. (2017). Predicted microbial secretomes and their target substrates in marine sediment. *Nat. Microbiol.* 3, 32–37. doi:10.1038/s41564-017-0047-9.
- Orth, J. D., Thiele, I., and Palsson, B. O. (2010). What is flux balance analysis? *Nat. Biotechnol.* 28, 245–248. doi:10.1038/nbt.1614.
- Pacheco, A. R., Moel, M., and Segrè, D. (2019). Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems. *Nat. Commun.* 10. doi:10.1038/s41467-018-07946-9.
- Pacheco, A. R., and Segrè, D. (2019). A multidimensional perspective on microbial interactions. *FEMS Microbiol. Lett.* 366, 1–11. doi:10.1093/femsle/fnz125.
- Pande, S., and Kost, C. (2017). Bacterial Unculturability and the Formation of Intercellular Metabolic Networks. *Trends Microbiol.* 25, 349–361. doi:10.1016/j.tim.2017.02.015.
- Pande, S., Merker, H., Bohl, K., Reichelt, M., Schuster, S., De Figueiredo, L. F., et al. (2014). Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. *ISME J.* 8, 953–962. doi:10.1038/ismej.2013.211.
- Pickard, J. M., Zeng, M. Y., Caruso, R., and Núñez, G. (2017). Gut microbiota: Role in pathogen colonization, immune responses, and inflammatory disease. *Immunol. Rev.* 279, 70–89. doi:10.1111/imr.12567.
- Pini, F., Galardini, M., Bazzicalupo, M., and Mengoni, A. (2011). Plant-bacteria association and symbiosis: Are there common genomic traits in alphaproteobacteria? *Genes (Basel).* 2, 1017–1032. doi:10.3390/genes2041017.
- Pinu, F. R., Granucci, N., Daniell, J., Han, T. L., Carneiro, S., Rocha, I., et al. (2018). Metabolite secretion in microorganisms: the theory of metabolic overflow put to the test. *Metabolomics* 14, 1–16. doi:10.1007/s11306-018-1339-7.
- Pocheville, A. (2015). “The Ecological Niche: History and Recent Controversies,” in *Handbook of Evolutionary Thinking in the Sciences*, eds. T. Heams, P. Huneman, G. Lecointre, and M. Silberstein (Dordrecht: Springer Netherlands), 1–910. doi:10.1007/978-94-017-9014-7_26.
- Porter, S. S., Chang, P. L., Conow, C. A., Dunham, J. P., and Friesen, M. L. (2017). Association mapping reveals novel serpentine adaptation gene clusters in a population of symbiotic Mesorhizobium. *ISME J.* 11, 248–262. doi:10.1038/ismej.2016.88.
- Powlson, D. S., Hirsch, P. R., and Brookes, P. C. (2001). The role of soil microorganisms in soil organic matter conservation in the tropics. *Nutr. Cycl. Agroecosystems* 61, 41–51. doi:10.1023/A:1013338028454.
- Price, G. W., Langille, M. G. I., and Yurgel, S. N. (2021). Microbial co-occurrence network analysis of soils receiving short- and long-term applications of alkaline treated biosolids. *Sci. Total Environ.* 751, 141687. doi:10.1016/j.scitotenv.2020.141687.
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One* 5. doi:10.1371/journal.pone.0009490.
- Prigent, S., Dittami, S. M., Thiele, S., Tonon, T., Eveillard, D., and Siegel, A. (2017). Meneco , a Topology-Based Gap-Filling Tool Applicable to Degraded Genome-Wide Metabolic Networks. *PLoS Comput. Biol.*, 1–32. doi:10.1371/journal.pcbi.1005276.
- Prosser, J. I., Bohannan, B. J. M., Curtis, T. P., Ellis, R. J., Firestone, M. K., Freckleton, R. P., et al. (2007). Prosser et al 2007 _The role of ecological theory in microbial ecology. *Nat. Rev. Microbiol.* 5, 384–392.

General bibliography

- Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., and Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* 35, 1211. doi:10.1038/nbt1217-1211b.
- R. Poudel, A. G. (2016). Microbiome network: A systems framework for identify candidate Microbial Assemblages for Disease Management. 106, 1083–1096. Available at: <http://apsjournals.apsnet.org/doi/pdf/10.1094/PHYTO-02-16-0058-FI>.
- Raina, J. B., Eme, L., Pollock, F. J., Spang, A., Archibald, J. M., and Williams, T. A. (2018). Symbiosis in the microbial world: from ecology to genome evolution. *Biol. Open* 7, 5–8. doi:10.1242/bio.032524.
- Rakoff-Nahoum, S., Foster, K. R., and Comstock, L. E. (2016). The evolution of cooperation within the gut microbiota. *Nature* 533, 255–259. doi:10.1038/nature17626.
- Ramette, A. (2007). Multivariate analyses in microbial ecology. *FEMS Microbiol Ecol* 62, 142–160. doi:10.1111/j.1574-6941.2007.00375.x.
- Ramirez, K. S., Geisen, S., Morriën, E., Snoek, B. L., and van der Putten, W. H. (2018). Network Analyses Can Advance Above-Belowground Ecology. *Trends Plant Sci.* 23, 759–768. doi:10.1016/j.tplants.2018.06.009.
- Ray, P., Lakshmanan, V., Labbé, J. L., and Craven, K. D. (2020). Microbe to Microbiome: A Paradigm Shift in the Application of Microorganisms for Sustainable Agriculture. *Front. Microbiol.* 11, 1–15. doi:10.3389/fmicb.2020.622926.
- Read, T., Fortun-Lamothe, L., Pascal, G., Boulch, M. Le, Cauquil, L., Gabinaud, B., et al. (2019). Diversity and Co-occurrence pattern analysis of cecal microbiota establishment at the onset of solid feeding in young rabbits. *Front. Microbiol.* 10, 1–13. doi:10.3389/fmicb.2019.00973.
- Richards, T. A., Massana, R., Pagliara, S., and Hall, N. (2019). Single cell ecology. *Philos. Trans. R. Soc. B Biol. Sci.* 374. doi:10.1098/rstb.2019.0076.
- Richelle, A., David, B., Demaegd, D., Dewerchin, M., Kinet, R., Morreale, A., et al. (2020). Towards a widespread adoption of metabolic modeling tools in biopharmaceutical industry: a process systems biology engineering perspective. *npj Syst. Biol. Appl.* 6. doi:10.1038/s41540-020-0127-y.
- Riera, J. L., and Baldo, L. (2020). Microbial co-occurrence networks of gut microbiota reveal community conservation and diet-associated shifts in cichlid fishes. *Anim. Microbiome* 2. doi:10.1186/s42523-020-00054-4.
- Rodionova, I. A., Li, X., Plymale, A. E., Motamedchaboki, K., Konopka, A. E., Romine, M. F., et al. (2015). Genomic distribution of B-vitamin auxotrophy and uptake transporters in environmental bacteria from the Chloroflexi phylum. *Environ. Microbiol. Rep.* 7, 204–210. doi:10.1111/1758-2229.12227.
- Rodríguez, P. A., Rothballer, M., Chowdhury, S. P., Nussbaumer, T., Gutjahr, C., and Falter-Braun, P. (2019). Systems Biology of Plant-Microbiome Interactions. *Mol. Plant* 12, 804–821. doi:10.1016/j.molp.2019.05.006.
- Rogel, A. M., Hernandez-Lucas, I., Kuykendall, D. L., Balkwill, D. L., and Martinez-romero, E. (2001). Nitrogen-Fixing Nodules with *Ensifer adhaerens* Harboring *Rhizobium tropici* Symbiotic Plasmids. *Appl. Environ. Microbiol.* 67, 3264–3268. doi:10.1128/AEM.67.7.3264.
- Röling, W. F. M., and Van Bodegom, P. M. (2014). Toward quantitative understanding on microbial community structure and functioning: A modeling-centered approach using degradation of marine oil spills as example. *Front. Microbiol.* 5, 1–12. doi:10.3389/fmicb.2014.00125.
- Romdhane, S., Spor, A., Aubert, J., Bru, D., Breuil, M. C., Hallin, S., et al. (2021). Unraveling negative biotic interactions determining soil microbial community assembly and functioning. *ISME J.*, 1–11. doi:10.1038/s41396-021-01076-9.

General bibliography

- Rosselló-Mora, R., and Amann, R. (2001). The species concept for prokaryotes. *FEMS Microbiol. Rev.* 25, 39–67. doi:10.1016/S0168-6445(00)00040-1.
- Röttjers, L., and Faust, K. (2019). Can we predict kestones? *Nat. Rev. Microbiol.* 17, 193. doi:10.1038/s41579-018-0132-y.
- Russel, J., Røder, H. L., Madsen, J. S., Burmølle, M., and Sørensen, S. J. (2017). Antagonism correlates with metabolic similarity in diverse bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 114, 10684–10688. doi:10.1073/pnas.1706016114.
- San Roman, M., and Wagner, A. (2018). An enormous potential for niche construction through bacterial cross-feeding in a homogeneous environment. *PLoS Comput. Biol.* 14, 1–29. doi:10.1371/journal.pcbi.1006340.
- Satola, B., Wübbeler, J. H., and Steinbüchel, A. (2013). Metabolic characteristics of the species *Variovorax paradoxus*. *Appl. Microbiol. Biotechnol.* 97, 541–560. doi:10.1007/s00253-012-4585-z.
- Schlatter, D. C., Yin, C., Burke, I., Hulbert, S., and Paulitz, T. (2018). Location, Root Proximity, and Glyphosate-Use History Modulate the Effects of Glyphosate on Fungal Community Networks of Wheat. *Microb. Ecol.* 76, 240–257. doi:10.1007/s00248-017-1113-9.
- Schöps, R., Goldmann, K., Herz, K., Lentendu, G., Schöning, I., Bruelheide, H., et al. (2018). Land-use intensity rather than plant functional identity shapes bacterial and fungal rhizosphere communities. *Front. Microbiol.* 9. doi:10.3389/fmicb.2018.02711.
- Schulz, F., Eloe-Fadrosh, E. A., Bowers, R. M., Jarett, J., Nielsen, T., Ivanova, N. N., et al. (2017). Towards a balanced view of the bacterial tree of life. *Microbiome* 5, 140. doi:10.1186/s40168-017-0360-9.
- Seif, Y., Choudhary, K. S., Hefner, Y., Anand, A., Yang, L., and Palsson, B. O. (2020). Metabolic and genetic basis for auxotrophies in Gram-negative species. *Proc. Natl. Acad. Sci. U. S. A.* 117, 6264–6273. doi:10.1073/pnas.1910499117.
- Seth, E. C., and Taga, M. E. (2014). Nutrient cross-feeding in the microbial world. *Front. Microbiol.* 5, 1–6. doi:10.3389/fmicb.2014.00350.
- Shah, H. A., Liu, J., Yang, Z., and Feng, J. (2021). Review of Machine Learning Methods for the Prediction and Reconstruction of Metabolic Pathways. *Front. Mol. Biosci.* 8, 1–11. doi:10.3389/fmolb.2021.634141.
- Shahzad, K., and Loor, J. J. (2012). Application of Top-Down and Bottom-up Systems Approaches in Ruminant Physiology and Metabolism. *Curr. Genomics* 13, 379–394. doi:10.2174/138920212801619269.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7. doi:10.1038/msb.2011.75.
- Simonet, C., and McNally, L. (2021). Kin selection explains the evolution of cooperation in the gut microbiota. *Proc. Natl. Acad. Sci. U. S. A.* 118, 1–10. doi:10.1073/pnas.2016046118.
- Sivadon, P., Barnier, C., Urios, L., and Grimaud, R. (2019). Biofilm formation as a microbial strategy to assimilate particulate substrates. *Environ. Microbiol. Rep.* 11, 749–764. doi:10.1111/1758-2229.12785.
- Smith, N. W., Shorten, P. R., Altermann, E., Roy, N. C., and McNabb, W. C. (2019). The Classification and Evolution of Bacterial Cross-Feeding. *Front. Ecol. Evol.* 7, 1–15. doi:10.3389/fevo.2019.00153.
- Smith, S. E., and Read, D. (2008). “INTRODUCTION,” in *Mycorrhizal Symbiosis (Third Edition)*, eds. S. E. Smith and D. Read (London: Academic Press), 1–9. doi:<https://doi.org/10.1016/B978-012370526-6.50002-7>.
- Snoep, J. L., and Westerhoff, H. V. (2005). From isolation to integration, a systems biology approach for building the Silicon Cell. *Syst. Biol. (Stevenage)*. 13, 13–30. doi:10.1007/b106456.

General bibliography

- Soininen, J., McDonald, R., and Hillebrand, H. (2007). The distance decay of similarity in ecological communities. *Ecography (Cop.)*. 30, 3–12. doi:10.1111/j.2006.0906-7590.04817.x.
- Song, S. J., Sanders, J. G., Delsuc, F., Metcalf, J., Amato, K., Taylor, M. W., et al. (2020). Comparative analyses of vertebrate gut microbiomes reveal convergence between birds and bats. *MBio* 11. doi:10.1128/mBio.02901-19.
- Sriswasdi, S., Yang, C. C., and Iwasaki, W. (2017). Generalist species drive microbial dispersion and evolution. *Nat. Commun.* 8. doi:10.1038/s41467-017-01265-1.
- Stacy, A., Everett, J., Jorth, P., Trivedi, U., Rumbaugh, K. P., and Whiteley, M. (2014). Bacterial fight-and-flight responses enhance virulence in a polymicrobial infection. *Proc. Natl. Acad. Sci. U. S. A.* 111, 7819–7824. doi:10.1073/pnas.1400586111.
- Stams, A. J. M., and Plugge, C. M. (2009). Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nat. Rev. Microbiol.* 7, 568–577. doi:10.1038/nrmicro2166.
- Stegen, J. C., Lin, X., Fredrickson, J. K., Chen, X., Kennedy, D. W., Murray, C. J., et al. (2013). Quantifying community assembly processes and identifying features that impose them. *ISME J.* 7, 2069–2079. doi:10.1038/ismej.2013.93.
- Stegen, J. C., Lin, X., Konopka, A. E., and Fredrickson, J. K. (2012). Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J.* 6, 1653–1664. doi:10.1038/ismej.2012.22.
- Stewart, P. S. (2003). Diffusion in Biofilm. *J. Bacteriol.* 185, 1485–1491. doi:10.1128/JB.185.5.1485.
- Stringlis, I. A., Yu, K., Feussner, K., De Jonge, R., Van Bentum, S., Van Verk, M. C., et al. (2018). MYB72-dependent coumarin exudation shapes root microbiome assembly to promote plant health. *Proc. Natl. Acad. Sci. U. S. A.* 115, E5213–E5222. doi:10.1073/pnas.1722335115.
- Sun, Z., Koffel, T., Stump, S. M., Grimaud, G. M., and Klausmeier, C. A. (2019). Microbial cross-feeding promotes multiple stable states and species coexistence, but also susceptibility to cheaters. *J. Theor. Biol.* 465, 63–77. doi:10.1016/j.jtbi.2019.01.009.
- Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., et al. (2015). Ocean plankton. Structure and function of the global ocean microbiome. *Science (80-)*. 348, 1–10. doi:10.1126/science.1261359.
- Sung, J., Kim, S., Cabatbat, J. J. T., Jang, S., Jin, Y. S., Jung, G. Y., et al. (2017). Global metabolic interaction network of the human gut microbiota for context-specific community-scale analysis. *Nat. Commun.* 8, 1–12. doi:10.1038/ncomms15393.
- Tang, J., and Zhou, S. (2011). The importance of niche differentiation for coexistence on large scales. *J. Theor. Biol.* 273, 32–36. doi:10.1016/j.jtbi.2010.12.025.
- Testa, S., Berger, S., Piccardi, P., Oechslin, F., Resch, G., and Mitri, S. (2019). Spatial structure affects phage efficacy in infecting dual-strain biofilms of *Pseudomonas aeruginosa*. *Commun. Biol.* 2, 1–12. doi:10.1038/s42003-019-0633-x.
- Theis, K. R., Dheilly, N. M., Klassen, J. L., Brucker, R. M., Baines, J. F., Bosch, T. C. G., et al. (2016). Getting the Hologenome Concept Right: an Eco-Evolutionary Framework for Hosts and Their Microbiomes. *mSystems* 1, 1–6. doi:10.1128/msystems.00028-16.
- Thiele, I., and Palsson, B. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 5, 93–121. doi:10.1038/nprot.2009.203.
- Thommes, M., Wang, T., Zhao, Q., Paschalidis, I. C., and Segré, D. (2019). Designing Metabolic Division of Labor in Microbial Communities. *mSystems* 4, 1–21. doi:10.1128/msystems.00263-18.

General bibliography

- Toju, H., Peay, K. G., Yamamichi, M., Narisawa, K., and Hiruma, K. (2018). Core microbiomes for sustainable agroecosystems. *Nat. Plants* 4. doi:10.1038/s41477-018-0139-4.
- Tonon, T., Eveillard, D., Prigent, S., Bourdon, J., Potin, P., Boyen, C., et al. (2011). Toward systems biology in brown algae to explore acclimation and adaptation to the shore environment. *Omi. A J. Integr. Biol.* 15, 883–892. doi:10.1089/omi.2011.0089.
- Trivedi, P., Leach, J. E., Tringe, S. G., Sa, T., and Singh, B. K. (2020). Plant–microbiome interactions: from community assembly to plant health. *Nat. Rev. Microbiol.* 18, 607–621. doi:10.1038/s41579-020-0412-1.
- Tropini, C., Earle, K. A., Huang, K. C., and Sonnenburg, J. L. (2017). The Gut Microbiome: Connecting Spatial Organization to Function. *Cell Host Microbe* 21, 433–442. doi:10.1016/j.chom.2017.03.010.
- Turner, T. R., James, E. K., and Poole, P. S. (2013). The Plant Microbiome. *Genome Biol.* 14, 209. doi:10.1016/B978-0-12-417163-3.00011-1.
- Ulrich, W., Zaplata, M. K., Winter, S., and Fischer, A. (2017). Spatial distribution of functional traits indicates small scale habitat filtering during early plant succession. *Perspect. Plant Ecol. Evol. Syst.* 28, 58–66. doi:10.1016/j.ppees.2017.08.002.
- Vandenkoornhuyse, P., Dufresne, A., Quaiser, A., Gouesbet, G., Binet, F., Francez, A. J., et al. (2010). Integration of molecular functions at the ecosystemic level: Breakthroughs and future goals of environmental genomics and post-genomics. *Ecol. Lett.* 13, 776–791. doi:10.1111/j.1461-0248.2010.01464.x.
- Vandenkoornhuyse, P., Quaiser, A., Duhamel, M., Van, A. Le, and Dufresne, A. (2015). The importance of the microbiome of the plant holobiont. *New Phytol.* 206. doi:10.1111/nph.13312 Key.
- Vázquez-Castellanos, J. F., Biclot, A., Vrancken, G., Huys, G. R., and Raes, J. (2019). Design of synthetic microbial consortia for gut microbiota modulation. *Curr. Opin. Pharmacol.* 49, 52–59. doi:10.1016/j.coph.2019.07.005.
- Vellend, M. (2010). Conceptual synthesis in community ecology. *Q. Rev. Biol.* 85, 183–206. doi:10.1086/652373.
- Vellend, M., Srivastava, D. S., Anderson, K. M., Brown, C. D., Jankowski, J. E., Kleynhans, E. J., et al. (2014). Assessing the relative importance of neutral stochasticity in ecological communities. *Oikos* 123, 1420–1430. doi:10.1111/oik.01493.
- Venturelli, O. S., Carr, A. V., Fisher, G., Hsu, R. H., Lau, R., Bowen, B. P., et al. (2018). Deciphering microbial interactions in synthetic human gut microbiome communities. *Mol. Syst. Biol.* 14, 1–19. doi:10.1525/msb.20178157.
- Vet, S., Gelens, L., and Gonze, D. (2020). Mutualistic cross-feeding in microbial systems generates bistability via an Allee effect. *Sci. Rep.* 10, 1–12. doi:10.1038/s41598-020-63772-4.
- Violle, C., Nemergut, D. R., Pu, Z., and Jiang, L. (2011). Phylogenetic limiting similarity and competitive exclusion. *Ecol. Lett.* 14, 782–787. doi:10.1111/j.1461-0248.2011.01644.x.
- Voges, M. J. E. E., Bai, Y., Schulze-Lefert, P., and Sattely, E. S. (2019). Plant-derived coumarins shape the composition of an *Arabidopsis* synthetic root microbiome. *Proc. Natl. Acad. Sci. U. S. A.* 116, 12558–12565. doi:10.1073/pnas.1820691116.
- Volkova, S., Matos, M. R. A., Mattanovich, M., and de Mas, I. M. (2020). Metabolic modelling as a framework for metabolomics data integration and analysis. *Metabolites* 10, 1–27. doi:10.3390/metabo10080303.
- Vrancken, G., Gregory, A. C., Huys, G. R. B., Faust, K., and Raes, J. (2019). Synthetic ecology of the human gut microbiota. *Nat. Rev. Microbiol.* 17, 754–763. doi:10.1038/s41579-019-0264-8.
- Wade, M. J., Harmand, J., Benyahia, B., Bouchez, T., Chaillou, S., Cloez, B., et al. (2016). Perspectives in mathematical modelling for microbial ecology. *Ecol. Modell.* 321, 64–74. doi:10.1016/j.ecolmodel.2015.11.002.

General bibliography

- Walker, R. W., Clemente, J. C., Peter, I., and Loos, R. J. F. (2017). The prenatal gut microbiome: are we colonized with bacteria in utero? *Pediatr. Obes.* 12, 3–17. doi:10.1111/ijpo.12217.
- Wampach, L., Heintz-Buschart, A., Hogan, A., Muller, E. E. L., Narayanasamy, S., Laczny, C. C., et al. (2017). Colonization and succession within the human gut microbiome by archaea, bacteria, and microeukaryotes during the first year of life. *Front. Microbiol.* 8, 1–21. doi:10.3389/fmicb.2017.00738.
- Wassermann, B., Cernava, T., Müller, H., Berg, C., and Berg, G. (2019). Seeds of native alpine plants host unique microbial communities embedded in cross-kingdom networks. *Microbiome* 7, 108. doi:10.1186/s40168-019-0723-5.
- Watson, E., Yilmaz, L. S., and Walhout, A. J. M. (2015). Understanding Metabolic Regulation at a Systems Level: Metabolite Sensing, Mathematical Predictions, and Model Organisms. *Annu. Rev. Genet.* 49, 553–575. doi:10.1146/annurev-genet-112414-055257.
- Welch, J. L. M., Hasegawa, Y., McNulty, N. P., Gordon, J. I., and Borisy, G. G. (2017). Spatial organization of a model 15-member human gut microbiota established in gnotobiotic mice. *Proc. Natl. Acad. Sci. U. S. A.* 114, E9105–E9114. doi:10.1073/pnas.1711596114.
- West, S. A., Griffin, A. S., Gardner, A., and Diggle, S. P. (2006). Social evolution theory for microorganisms. *Nat. Rev. Microbiol.* 4, 597–607. doi:10.1038/nrmicro1461.
- Whitman, W. B., Coleman, D. C., and Wiebe, W. J. (1998). Prokaryotes: The unseen majority. *Proc. Natl. Acad. Sci. U. S. A.* 95, 6578–6583. doi:10.1073/pnas.95.12.6578.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York Available at: <https://ggplot2.tidyverse.org>.
- Widder, S., Allen, R. J., Pfeiffer, T., Curtis, T. P., Wiuf, C., Sloan, W. T., et al. (2016). Challenges in microbial ecology: Building predictive understanding of community function and dynamics. *ISME J.* 10, 2557–2568. doi:10.1038/ismej.2016.45.
- Wilkinson, D. M., and Sherratt, T. N. (2001). Horizontally acquired mutualisms , an unsolved problem in ecology ? *Oikos* 92, 377–384.
- Winter, C., Bouvier, T., Weinbauer, M. G., and Thingstad, T. F. (2010). Trade-Offs between Competition and Defense Specialists among Unicellular Planktonic Organisms: the “Killing the Winner” Hypothesis Revisited. *Microbiol. Mol. Biol. Rev.* 74, 42–57. doi:10.1128/mmbr.00034-09.
- Wippel, K., Tao, K., Niu, Y., Zgadzaj, R., Kiel, N., Guan, R., et al. (2021). Host preference and invasiveness of commensal bacteria in the Lotus and Arabidopsis root microbiota. *Nat. Microbiol.* 6, 1150–1162. doi:10.1038/s41564-021-00941-9.
- Womack, A. M., Bohannan, B. J. M., and Green, J. L. (2010). Biodiversity and biogeography of the atmosphere. *Philos. Trans. R. Soc. B Biol. Sci.* 365, 3645–3653. doi:10.1098/rstb.2010.0283.
- Woo, S. L., and Pepe, O. (2018). Microbial consortia: Promising probiotics as plant biostimulants for sustainable agriculture. *Front. Plant Sci.* 9, 7–12. doi:10.3389/fpls.2018.01801.
- Wu, B. W., Gao, C., Chen, L., Buscot, F., Goldmann, K., Purahong, W., et al. (2018). Host phylogeny is a major determinant of fagaceae-associated ectomycorrhizal fungal community assembly at a regional scale. *Front. Microbiol.* 9, 1–12. doi:10.3389/fmicb.2018.02409.
- Wu, M., and Eisen, J. A. (2008). A simple, fast, and accurate method of phylogenomic inference. *Genome Biol.* 9. doi:10.1186/gb-2008-9-10-r151.

General bibliography

- Xiong, C., Zhu, Y. G., Wang, J. T., Singh, B., Han, L. L., Shen, J. P., et al. (2021). Host selection shapes crop microbiome assembly and network complexity. *New Phytol.* 229, 1091–1104. doi:10.1111/nph.16890.
- Xu, Q., Ling, N., Chen, H., Duan, Y., Wang, S., Shen, Q., et al. (2020). Long-Term Chemical-Only Fertilization Induces a Diversity. *mSystems* 5, 1–10.
- Yadav, A. N., Verma, P., Kour, D., Rana, K. L., Kumar, V., Singh, B., et al. (2017). Plant Microbiomes and Its Beneficial Multifunctional Plant Growth Promoting Attributes. *Int. J. Environ. Sci. Nat. Resour.* 3. doi:10.19080/ijesnr.2017.03.555601.
- Yan, Z.-Z., Chen, Q.-L., Li, C.-Y., Thi Nguyen, B.-A., Zhu, Y.-G., He, J.-Z., et al. (2021). Biotic and abiotic factors distinctly drive contrasting biogeographic patterns between phyllosphere and soil resistomes in natural ecosystems. *ISME Commun.* 1. doi:10.1038/s43705-021-00012-4.
- Yang, J., Wang, Y., Cui, X., Xue, K., Zhang, Y., and Yu, Z. (2019). Habitat filtering shapes the differential structure of microbial communities in the Xilingol grassland. *Sci. Rep.* 9, 1–12. doi:10.1038/s41598-019-55940-y.
- Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., et al. (2012). Human gut microbiome viewed across age and geography. *Nature* 486, 222–227. doi:10.1038/nature11053.
- Yoshida, T., Ellner, S. P., Jones, L. E., Bohannan, B. J. M., Lenski, R. E., and Hairston, N. G. (2007). Cryptic population dynamics: Rapid evolution masks trophic interactions. *PLoS Biol.* 5, 1868–1879. doi:10.1371/journal.pbio.0050235.
- Young, J. P. W. (2016). Bacteria Are Smartphones and Mobile Genes Are Apps. *Trends Microbiol.* 24, 931–932. doi:10.1016/j.tim.2016.09.002.
- Yu, X. J., Walker, D. H., Liu, Y., and Zhang, L. (2009). Amino acid biosynthesis deficiency in bacteria associated with human and animal hosts. *Infect. Genet. Evol.* 9, 514–517. doi:10.1016/j.meegid.2009.02.002.
- Zélé, F., Magalhães, S., Kéfi, S., and Duncan, A. B. (2018). Ecology and evolution of facilitation among symbionts. *Nat. Commun.* 9, 1–12. doi:10.1038/s41467-018-06779-w.
- Zelezniak, A., Andrejev, S., Ponomarova, O., Mende, D. R., Bork, P., and Patil, K. R. (2015). Metabolic dependencies drive species co-occurrence in diverse microbial communities. 112. doi:10.1073/pnas.1522642113.
- Zengler, K., and Zaramela, L. S. (2018). The social network of microorganisms - How auxotrophies shape complex communities. *Nat. Rev. Microbiol.* 16, 383–390. doi:10.1038/s41579-018-0004-5.
- Zhelnina, K., Louie, K. B., Hao, Z., Mansoori, N., Da Rocha, U. N., Shi, S., et al. (2018a). Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nat. Microbiol.* 3, 470–480. doi:10.1038/s41564-018-0129-3.
- Zhelnina, K., Zengler, K., Newman, D., and Northen, T. R. (2018b). Need for Laboratory Ecosystems To Unravel the Structures. *MBio* 9, 1–8.
- Zhang, H., Zhu, S., John, R., Li, R., Liu, H., and Ye, Q. (2017). Habitat filtering and exclusion of weak competitors jointly explain fern species assemblage along a light and water gradient. *Sci. Rep.* 7, 1–10. doi:10.1038/s41598-017-00429-9.
- Zhang, W., Corwin, J. A., Copeland, D. H., Feusier, J., Eshbaugh, R., Cook, D. E., et al. (2019). Plant–necrotroph co-transcriptome networks illuminate a metabolic battlefield. *eLife* 8, 1–32. doi:10.7554/eLife.44279.
- Zhou, J., and Ning, D. (2017). Stochastic Community Assembly : Does It Matter in Microbial Ecology? *Microbiol. Mol. Biol. Rev.* 81, 1–32.
- Zhou, Z., Wang, C., and Luo, Y. (2018). Response of soil microbial communities to altered precipitation: A global synthesis. *Glob. Ecol. Biogeogr.* 27, 1121–1136. doi:10.1111/geb.12761.

General bibliography

- Zhu, X., Campanaro, S., Treu, L., Seshadri, R., Ivanova, N., Kougias, P. G., et al. (2020). Metabolic dependencies govern microbial syntrophies during methanogenesis in an anaerobic digestion ecosystem. *Microbiome* 8, 1–14. doi:10.1186/s40168-019-0780-9.
- Zilber-Rosenberg, I., and Rosenberg, E. (2008). Role of microorganisms in the evolution of animals and plants: The hologenome theory of evolution. *FEMS Microbiol. Rev.* 32, 723–735. doi:10.1111/j.1574-6976.2008.00123.x.
- Zogg, G. P., Zak, D. R., Ringelberg, D. B., White, D. C., MacDonald, N. W., and Pregitzer, K. S. (1997). Compositional and Functional Shifts in Microbial Communities Due to Soil Warming. *Soil Sci. Soc. Am. J.* 61, 475. doi:10.2136/sssaj1997.03615995006100020015x.
- Zomorrodi, A. R., and Segrè, D. (2016). Synthetic Ecology of Microbes: Mathematical Models and Applications. *J. Mol. Biol.* 428, 837–861. doi:10.1016/j.jmb.2015.10.019.
- Zomorrodi, A. R., and Segrè, D. (2017). Genome-driven evolutionary game theory helps understand the rise of metabolic interdependencies in microbial communities. *Nat. Commun.* 8, 1–11. doi:10.1038/s41467-017-01407-5.
- Zuñiga, C., Li, C. T., Yu, G., Al-Bassam, M. M., Li, T., Jiang, L., et al. (2019). Environmental stimuli drive a transition from cooperation to competition in synthetic phototrophic communities. *Nat. Microbiol.* 4, 2184–2191. doi:10.1038/s41564-019-0567-6.
- Zuppinger-Dingley, D., Schmid, B., Petermann, J. S., Yadav, V., De Deyn, G. B., and Flynn, D. F. B. (2014). Selection for niche differentiation in plant communities increases biodiversity effects. *Nature* 515, 108–111. doi:10.1038/nature13869.

Résumé

Les communautés microbiennes forment un réseau complexe d’interactions entre organismes, qui façonnent leur structure. Parmi ces interactions, les échanges de métabolites entre organismes, nommés ‘cross-feeding’, sont considérés comme fréquents et importants. Ces échanges peuvent être prédits *in silico* avec des réseaux métaboliques inférés à partir de génomes. Ces modèles sont intégrés dans un cadre de travail nommé ‘microbial system ecology’, qui rassemble plusieurs méthodes de modélisation ainsi que leur validation expérimentale, à différentes échelles d’étude. Dans cette thèse, nous avons utilisé ce cadre pour prédire de nombreuses et hypothétiques associations de bactéries du microbiote racinaire d'*Arabidopsis thaliana* permettant le production de métabolites non productibles par des bactéries seules. Nous avons également modélisé l’impact des nutriments disponibles dans le milieu, et prédit que ces contraintes nutritionnelles sont largement compensées par des comportements de cross-feeding. En parallèle, nous avons également utilisé des métriques simples pour corrélérer le potentiel de cross-feeding et de compétition avec le métabolisme des bactéries et leur distance phylogénétique. Ces métriques ont été testées expérimentalement avec un ensemble de petites communautés synthétiques. La compétition semblait dominer dans la plupart des communautés, mais a été déterminée comme étant plus faible dans les communautés abritant des souches à métabolisme différent. Nous avons identifié quelques bactéries profitant des exsudats d’autres espèces, particulièrement une *Achromobacter sp.*. Finalement, les résultats suggèrent que le cross-feeding est courant mais potentiellement masqué par de la forte compétition pour d’autres ressources, par exemple spatiales.

1 Introduction

Les microorganismes sont reconnus comme des éléments majeurs de nombreux processus environnementaux. Lorsqu’ils sont associés à un organisme hôte (animal ou végétal), ils constituent son microbiote, et sont impliqués dans de nombreux processus qui lui sont bénéfiques (nutrition, résistance aux maladies ou à divers stress environnementaux) (Larimer et al., 2010; Fierer, 2017; Amor and Bello, 2019). Les nombreuses fonctions associées aux communautés microbiennes ont également conduit au design de petits communautés synthétiques, optimisées pour assurer une fonction donnée une fois inoculée dans un environnement. Il peut par exemple s’agir de protection des plantes, de médecine, ou de dépollution (Liu et al., 2017; Chubukov et al., 2018; Liu and Nielsen, 2019).

1.1 Les règles d'assemblages des communautés microbiennes

Un aspect fondamental des communautés microbiennes, est l'étude des mécanismes d'assemblage qui leur donnent leur structure (i.e. la composition et abondances des espèces) et leur dynamique temporelle. L'écologie des communautés est la discipline qui cherche à caractériser les assemblages d'espèces. Ces mécanismes sont nombreux et regroupés en quatre processus distincts (Vellend, 2010) :

- La dérive : des changements aléatoires d'abondances des espèces (natalité, mortalité)
- La dispersion (migration ...)
- La spéciation, i.e. l'apparition de nouvelles espèces
- La sélection, seul processus déterministe, où seuls les individus / espèces les plus adaptés sont sélectionnés survivent, et se reproduisent.

L'importance relative de chaque processus, particulièrement entre ceux stochastiques et ceux déterministes, est encore relativement peu connue, malgré d'abondantes connaissances accumulées (Morrison-Whittle and Goddard, 2015). D'une manière générale, il est établi que les processus stochastiques sont plus déterminants dans la détermination de pool d'espèces globaux, tandis que les processus déterministes sont plus importants à l'échelle des pools d'espèces locaux (Götzenberger et al., 2012). Les effets déterministes impliqués dans la sélection sont variés, et impliquent la réponse des organismes à un ou des facteurs environnementaux, qu'ils soient abiotiques (humidité, température, salinité, acidité ...), ou biotiques, c'est-à-dire comment les espèces interagissent les unes avec les autres.

1.1.1 La théorie des niches

La théorie des niches considère que chaque espèce est définie par sa « niche » écologique, un ensemble de conditions biotiques et abiotiques qui, une fois considérées comme des « axes », définissent le volume dans lequel cette espèce peut s'implanter et survivre. La coexistence de deux espèces est possible car leur niches sont différentes, ce qui limite la compétition entre espèces : il s'agit de la différenciation de niches, qui peut s'opérer de nombreuses manières différentes (occupation d'espaces différents, consommation de ressources différentes..) (Brochet et al., 2021). La différenciation de niche entraîne un décalage entre la niche fondamentale d'une espèce et sa niche réelle, c'est-à-dire la différences entre les dimensions qu'elle peut potentiellement occuper dans un environnement et celle qu'elle occupe réellement. Selon la théorie des niches, les espèces que l'on trouve au sein d'une communauté sont celles qui ont su différencier suffisamment leurs niches par rapport à celles des autres.

A l'opposé de la différentiation de niches, la théorie de « l'habitat filtering » propose que les traits les plus aptes pour survivre à un environnement sont sélectionnés. En conséquence, les espèces présentes dans un même biotope ont toutes des traits similaires (Zhang et al., 2017).

1.1.2 Facteurs externes impliqués dans les règles d'assemblages des communautés

La structure d'une communauté microbienne est affectée par de nombreux facteurs stochastiques, tels que les précipitations, la dispersion par le vent, l'ordre d'arrivée des espèces dans un milieu (Fukami et al., 2010). Les facteurs physico-chimiques déjà évoqués de l'environnement ont également un poids certain (salinité, acidité, composition en nutriments) (Cui et al., 2021). D'autres facteurs tels que l'utilisation des sols, la connectivité à d'autres communautés, le transport par des vecteurs animaux, ainsi que les plantes et animaux voisins sont également en jeu (Hacquard, 2016; Griggs et al., 2021). Enfin, dans le cas des microbiotes, l'hôte exerce une sélection. Par exemple, les plantes conditionnent fortement quels microbes du sol qui pourront coloniser ses différents compartiments, grâce à leurs exsudats racinaires (Bulgarelli et al., 2013; Griggs et al., 2021). Le transfert peut également être vertical, de génération en génération (Walker et al., 2017).

1.1.2 Les interactions biotiques entre organismes

En comparaison des facteurs abiotiques, les différentes interactions entre les membres d'une même communauté microbienne sont moins connues. Elles sont en effet particulièrement difficiles à étudier en raison de la forte concentration de nombreux individus de nombreuses espèces au même endroit, les patterns observés étant le résultat de multiples interactions. De plus, elles dépendent également d'autres paramètres, tels que les échelles spatiales et temporelles, ou encore les ressources disponibles (Kelsic et al., 2015; Zuñiga et al., 2019; Fritts et al., 2021). Les interactions précises par paires d'espèces sont donc moins bien connues. Elles sont néanmoins variées, allant de l'antagonisme et la compétition (prédatation, sécrétion d'antibiotiques, compétition pour une ressource) au mutualisme (coopération pour des bénéfices réciproques). Les interactions peuvent également être commensales (sans effet pour une espèce, bénéfique pour l'autre) ou ammenses (sans effet pour une espèce, négative pour l'autre) (Zélé et al., 2018; Coyte and Rakoff-Nahoum, 2019; Pacheco and Segrè, 2019).

1.1.3 Une interaction métabolique : le cross-feeding

Une interaction de « cross-feeding » entre deux microorganismes désigne la situation où l'un d'entre eux (le bénéficiaire) assimile un exsudat issu du métabolisme de l'autre (le fournisseur), et l'utilise comme nutriment pour son propre métabolisme. Une telle interaction peut être tantôt bénéfique dans le cas où l'échange est mutuel, tantôt négatif pour le fournisseur (dans un cas de parasitisme), tantôt sans effet pour le fournisseur (relation commensale) (figure 1A). Il en existe

différents types (figure 1B), et peut être obligatoire ou non à la survie du bénéficiaire, ou bien impliquer un bénéficiaire spécifiquement associé à un fournisseur précis ou généraliste, ou encore restreint ou non aux déchets du bénéficiaire (Morris et al., 2013; Pande and Kost, 2017; Hillesland, 2018; Zengler and Zaramela, 2018; Smith et al., 2019). Cette interaction repose sur la sécrétion de « biens communs » par les fournisseurs, qui peuvent être des enzymes, protéines, sidérophores, cofacteurs, acides aminés, vitamines (etc) (Fritts et al., 2021), mis à disposition de tous les organismes qui peuvent les assimiler, sous condition qu'ils puissent être acheminés du fournisseur au bénéficiaire (figure 1C) (Sung et al., 2017; D'Souza et al., 2018; Zengler and Zaramela, 2018). Les nutriments déjà disponibles dans le milieu vont notamment dicter le métabolisme des microorganismes, ainsi que quelles interactions de cross-feeding seront indispensables ou non à la survie de certains (figure 1D) (Magnúsdóttir et al., 2017). En effet, sous de pauvres conditions nutritionnelles, il a été démontré que de nombreux microorganismes compensent ces contraintes par de nombreux comportements de cross-feeding, même entre des espèces habituellement compétitrices (Zuñiga et al., 2019; Fritts et al., 2021). Pour finir, les échanges de composés entre microorganismes permettent la production de métabolites non productibles par des organismes isolés, donnant à la communauté microbienne une « plus-value ».

De nombreuses approches expérimentales ont conclu à la prépondérance de la compétition dans les interactions microbiennes. Cependant, des travaux récents tendent à montrer que l'échange de métabolites, qu'ils soient commensaux ou mutualistes, sont également d'une grande importance, et suggèrent l'existence de compromis entre compétition et coopération (Zelezniak et al., 2015; Coyte and Rakoff-Nahoum, 2019; Thommes et al., 2019; Machado et al., 2021). Parmi ces études, un grand nombre s'appuie sur la modélisation du métabolisme à partir des génomes et la prédiction des interactions métaboliques, dont un grand nombre est encore à valider expérimentalement. De telles approches impliquent de nombreuses méthodes, et ont été regroupées dans un cadre de travail dénommé « microbial system ecology »

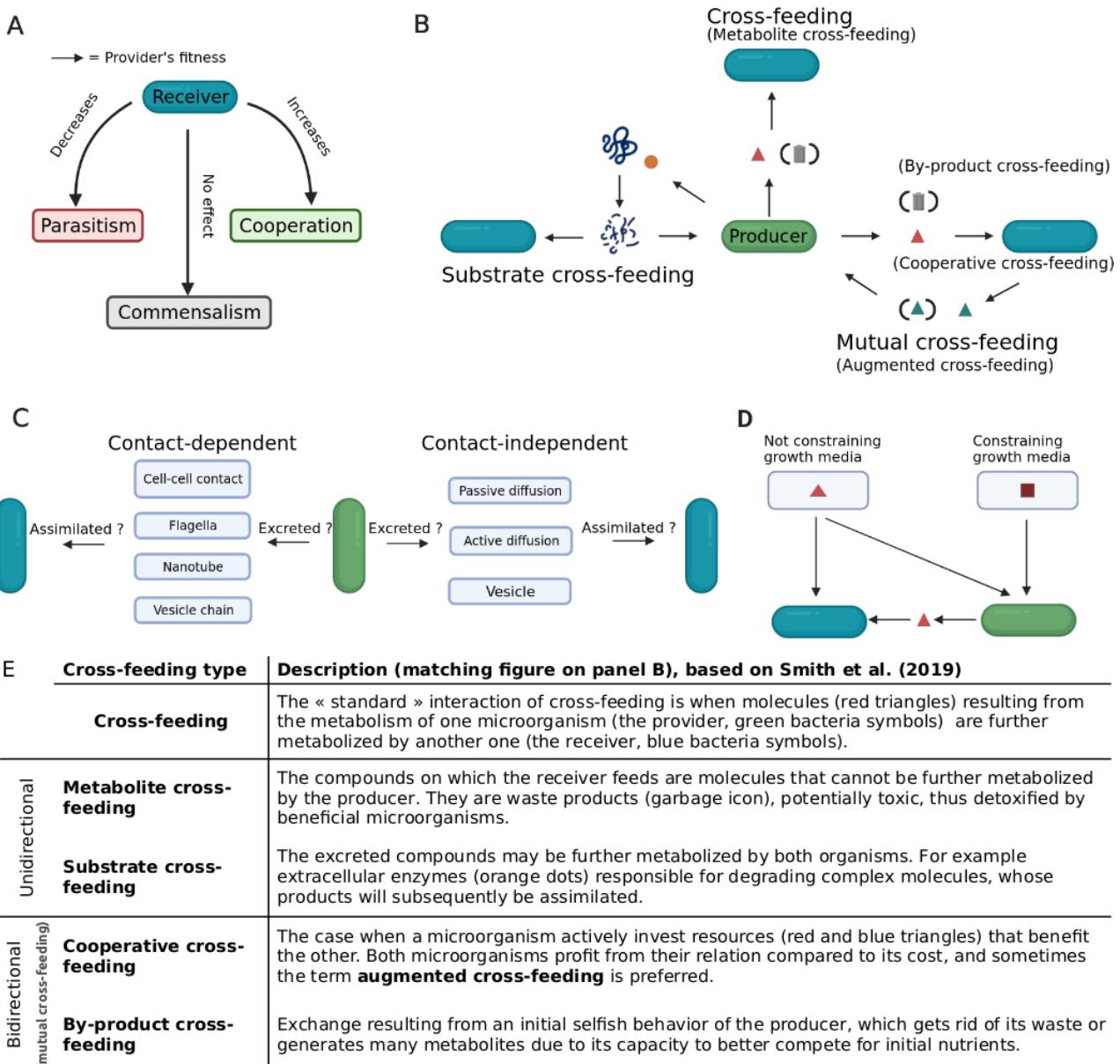


Figure 1: (A) Les interactions de cross-feeding ne sont pas nécessairement mutualistes. Selon l'effet du bénéficiaire sur le fournisseur, l'interaction peut être mutualiste, commensale ou parasitaire. (B) Il existe différents types de cross-feeding (Smith et al. 2019), selon la nature du composé échangé et sur sa directionnalité (échange mutuel ou unidirectionnel). (C) L'existence de cross-feeding est conditionné à la sécrétion au transport, et à l'assimilation du « bien commun » échangé (D'Souza et al. 2018). (D) Les interactions métaboliques dépendent du contexte environnemental. Si un nutriment est déjà disponible dans le milieu, le cross-feeding n'est pas obligatoire pour l'organisme bénéficiaire. Si un nutriment n'est disponible que par excrétion d'un organisme producteur, le cross-feeding devient obligatoire pour le bénéficiaire (E) Définitions des différents types de cross-feeding représentées en B).

1.2 L'approche par ‘microbial system ecology’

La biologie des systèmes est la discipline qui s'applique à modéliser mathématiquement et informatiquement le fonctionnement des différents systèmes biologiques : cellules, tissus, organes (etc). Le but est d'obtenir les propriétés émergentes du système par la connexion de ses différents composants, plutôt que les propriétés de ses composants seuls : le système est vu comme un réseau. La « microbial system ecology » (‘MSE’) reprend les principes de la biologie des systèmes en les appliquant à l’écologie des communautés (Muller et al., 2018).

L'échelle des communautés microbiennes présente un avantage : la plupart des organismes d'un microbiote sont unicellulaires (les bactéries par exemple), et l'application des méthodes de MSE permet un croisement entre outils de biologie cellulaire et d'écologie des communautés. Une majeure partie des modèles se base sur les réseaux métaboliques des organismes reconstruits à partir de leurs génomes séquencés puis annotés (ce qu'on appelle la génomique) (Mendoza et al., 2019; Jansma and Aidy, 2020). Le croisement avec d'autres données -omiques (transcriptomique, protéomique, métabolomique) permet de renforcer les modèles. Les différentes interactions biotiques entre organismes peuvent ainsi être prédits à partir de la simulation de leur mode de leur écologie (Muller et al., 2018)(Mee et al., 2014; Kumar et al., 2019). Néanmoins, nombre de ces modèles sont encore vulnérables à de nombreux biais, et servent souvent de générateurs d'hypothèses à tester expérimentalement (Coyte and Rakoff-Nahoum, 2019; Goyal et al., 2021).

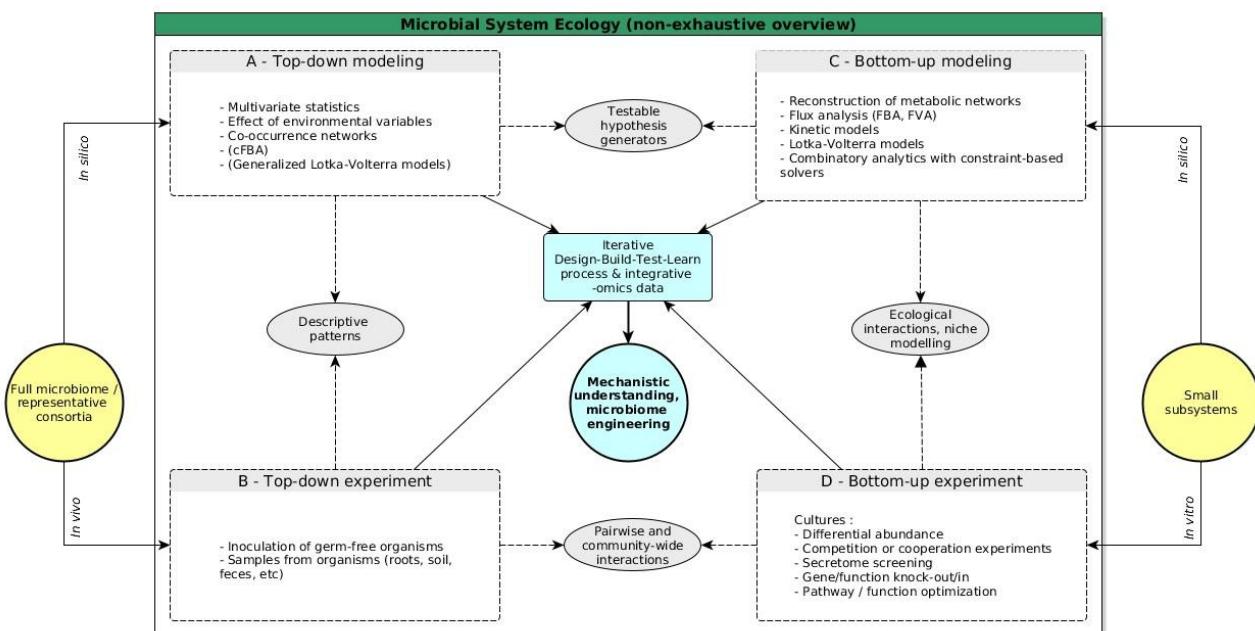


Figure 2: résumé schématique des approches top-down et bottom-up en microbial system ecology, avec une liste non-exhaustive de méthodes, techniques, et analyses. Sous ce cadre de travail, analyser la structure et la dynamique d'un microbiote implique un constant va-et-vient entre les approches top-down / bottom-up et *in silico* / *in vitro* / *in vivo*, de façon itérative en direction du but recherché (par exemple le design d'une communauté synthétique). La modélisation top-down (A) utilise des données -omiques à grande échelle associé à des analyses descriptives et statistiques multivariées pour détecter des patterns (spatiaux, temporels, environnementaux, fonctionnels ...), par exemple à partir de l'abondance des espèces ou de la diversité de leurs gènes. Ces données sont souvent issues d'expériences à grande échelle (B), comme l'échantillonnage d'une communauté entière. La modélisation bottom-up préfère une approche réductionniste, où des petits sous-système sont analysés de façon plus détaillée et mécanistique, notamment concernant les interactions entre microorganismes. De nombreux modèles se basent sur des réseaux métaboliques construits à partir de génomes annotés, pouvant prédire de la compétition pour un nutriment ou bien son échange par cross-feeding. De plus, la dynamique de ces interactions peut être modélisé avec des modèles de type Lotka-Volterra. (D) idéalement, les prédictions des modèles doivent être testées expérimentalement. Chaque méthode présente ses avantages et inconvénients, et fournit des connaissances précises qui doivent être synthétisées en un ensemble cohérent afin d'obtenir une analyse complète (plus de détails dans Shahzad 2012, Fransoza et al. 2015, Amor and Bello 2019, Lawson et al. 2019, Lloyd-Price et al. 2019, Vrancken et al. 2019).

Les différentes approches de la MSE s'articulent autour de deux grandes catégories de méthodes. Les méthodes dites « top-down » (figure A&B), qui sont essentiellement descriptives et considèrent de grandes communautés et tentent d'y découvrir des patterns d'organisation. A l'opposé, les méthodes dites « bottom-up » (figure C&D) se basent sur de très petites sous-communautés afin d'étudier finement les interactions entre microorganismes. Toutes peuvent être modélisatrices ou expérimentales, et se nourrissent et s'interrogent les unes les autres dans un cadre de progression itératif dénommé « Design-Build-Test-Learn » (figure 2) (Lawson et al., 2019).

1.3 Objectifs de la thèse

Cette thèse de doctorat a pour but d'étudier les interactions métaboliques par cross-feeding entre des bactéries du microbiote racinaire d'*Arabidopsis thaliana*, représenté par 193 génomes annotés et considérés comme représentatif du microbiote principal de la plante.

Tout d'abord, l'introduction générale a été enrichie d'une mini-review (soumise à *Frontiers in Microbiology*) pensée comme une porte d'entrée à la microbial system ecology pour les chercheurs non familiers avec cette discipline.

Ensuite, un travail important a été dédié à la prédiction d'interactions de cross-feeding entre les différentes bactéries, sous différentes contraintes nutritionnelles, incluant notamment l'effet des exsudats racinaires. Les échanges de métabolites à l'échelle de la communauté entière compensaient largement les différentes contraintes nutritionnelles. Après avoir ciblé la production de métabolites en particulier (acides aminés, vitamines, phytohormones), de nombreuses interactions hypothétiques ont été recensées, avec une redondance importante de certaines bactéries parmi les combinaisons proposées. En parallèle, nous avons établi une corrélation entre le potentiel à coopérer et la distance phylogénétique entre deux souches bactériennes.

Finalement, un protocole expérimental simple a été mis au point pour tester la vraisemblance de la corrélation établie entre distance phylogénétique (traduite en diversité de métabolites productibles, car plus des bactéries sont distantes, plus leur production est variée) et potentiel de cross-feeding. 41 petits communautés synthétiques de 4 bactéries (dont une commune à toutes les communautés) ont été construites le long d'un gradient de diversité métabolique, puis cultivées en laboratoire. La compétition semblait moins forte entre les espèces avec une plus grande diversité métabolique (donc une plus grande distance phylogénétique). De plus, nous avons identifié quelques souches bactériennes capables de se nourrir des exsudats des autres, notamment une *Achromobacter* sp.

2. Matériel et méthodes

Les données de départ consistaient en un pool de 193 génomes annotés de souches bactériennes considérées comme représentatives du microbiote racinaire d'*Arabidopsis thaliana*.

2.1 Modélisation et prédictions

Pour chaque génome (et donc chaque souche), un réseau métabolique (« Genome-scale metabolic model » : GEM) a été reconstruit à l'aide des outils *Pathway Tools* (Karp et al., 2011; Belcour et al., 2020), *AuReMe* (Aite et al., 2018), et la base de données *MetaCyc* (Caspi et al., 2008). Un arbre phylogénétique était également disponible. Plusieurs métriques pouvaient être extraites de chaque GEM :

- Le nombre et la liste des métabolites productibles (« Predicted Producible Metabolites » : PPM).
- Plusieurs PPM de différents GEMs peuvent être calculés en même temps, et ainsi inclure les métabolites uniquement productibles grâce aux échanges de composés intermédiaires entre GEMs. C'est la « plus-value » apportée par la communauté.
- Dans le cas de plusieurs GEMs à la fois, l'intersection des PPM des GEMs, c'est-à-dire les métabolites productibles par tous les GEMs considérés (« Core Predicted Producible Metabolites » : CPPM).
- Le nombre de métabolites cibles (acides aminés, vitamines, phytohormones) productibles (« Targeted Predicted Producible Metaoblitites » : TPPM), par un ou plusieurs GEMs à la fois (en incluant les composés productibles par complétion métaboliques grâce au cross-feeding).

Ces trois métriques peuvent être calculées sous une contrainte nutritionnelle, c'est-à-dire une simulation des nutriments directement disponibles dans l'environnement, ou sans contrainte. Sous contrainte, la production des GEM est donc limitée par les nutriments disponibles, et sans contrainte, c'est leur production théorique complète qui est obtenue. Plusieurs contraintes ont été utilisées en simulant des milieux de culture, réparties en deux catégories : sévère (milieux pauvres en nutriments, essentiellement minéraux) et modérée (milieux riches en nutriments, notamment carbonés). Chaque contrainte existait en une version alternative, enrichie en composants simulant des exsudats racinaires (« Artificial Root Exudates » : ARE). Finalement, les prédictions de combinaisons de GEM pour produire les TPPM ont été calculées avec un solveur d'Answer Set programming (ASP, programmation par ensemble-réponse), *Miscoto* (Frioux et al., 2018)

2.2 Approche expérimentale

Les 41 communautés synthétiques (SynComs) testées étaient réparties le long d'un gradient de PPM (également corrélé à la distance phylogénétique moyenne entre les membres des SynComs), parmi un pool de 50000 SynComs générées aléatoirement. Chaque SynCom contenait 4 membres, ce qui correspond à la taille de la plupart des combinaisons renvoyées par le solveur ASP auxquelles a été ajouté une souche supplémentaire ('Root695', *mesorhizobium sp*), constante dans toutes les SynComs. Cette souche était peu compétitrice avec une croissance lente : son ajout a donc été estimé comme peu impactant sur les autres membres, et sa constance avait pour but d'obtenir un éventail d'interactions plus précis pour une souche en particulier. Chaque SynCom a été cultivée en incubateur à 20°C, en milieu liquide minimal « M9 » auquel a été ajouté un mix mimant les exsudats racinaires d'*A. thaliana*.

2. Résultats et discussion

2.1 Effet de différentes contraintes nutritionnelles sur le métabolisme du microbiote racinaire d'*Arabidopsis thaliana*

Dans cette analyse prédictive, le nombre maximal de PPM de chaque GEM, c'est-à-dire sans contrainte nutritionnelle, a été comparé au PPM de ces mêmes GEM sous différentes contraintes nutritionnelles. Il est apparu que les différentes contraintes, qu'elles soient sévères ou modérées, impactaient très fortement le nombre de PPM de chaque GEM (figure 3A) par rapport au PPM non contraints. L'effet des milieux pauvres était plus important que celui des milieux riches. Cependant, l'ajout des ARE compensaient en partie ces contraintes, en augmentant significativement les PPM contraints (figure 3A). De plus, les ARE débloquaient la production des mêmes métabolites d'un milieu à l'autre : en effet, une classification hiérarchique de la liste des métabolites productibles rassemblaient les milieux avec ARE dans une même branche, peu importe la sévérité de la contrainte appliquée (figure 3B). A l'échelle de la communauté, sous l'hypothèse forte que tous les métabolites soient secrétaires et échangeables entre GEM, la plus-value apportée par la communauté compensait largement toutes les contraintes nutritionnelles, en ramenant le PPM de la communauté à une valeur similaire quel que soit la contrainte (figure 3C). Finalement, une attention particulière sur les TPPM a permis de mettre en évidence la sévérité des contraintes nutritionnelles sur le métabolisme des bactéries. La majorité des TPPM était théoriquement productible sans contrainte, alors qu'une faible partie d'entre eux l'étaient sous les différentes contraintes, et ce par un nombre réduit de GEM. Les ARE avaient de nouveau un effet positif, débloquant la production de plusieurs TPPM pour un nombre relativement important de GEM (figure 1D).

Ainsi, nous avons démontré que le métabolisme des 193 bactéries est dépendant des nutriments présents dans l'environnement, et sont particulièrement adaptés pour utiliser les composés carbonés excrétés par les racines de la plante hôte. Nous avons également démontré que des échanges métaboliques entre bactéries (à l'échelle de la communauté entière) peuvent compenser fortement la contrainte appliquée par des limitations en nutriments.

2.2 Corrélation entre distance phylogénétique et diversité métabolique

Dans cette analyse, nous avons cherché à corrélérer différentes métriques issues des génomes et des GEM associés à la diversité (PPM) et redondance (CPPM) métabolique des différentes souches bactériennes, et par extension leur potentiel de coopération ou compétition. Pour différentes tailles de SynCom allant de 2 à 20, et 500 SynComs par taille, nous avons calculé pour chaque SynCom le PPM, le CPPM, la distance phylogénétique moyenne et la taille de génome moyenne.

Nous avons observé que :

- Les PPM augmentent et les CPPM diminuent significativement avec la taille des SynComs. La différence d'une taille à l'autre se réduit au fur et à mesure que la taille grandit (figure 4). De plus, l'étendue des valeurs prise par les différentes métriques se réduit également, témoignant d'un rapprochement progressif de la communauté entière, donc une plus grande ressemblance entre les SynComs.
- Les PPM augmentent et les CPPM diminuent significativement avec la distance phylogénétique moyenne des SynComs. Pour les SynComs de petite taille, un optimum de PPM est atteint vers une distance phylogénétique d'environ 1.3. Cet optimum se transforme en plateau au fur et à mesure que la taille des SynComs augmente (figure 4A&B).
- Les PPM et les CPPM augmentent significativement avec la taille moyenne des génomes (figure 4C&D).
- Bien qu'une corrélation soit établie, la plupart des PPM et CPPM des SynComs sont concentrées autour d'une distance phylogénétique précise (environ 1.3) au lieu de se répartir le long des axes. Cela était dû à la composition taxonomique du pool de bactéries, majoritairement composé de proteobacteria et d'actinobacteria : les combinaisons entre ces souches renvoyaient des métriques autour de ces valeurs. Les valeurs extrêmes correspondent aux combinaisons contenant d'autres phyla, ce qui est également la cause du pic de PPM des petites SynComs : les combinaisons de phyla donnant des SynComs à grande distance phylogénétique représentaient des souches avec des PPM plus faibles que les proteobacteria et actinobacteria.

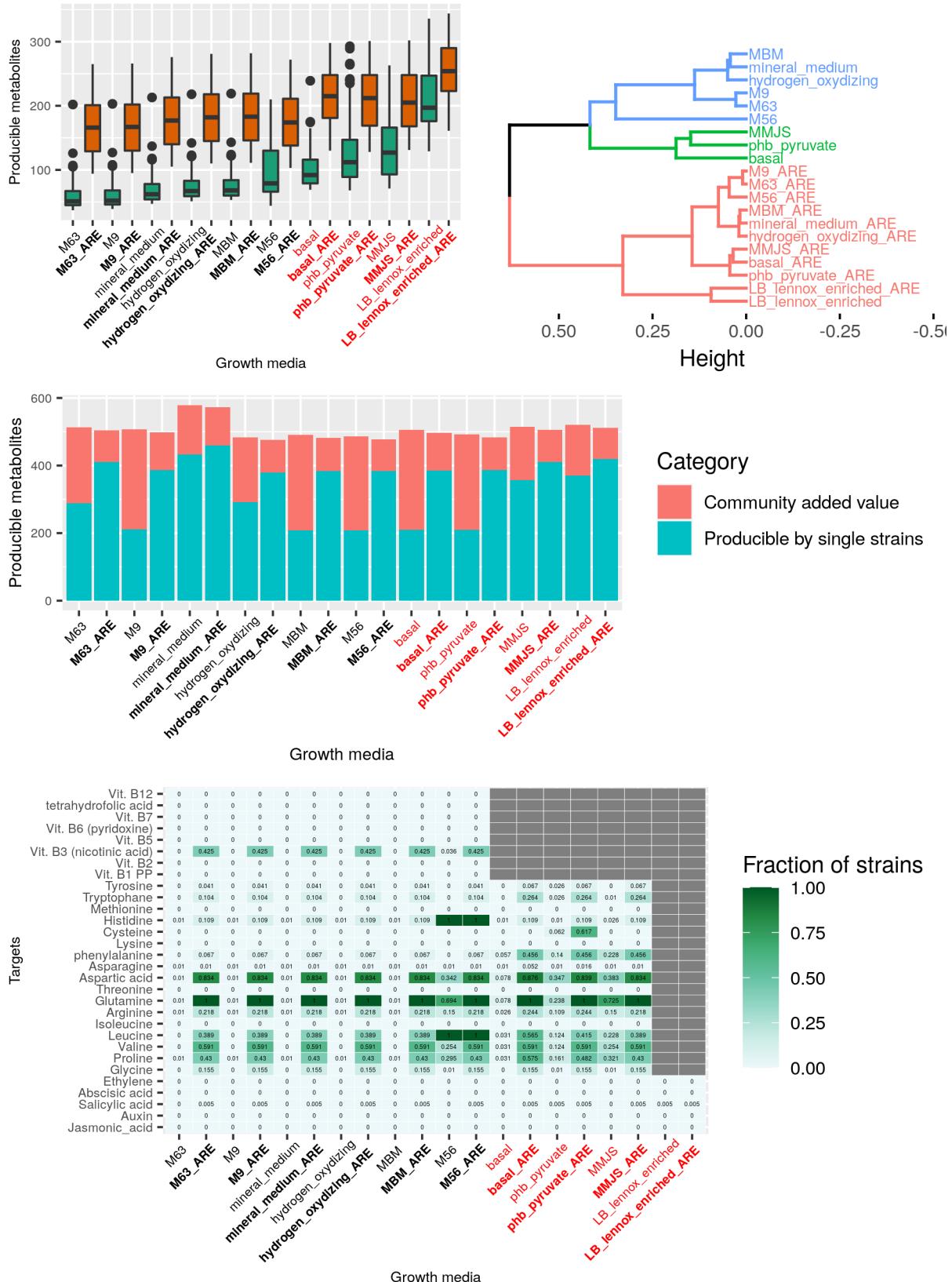


Figure 3: Effet des contraintes nutritionnelles sur les PPM et TPPM. (A) Boxplots du nombre de PPM par GEM selon les contraintes nutritionnelles, comparée au nombre de PPM sans contrainte. (B) Classification hiérarchique des contraintes nutritionnelles selon la composition des PPM de chaque GEM qui en résultent. (C) PPM de la communauté entière selon la contrainte nutritionnelle, avec la plus-value apportée par des interactions de cross-feeding entre les 193 GEM. (D) Détail de la production de TPPM selon les contraintes nutritionnelles. Les nombres indiquent la proportion de GEM capables de produire les TPPM. Les cellules grises sont exclues de l'analyse et correspondent au cas où un milieu de culture modélisé contient déjà un TPPM. Sur toutes les figures, les contraintes sévères sont labellisées en noir, les contraintes modérées en rouge, et les contraintes avec ajout d'ARE sont en gras.

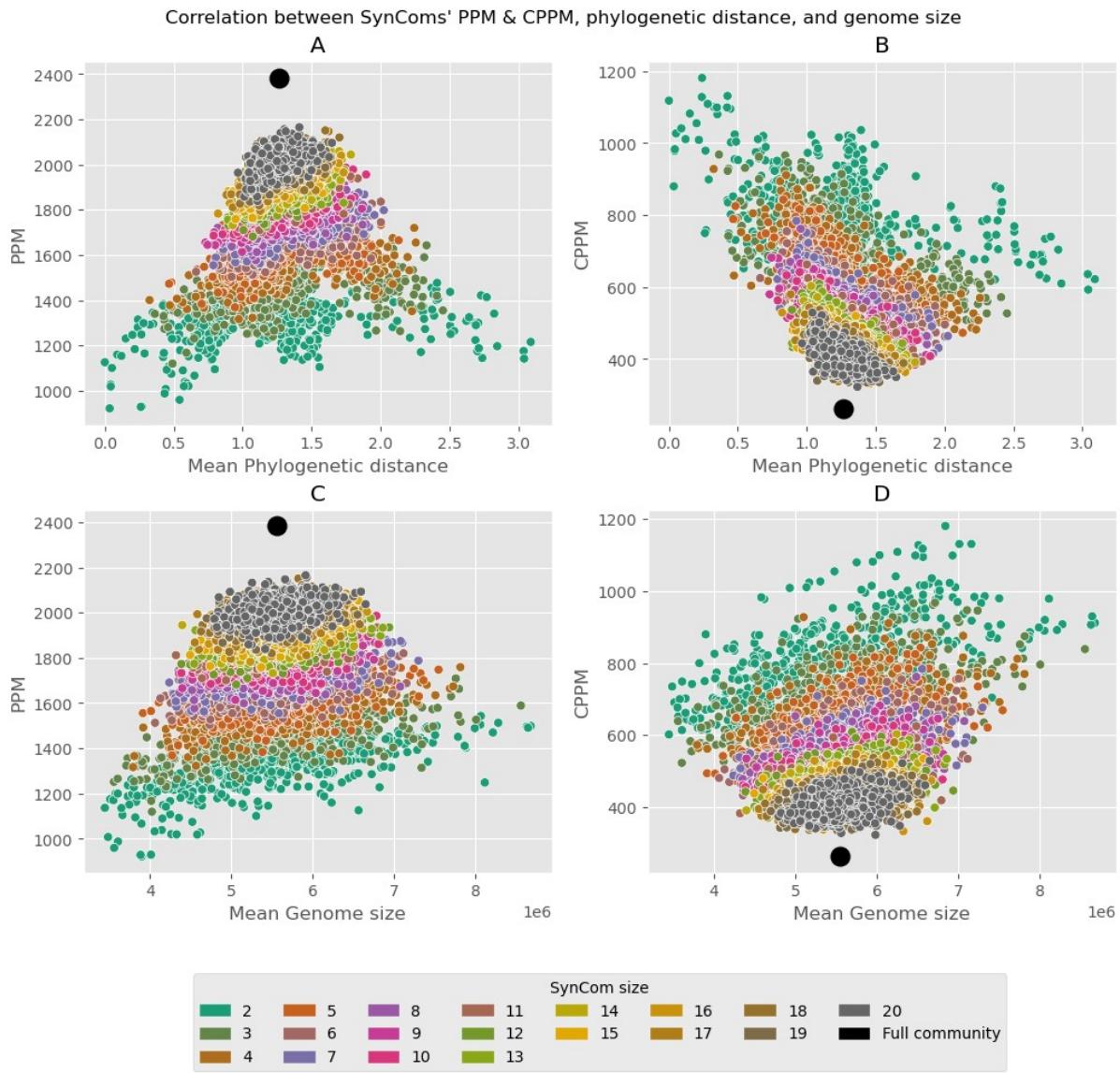


Figure 4: Corrélations entre les PPM (A&C) / CPPM (B&D) et les différentes métriques explicatives : distance phylogénétique moyenne (A&B), taille moyenne des génomes (C&D), et taille des SynComs (code couleur). Les points noirs représentent les valeurs de la communauté entière (193 GEM).

D'après ces résultats, la capacité fonctionnelle de toute la communauté est donc atteignable par un nombre relativement réduit de GEM (une combinaisons d'une dizaine de GEM étant déjà plus proches de la communauté entière qu'une combinaisons de 2 GEM). Ceci illustre une certaine redondance métabolique parmi les membres de la communauté, ce qui fait écho à certaines études affirmant l'existence de « core » microbiotes pouvant assurer certaines fonctions aussi bien que toute la communauté (Toju et al., 2018). Les souches phylogénétiquement distantes présentent un métabolisme plus diversifié et moins redondant, ce qui est souvent interprété comme un pattern limitant la compétition pour les mêmes nutriments et favorisant la coopération pour certains métabolites (Hester et al., 2019; Lam et al., 2020). Cependant, la forte concentration des SynComs à des valeurs intermédiaires suggère l'existence d'un compromis entre compétition et coopération (Machado et al., 2021), où les différentes espèces seraient suffisamment proches pour s'échanger

des métabolites dont elles ont besoin (car proximité phylogénétique implique généralement proximité fonctionnelle), et suffisamment éloignées pour éviter une forte compétition.

2.3 Prédictions d'interactions de cross-feeding pour la production de métabolites clés

Dans cette dernière analyse prédictive, le programme *Miscoto* a été utilisé pour prédire toutes les combinaisons minimales (i.e. les plus petites) de GEM capables de produire le plus de TPPM possibles, sous les différentes contraintes nutritionnelles.

Les combinaisons renvoyées par le programme étaient toutes de petite taille (2 GEM pour les contraintes modérées, 3 GEM pour les contraintes sévères). Elles étaient très nombreuses sous contraintes sévères et peu nombreuses sous contraintes modérées, et permettaient la production de plusieurs TPPM impossibles à produire par des GEM isolés. Toutefois, le gain de TPPM était mineur sous contraintes modérées. Ces patterns se maintenaient en réduisant le nombre de TPPM uniquement les acides aminés, ou bien uniquement les vitamines, ou bien uniquement les phytohormones). L'ajout d'ARE diminuait le nombre de combinaisons ainsi que la plus-value en TPPM qu'elles apportaient, sauf dans l'analyse qui considérait toutes les TPPM à la fois, où elles les augmentaient lorsque ajoutées à des milieux pauvres. (figure 5A). Cependant, le nombre total de GEM impliqués variaient fortement (de 67 à 193, en fonction de la contrainte nutritionnelle et des TPPM considérés). Certains GEM étaient très fréquents dans les combinaisons, tandis que d'autres très peu (figure 5), ce qui fait que l'ensemble des solutions peut être résumé avec une quinzaine de GEM. La taille des génomes correspondant à chaque GEM n'avait pas d'effet sur cette fréquence d'occurrence, écartant une analyse complémentaire entre opportunité de cross-feeding et comportement généraliste (généralement synonyme de gros génome) (figure 5).

Cette analyse a permis de montrer que les interactions de cross-feeding sont dépendantes du milieu, plus particulièrement des nutriments disponibles. En effet, moins de combinaisons sous de sévères contraintes nutritionnelles signifie que les GEM de la communauté sont capables de produire plus de TPPM de façon autonome, alors que de la coopération est nécessaire sous contraintes sévères. Les ARE réduisent le nombre de combinaisons possibles en rendant les souches autonomes, sauf sous les contraintes les plus sévères, où elles débloquent suffisamment de réactions pour assurer un métabolisme minimal et les interactions de cross-feeding qui en découlent. Le nombre élevé de combinaisons suggère un degré de redondance fonctionnelle au sein de la communauté, une assurance de pouvoir accomplir de nombreuses fonctions par diverses routes métaboliques pour faire face aux variations environnementales (pour les TPPM étudiées). Néanmoins, la prépondérance de certaines souches suggèrent des différences d'importance entre

souches, l'essentiel du métabolisme étudié pouvant être accompli par un « core » microbiote relativement réduit.

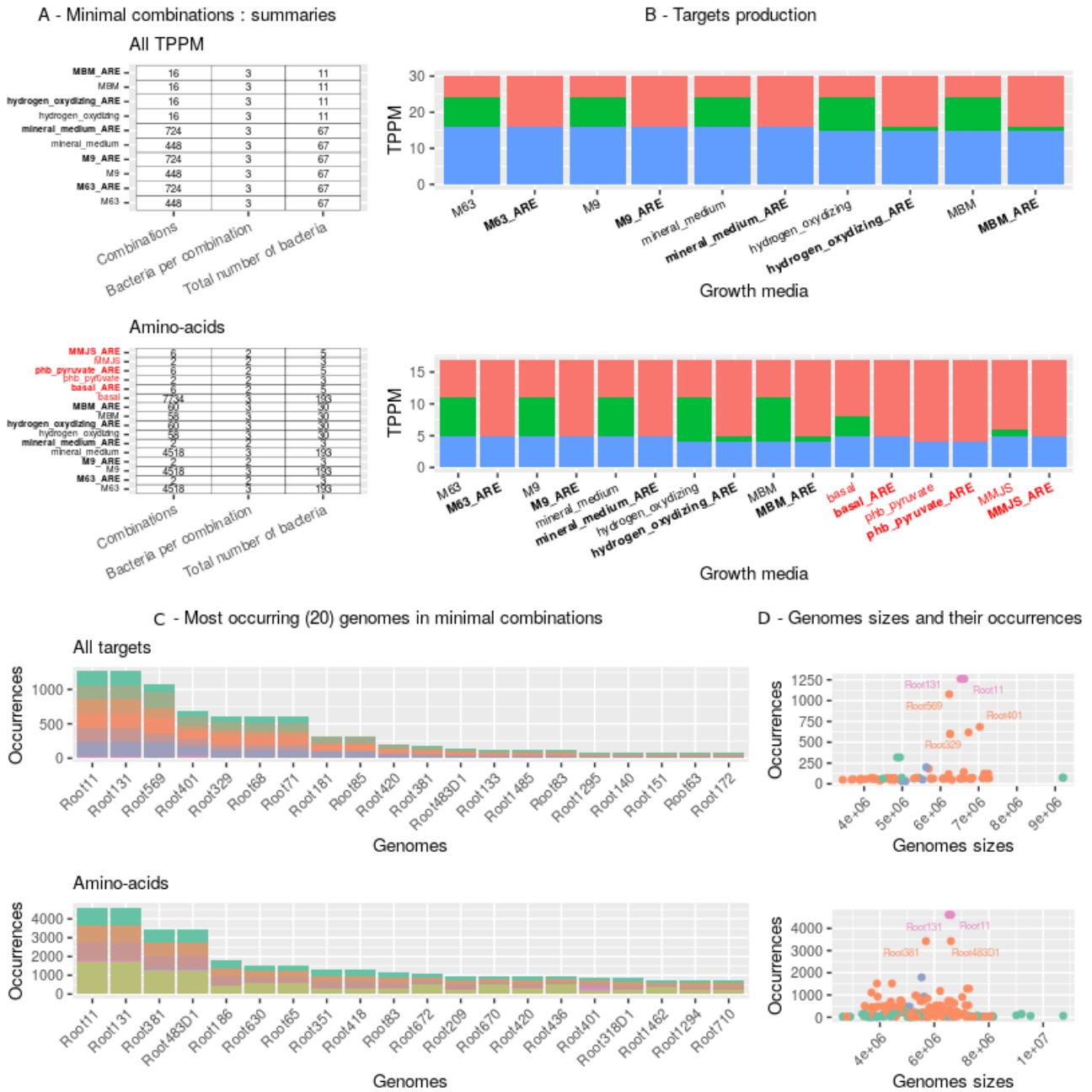


Figure 5: Combinations minimales de GEM pour produire les TPPM (toutes catégories confondues et acides aminés seulement). (A) Nombre de combinaisons, nombre de GEM par combinaison, et nombre total de GEM impliqués, selon les contraintes nutritionnelles. (B) Nombre de TPPM productibles par des GEM seuls, productibles uniquement par cross-feeding, et improductibles, selon la contrainte nutritionnelle. (C) Top 20 des GEM les plus fréquents dans les combinaisons, selon la contrainte nutritionnelle. (D) Il n'y a pas de corrélation entre la fréquence d'un GEM dans les combinaisons et la taille des génomes à partir desquels ils ont été modélisés.

2.4 Test expérimentaux basés sur les métriques prédictives

L'analyse prédictive par modélisation à partir des GEM a été suivie par une phase expérimentale. 41 communautés synthétiques (SynComs) de 4 souches ont été construites suivant un gradient de PPM. Une bactérie constante était commune à toutes les SynComs afin d'étudier plus précisément son profil d'interaction. Nous avions formulé l'hypothèse que des SynComs avec des PPM élevés (donc composées de bactéries suffisamment distantes phylogénétiquement) permettraient une meilleure co-existence des différentes bactéries, grâce à une compétition réduite (avec des niches moins chevauchantes), et potentiellement grâce à du cross-feeding.

Dans de nombreuses SynComs, la compétition pour les ressources entre souches semblait être la force principale dirigeant la dynamique des différentes abondances. En effet, la plupart renfermaient une ou deux souches dominantes, excluant les autres (par exclusion compétitive), en restant toutefois moins abondantes que dans leurs monocultures. L'analyse des abondances totales par SynComs a cependant suggéré une compétition moins forte dans les SynComs avec un PPM élevé. La souche constante s'est avéré avoir un taux de croissance très faible, en plus d'être une faible compétitrice dans toutes les SynComs, sauf une (C42), où il existait un pattern de croissance clair. Malgré la forte compétition apparente, plusieurs SynComs présentaient des indices de cross-feeding potentiel, avec des souches plus abondantes en SynComs qu'en monocultures. Une de ces souches, labellisée Root565 (*Achromobacter sp*) et présente dans trois SynComs (C15, C31 et C37), présentait l'effet le plus fort et ses cultures ont été répliquées.

Ces cultures supplémentaires ont confirmé la capacité de Root565 à mieux se développer en présence d'autres souches plutôt qu'en monoculture (figure 6A), donc à profiter des sécrétions d'autres espèces. Cette capacité ne semblait pas spécifique à certains fournisseurs, mais est apparue comme dépendante du taux de croissance du fournisseur (figure 1C, 1E et 1F) : plus le fournisseur croît vite, plus il fournit de sécrétions, et plus Root565 peut en profiter. Une culture supplémentaire de la Syncom C42 n'a pas permis de mettre en évidence de cross-feeding en faveur de la bactérie constante Root695, sa croissance étant moins forte qu'en monoculture (figure 1B). L'effet individuel de chaque souche sur la croissance de Root695 était significatif pour une seule des trois souches co-cultivées (figure 1D). Toutefois, des cultures dans des compartiments séparés, où R695 bénéficiait du flux d'exsudats des autres souches sans être en contact avec elles, ont suggéré la possibilité de cross-feeding sans besoin de contact entre bactéries, un mécanisme pouvant également limiter la compétition. Ces résultats restent cependant à répliquer et à confirmer.

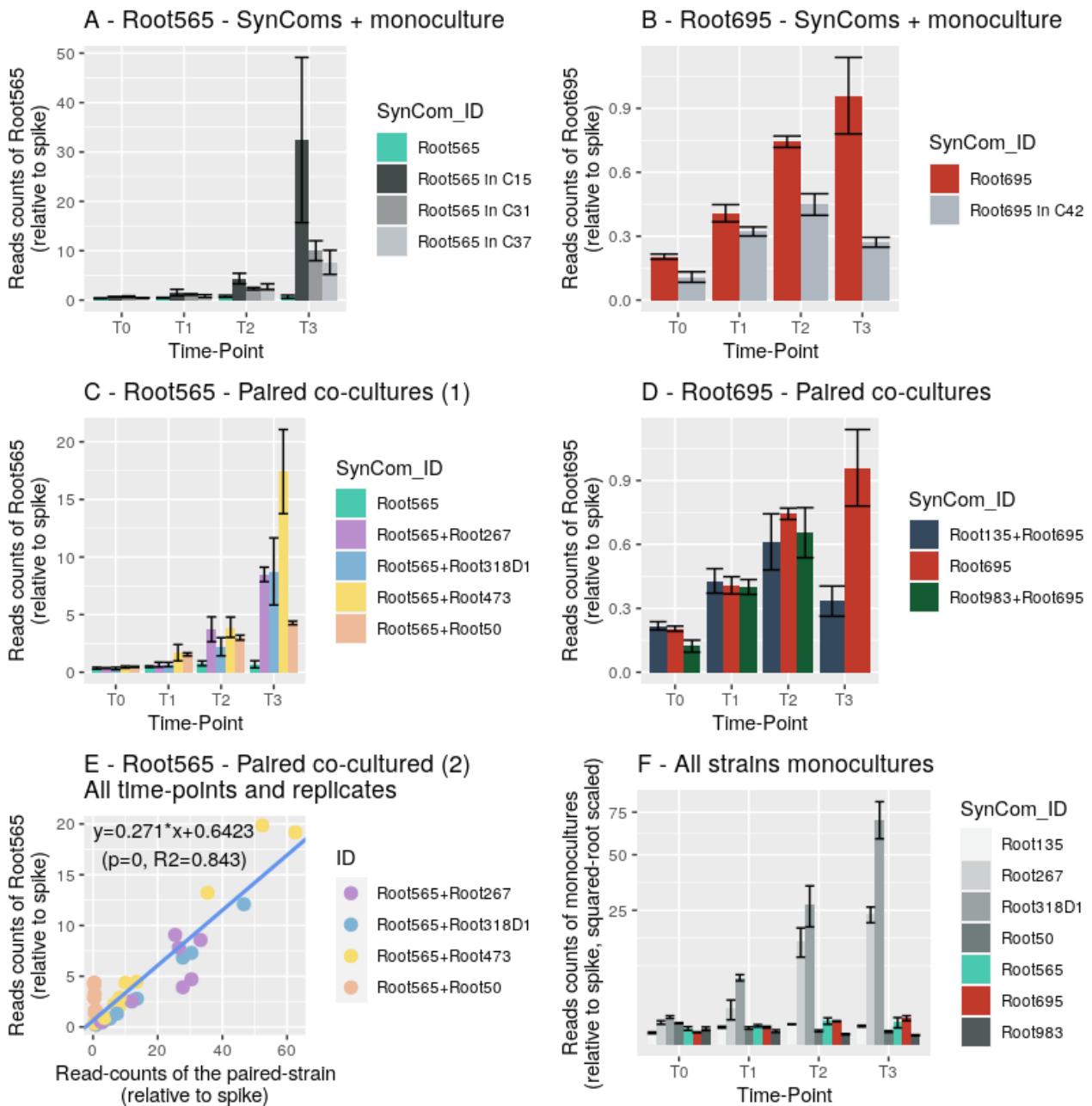


Figure 6: Abondances (en reads relatifs à la quantité de spike-in) des souches Root565 et Root695 dans 4 des 41 SynComs testées (C15, C31, C37, C42), en fonction des différents time-points. (A) Abondance de Root565 dans les SynComs C15, C31, C37 et sa monoculture. (B) Abondance de Root695 dans C42 et sa monoculture. (C) Abondances de Root565 en paires de co-cultures avec toutes les souches possibles des SynComs C15, C31, C37 (sauf Root695). (D)Abondances de Root695 en paires de co-cultures avec toutes les souches possibles de la SynCom C42. (E) Corrélation entre l'abondance de ROot565 (comme en (C)) avec l'abondance de la souche co-cultivée en paire (tous time-points et tous réplicats). F) Abondances de toutes les souches en monoculture. Root29 et Root473 ne sont pas incluses à cause de contaminations.

3. Discussion générale

Les travaux effectués pendant cette thèse ont permis de démontrer, *in silico* :

- L'effet des nutriments disponibles sur le métabolisme des bactéries, ainsi que l'importance des interactions de cross-feeding pour compenser les milieux imposant de sévères contraintes nutritionnelles. Les exsudats racinaires de la plante hôte, *A. thaliana*, jouent un rôle prépondérant pour que les différentes souches de bactéries réalisent leur métabolisme et certaines interactions, ce qui témoigne d'un certain degré d'adaptation du microbiote à son hôte.
- Ces interactions de cross-feeding sont importantes autant à l'échelle de la communauté qu'à l'échelle de petits assemblages. Cela assure une redondance fonctionnelle et permet la production de métabolites clés par diverses routes métaboliques, permettant potentiellement de faire face à un environnement fluctuant. De cette manière, de petits assemblages peuvent se rapprocher -pour les voies métaboliques étudiées-, de la capacité de la communauté tout entière.
- Une corrélation entre distance phylogénétique, compétition et coopération, déjà établi dans la littérature scientifique (Zelezniak et al., 2015; Russel et al., 2017; Lam et al., 2020) ainsi que l'existence hypothétique d'un compromis entre compétition et coopération parmi les membres du microbiote étudié (Machado et al., 2021).

Expérimentalement, il a été observé que:

- La compétition est moins forte entre les bactéries dont le métabolisme est suffisamment différent (PPM élevé, niches distinctes)
- Des interactions de cross-feeding existent -sans corrélation apparente avec les PPM- et peuvent avoir un effet important sur la croissance des souches bénéficiaires des sécrétions des autres

D'après les modèles, le cross-feeding est une interaction essentielle pour expliquer la coexistence de différentes espèces dans un microbiote, particulièrement sous certaines conditions environnementales (Widder et al., 2016; Magnúsdóttir et al., 2017; Pande and Kost, 2017). En effet, il est par exemple connu que dans un environnement très pauvre en nutriments, des bactéries habituellement compétitrices peuvent établir des interactions mutualistes pour maintenir leurs populations (Coyte and Rakoff-Nahoum, 2019). Ces résultats s'accordent avec une partie relativement importante de la littérature, où les interactions de cross-feeding sont prédits comme très fréquentes dans de nombreuses approches de modélisation, sans qu'elles soient confirmées expérimentalement (Fritts et al., 2021), la compétition étant souvent à l'œuvre (Foster and Bell, 2012). Un tel écart peut cependant s'expliquer par plusieurs hypothèses. Les métriques que nous

avons utilisées pour le protocole expérimental sont simples, et peuvent mal représenter à elles-seules les interactions de cross-feeding car elles ne permettent pas de l'isoler des autres interactions. En effet, la compétition sur d'autres axes des niches écologiques peut surpasser un cross-feeding existant, masquant sa détection (Hillesland and Stahl, 2010). Il est également possible qu'il y ait compétition entre plusieurs souches pour les métabolites secrétés (McCully et al., 2017). D'autres hypothèses avancent que le cross-feeding est une interaction principalement sous-jacente qui permettant de compenser la compétition (Goldford et al., 2018) ou de maintenir de nombreuses espèces à un niveau de métabolisme minimal, les préservant de l'extinction mais n'améliorant pas significativement leur croissance (McCully et al., 2017; Fritts et al., 2021). Finalement, la compétition elle-même pourrait être à l'origine d'interactions par cross-feeding, grâce à un mécanisme appelé « overflow metabolism ». Ce mécanisme décrit le métabolisme d'une espèce très compétitrice qui, en conséquence, à une activité métabolique importante et secrète de nombreux métabolites secondaires et déchets, dont peuvent profiter de nombreuses autres espèces (Basan et al., 2015; Carlson et al., 2018).

identifier toutes les interactions prenant place au sein d'une communauté microbienne est extrêmement difficile, particulièrement concernant les interactions d'échanges de métabolites. La recherche doit placer ses efforts dans plusieurs aspects principaux :

- L'identification du secrétome (ou examétabolome) des différents microbes, pour connaître précisément quels métabolites sont secrétés (Jacoby and Kopriva, 2019; Douglas, 2020).
- Les mécanismes de transport dans le milieu et d'assimilation par les bénéficiaires de ces mêmes métabolites (D'Souza et al., 2018).
- L'hétérogénéité du milieu, notamment spatiale, ainsi que les divers gradients qui en découlent (par exemple des gradients de concentration en nutriments) (Carlson et al., 2018; Ebrahimi et al., 2019; Gralka et al., 2020).
- La composition et l'effet des exsudats de l'organisme hôte, dans le cas de microbiotes (Monchgesang et al., 2016; Zhelnina et al., 2018).
- La qualité d'annotation des génomes (Raina et al., 2018).

Les perspectives de ces travaux sont nombreux, et le projet le plus prometteur serait de se focaliser sur l'écologie de la souche bactérienne Root565, qui a présenté d'importantes capacités à se nourrir par cross-feeding. Afin de déterminer quels secrétions sont bénéfiques à cette souche, des expériences supplémentaires sont envisageables, tel que des cultures de la souche seule avec des secrétions bactériennes en seule source de nutriments, et des modèles métaboliques plus perfectionnés, tel que des analyses de flux, sont possibles.

Bibliographie

- Aite, M., Chevallier, M., Frioux, C., Trottier, C., Got, J., Cortés, M. P., et al. (2018). Traceability, reproducibility and wiki-exploration for “à-la-carte” reconstructions of genome-scale metabolic models. *PLoS Comput. Biol.* 14, 1–25. doi:10.1371/journal.pcbi.1006146.
- Amor, D. R., and Bello, M. D. (2019). Bottom-up approaches to synthetic cooperation in microbial communities. *Life* 9. doi:10.3390/life9010022.
- Basan, M., Hui, S., Okano, H., Zhang, Z., Shen, Y., Williamson, J. R., et al. (2015). Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nature* 528, 99–104. doi:10.1038/nature15765.
- Belcour, A., Frioux, C., Aite, M., Hildebrand, F., and Siegel, A. (2020). Metage2Metabo, microbiota-scale metabolic complementarity for the identification of key species. *Elife* 9, 1–33.
- Brochet, S., Quinn, A., Mars, R. A., Neuschwander, N., Sauer, U., and Engel, P. (2021). Niche partitioning facilitates coexistence of closely related honey bee gut bacteria. *Elife* 10. doi:10.7554/elife.68583.
- Bulgarelli, D., Schlaepi, K., Spaepen, S., Van Themaat, E. V. L., and Schulze-Lefert, P. (2013). Structure and functions of the bacterial microbiota of plants. *Annu. Rev. Plant Biol.* 64, 807–838. doi:10.1146/annurev-arplant-050312-120106.
- Carlson, R. P., Beck, A. E., Phalak, P., Fields, M. W., Gedeon, T., Hanley, L., et al. (2018). Competitive resource allocation to metabolic pathways contributes to overflow metabolisms and emergent properties in cross-feeding microbial consortia. *Biochem. Soc. Trans.* 46, 269–284. doi:10.1042/BST20170242.
- Caspi, R., Foerster, H., Fulcher, C. A., Kaipa, P., Krummenacker, M., Latendresse, M., et al. (2008). The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 36, 623–631. doi:10.1093/nar/gkm900.
- Chubukov, V., Mukhopadhyay, A., Petzold, C. J., Keasling, J. D., and Martín, H. G. (2018). Synthetic and systems biology for microbial production of commodity chemicals. *npj Syst. Biol. Appl.* 2, 1–11. doi:10.1038/npjsba.2016.9.
- Coyte, K. Z., and Rakoff-Nahoum, S. (2019). Understanding Competition and Cooperation within the Mammalian Gut Microbiome. *Curr. Biol.* 29, R538–R544. doi:10.1016/j.cub.2019.04.017.
- Cui, J., Yuan, X., Zhang, Q., Zhou, J., Lin, K., Xu, J., et al. (2021). Nutrient availability is a dominant predictor of soil bacterial and fungal community composition after nitrogen addition in subtropical acidic forests. *PLoS One* 16, 1–20. doi:10.1371/journal.pone.0246263.
- D’Souza, G., Shitut, S., Preussger, D., Yousif, G., Waschina, S., and Kost, C. (2018). Ecology and evolution of metabolic cross-feeding interactions in bacteria. *Nat. Prod. Rep.* 35, 455–488. doi:10.1039/c8np00009c.
- Douglas, A. E. (2020). The microbial exometabolome: Ecological resource and architect of microbial communities. *Philos. Trans. R. Soc. B Biol. Sci.* 375. doi:10.1098/rstb.2019.0250.
- Ebrahimi, A., Schwartzman, J., and Cordero, O. X. (2019). Multicellular behaviour enables cooperation in microbial cell aggregates. *Philos. Trans. R. Soc. B Biol. Sci.* 374. doi:10.1098/rstb.2019.0077.
- Fierer, N. (2017). Embracing the unknown: Disentangling the complexities of the soil microbiome. *Nat. Rev. Microbiol.* 15, 579–590. doi:10.1038/nrmicro.2017.87.
- Foster, K. R., and Bell, T. (2012). Competition, not cooperation, dominates interactions among culturable microbial species. *Curr. Biol.* 22, 1845–1850. doi:10.1016/j.cub.2012.08.005.
- Frioux, C., Fremy, E., Trottier, C., and Siegel, A. (2018). Scalable and exhaustive screening of metabolic functions carried out by microbial consortia. *Bioinformatics* 34, i934–i943. doi:10.1093/bioinformatics/bty588.

- Fritts, R. K., McCully, A. L., and McKinlay, J. B. (2021). Extracellular Metabolism Sets the Table for Microbial Cross-Feeding. *Microbiol. Mol. Biol. Rev.* 85, 1–29. doi:10.1128/mmbr.00135-20.
- Fukami, T., Dickie, I. A., Paula Wilkie, J., Paulus, B. C., Park, D., Roberts, A., et al. (2010). Assembly history dictates ecosystem functioning: Evidence from wood decomposer communities. *Ecol. Lett.* 13, 675–684. doi:10.1111/j.1461-0248.2010.01465.x.
- Goldford, J. E., Lu, N., Bajić, D., Estrela, S., Tikhonov, M., Sanchez-Gorostiaga, A., et al. (2018). Emergent simplicity in microbial community assembly. *Science* (80-.). 361, 469–474. doi:10.1126/science.aat1168.
- Götzenberger, L., Bello, F. De, Brathen, K. A., Davison, J., Dubuis, A., Guisan, A., et al. (2012). Ecological assembly rules in plant communities — approaches , patterns and prospects. *Biol. Rev.* 87, 111–127. doi:10.1111/j.1469-185X.2011.00187.x.
- Goyal, A., Wang, T., Dubinkina, V., and Maslov, S. (2021). Ecology-guided prediction of cross-feeding interactions in the human gut microbiome. *Nat. Commun.* 12, 1–10. doi:10.1038/s41467-021-21586-6.
- Gralka, M., Szabo, R., Stocker, R., and Cordero, O. X. (2020). Trophic Interactions and the Drivers of Microbial Community Assembly. *Curr. Biol.* 30, R1176–R1188. doi:10.1016/j.cub.2020.08.007.
- Griggs, R. G., Steenwerth, K. L., Mills, D. A., and Bokulich, N. A. (2021). Sources and Assembly of Microbial Communities in Vineyards as a Functional Component of Winegrowing. 12. doi:10.3389/fmicb.2021.673810.
- Hacquard, S. (2016). Disentangling the factors shaping microbiota composition across the plant holobiont. *New Phytol.* 209, 454–457. doi:10.1111/nph.13760.
- Hester, E. R., Jetten, M. S. M., Welte, C. U., and Lücker, S. (2019). Metabolic overlap in environmentally diverse microbial communities. *Front. Genet.* 10, 1–10. doi:10.3389/fgene.2019.00989.
- Hillesland, K. L. (2018). Evolution on the bright side of life: microorganisms and the evolution of mutualism. *Ann. N. Y. Acad. Sci.* 1422, 88–103. doi:10.1111/nyas.13515.
- Hillesland, K. L., and Stahl, D. A. (2010). Rapid evolution of stability and productivity at the origin of a microbial mutualism. *Proc. Natl. Acad. Sci. U. S. A.* 107, 2124–2129. doi:10.1073/pnas.0908456107.
- Jacoby, R. P., and Kopriva, S. (2019). Metabolic niches in the rhizosphere microbiome: New tools and approaches to analyse metabolic mechanisms of plant-microbe nutrient exchange. *J. Exp. Bot.* 70, 1087–1094. doi:10.1093/jxb/ery438.
- Jansma, J., and Aidy, S. El (2020). Understanding the host-microbe interactions using metabolic modeling. *bioRxiv*, 1–14. doi:10.1101/2020.06.12.147918.
- Karp, P. D., Latendresse, M., and Caspi, R. (2011). The Pathway Tools Pathway Prediction Algorithm. 424–429. doi:10.4056/sigs.1794338.
- Kelsic, E. D., Zhao, J., Vetsigian, K., and Kishony, R. (2015). Counteraction of antibiotic production and degradation stabilizes microbial communities. *Nature* 521, 516–519. doi:10.1038/nature14485.
- Kumar, M., Ji, B., Zengler, K., and Nielsen, J. (2019). Modelling approaches for studying the microbiome. *Nat. Microbiol.* 4, 1253–1267. doi:10.1038/s41564-019-0491-9.
- Lam, T. J., Stamboulian, M., Han, W., and Ye, Y. (2020). Model-based and phylogenetically adjusted quantification of metabolic interaction between microbial species. *PLoS Comput. Biol.* 16, 1–17. doi:10.1371/journal.pcbi.1007951.
- Larimer, A. L., Bever, J. D., and Clay, K. (2010). The interactive effects of plant microbial symbionts: A review and meta-analysis. *Symbiosis* 51, 139–148. doi:10.1007/s13199-010-0083-1.
- Lawson, C. E., Harcombe, W. R., Hatzenpichler, R., Lindemann, S. R., Löffler, F. E., O’Malley, M. A., et al. (2019). Common principles and best practices for engineering microbiomes. *Nat. Rev. Microbiol.* 17, 725–741. doi:10.1038/s41579-019-0255-9.

- Liu, X., Ding, W., and Jiang, H. (2017). Engineering microbial cell factories for the production of plant natural products: From design principles to industrial-scale production. *Microb. Cell Fact.* 16, 1–9. doi:10.1186/s12934-017-0732-7.
- Liu, Y., and Nielsen, J. (2019). Recent trends in metabolic engineering of microbial chemical factories. *Curr. Opin. Biotechnol.* 60, 188–197. doi:10.1016/j.copbio.2019.05.010.
- Machado, D., Maistrenko, O. M., Andrejev, S., Kim, Y., Bork, P., Patil, K. R., et al. (2021). Polarization of microbial communities between competitive and cooperative metabolism. *Nat. Ecol. Evol.* 5, 195–203. doi:10.1038/s41559-020-01353-4.
- Magnúsdóttir, S., Heinken, A., Kutt, L., Ravcheev, D. A., Bauer, E., Noronha, A., et al. (2017). Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* 35, 81–89. doi:10.1038/nbt.3703.
- McCully, A. L., LaSarre, B., and McKinlay, J. B. (2017). Growth-independent cross-feeding modifies boundaries for coexistence in a bacterial mutualism. *Environ. Microbiol.* 19, 3538–3550. doi:10.1111/1462-2920.13847.
- Mee, M. T., Collins, J. J., Church, G. M., and Wang, H. H. (2014). Syntrophic exchange in synthetic microbial communities. 2149–2156. doi:10.1073/pnas.1405641111.
- Mendoza, S. N., Olivier, B. G., Molenaar, D., and Teusink, B. (2019). A Systematic Assessment Of Current Genome-Scale Metabolic Reconstruction Tools. *Genome Biol.* 20, 1–20. doi:10.1101/558411.
- Monchgesang, S., Strehmel, N., Schmidt, S., Westphal, L., Taruttis, F., Muller, E., et al. (2016). Natural variation of root exudates in *Arabidopsis thaliana*-linking metabolomic and genomic data. *Sci. Rep.* 6, 1–11. doi:10.1038/srep29033.
- Morris, B. E. L., Henneberger, R., Huber, H., and Moissl-Eichinger, C. (2013). Microbial syntrophy: Interaction for the common good. *FEMS Microbiol. Rev.* 37, 384–406. doi:10.1111/1574-6976.12019.
- Morrison-Whittle, P., and Goddard, M. R. (2015). Quantifying the relative roles of selective and neutral processes in defining eukaryotic microbial communities. *ISME J.* 9, 2003–2011. doi:10.1038/ismej.2015.18.
- Muller, E. E. L., Faust, K., Widder, S., Herold, M., Martínez Arbas, S., and Wilmes, P. (2018). Using metabolic networks to resolve ecological properties of microbiomes. *Curr. Opin. Syst. Biol.* 8, 73–80. doi:10.1016/j.coisb.2017.12.004.
- Pacheco, A. R., and Segrè, D. (2019). A multidimensional perspective on microbial interactions. *FEMS Microbiol. Lett.* 366, 1–11. doi:10.1093/femsle/fnz125.
- Pande, S., and Kost, C. (2017). Bacterial Unculturability and the Formation of Intercellular Metabolic Networks. *Trends Microbiol.* 25, 349–361. doi:10.1016/j.tim.2017.02.015.
- Raina, J. B., Eme, L., Pollock, F. J., Spang, A., Archibald, J. M., and Williams, T. A. (2018). Symbiosis in the microbial world: from ecology to genome evolution. *Biol. Open* 7, 5–8. doi:10.1242/bio.032524.
- Russel, J., Røder, H. L., Madsen, J. S., Burmølle, M., and Sørensen, S. J. (2017). Antagonism correlates with metabolic similarity in diverse bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 114, 10684–10688. doi:10.1073/pnas.1706016114.
- Smith, N. W., Shorten, P. R., Altermann, E., Roy, N. C., and McNabb, W. C. (2019). The Classification and Evolution of Bacterial Cross-Feeding. *Front. Ecol. Evol.* 7, 1–15. doi:10.3389/fevo.2019.00153.
- Sung, J., Kim, S., Cabatbat, J. J. T., Jang, S., Jin, Y. S., Jung, G. Y., et al. (2017). Global metabolic interaction network of the human gut microbiota for context-specific community-scale analysis. *Nat. Commun.* 8, 1–12. doi:10.1038/ncomms15393.
- Thommes, M., Wang, T., Zhao, Q., Paschalidis, I. C., and Segrè, D. (2019). Designing Metabolic Division of Labor in Microbial Communities. *mSystems* 4. doi:10.1128/msystems.00263-18.

- Toju, H., Peay, K. G., Yamamichi, M., Narisawa, K., and Hiruma, K. (2018). Core microbiomes for sustainable agroecosystems. *Nat. Plants* 4. doi:10.1038/s41477-018-0139-4.
- Vellend, M. (2010). Conceptual synthesis in community ecology. *Q. Rev. Biol.* 85, 183–206. doi:10.1086/652373.
- Walker, R. W., Clemente, J. C., Peter, I., and Loos, R. J. F. (2017). The prenatal gut microbiome: are we colonized with bacteria in utero? *Pediatr. Obes.* 12, 3–17. doi:10.1111/ijpo.12217.
- Widder, S., Allen, R. J., Pfeiffer, T., Curtis, T. P., Wiuf, C., Sloan, W. T., et al. (2016). Challenges in microbial ecology: Building predictive understanding of community function and dynamics. *ISME J.* 10, 2557–2568. doi:10.1038/ismej.2016.45.
- Zélé, F., Magalhães, S., Kéfi, S., and Duncan, A. B. (2018). Ecology and evolution of facilitation among symbionts. *Nat. Commun.* 9, 1–12. doi:10.1038/s41467-018-06779-w.
- Zeleznak, A., Andrejev, S., Ponomarova, O., Mende, D. R., Bork, P., and Patil, K. R. (2015). Metabolic dependencies drive species co-occurrence in diverse microbial communities. 112. doi:10.1073/pnas.1522642113.
- Zengler, K., and Zaramela, L. S. (2018). The social network of microorganisms - How auxotrophies shape complex communities. *Nat. Rev. Microbiol.* 16, 383–390. doi:10.1038/s41579-018-0004-5.
- Zhelnina, K., Zengler, K., Newman, D., and Northen, T. R. (2018). Need for Laboratory Ecosystems To Unravel the Structures. *MBio* 9, 1–8.
- Zhang, H., Zhu, S., John, R., Li, R., Liu, H., and Ye, Q. (2017). Habitat filtering and exclusion of weak competitors jointly explain fern species assemblage along a light and water gradient. *Sci. Rep.* 7, 1–10. doi:10.1038/s41598-017-00429-9.
- Zuñiga, C., Li, C. T., Yu, G., Al-Bassam, M. M., Li, T., Jiang, L., et al. (2019). Environmental stimuli drive a transition from cooperation to competition in synthetic phototrophic communities. *Nat. Microbiol.* 4, 2184–2191. doi:10.1038/s41564-019-0567-6.

Titre : Modélisation, prédition, et test expérimental d'interactions de cross-feeding au sein du microbiote d'*Arabidopsis thaliana*: une approche par microbial system ecology.

Mots clés : Microbiote, cross-feeding, Microbial system ecology

Résumé : les communautés microbiennes forment un réseau complexe d'interactions entre organismes, qui façonnent leur structure. Parmi ces interactions, les échanges de métabolites entre organismes, nommés 'cross-feeding', sont considérés comme fréquents et importants. Ces échanges peuvent être prédits *in silico* avec des réseaux métaboliques inférés à partir de génomes. Ces modèles sont intégrés dans un cadre de travail nommé 'microbial system ecology', qui rassemble plusieurs méthodes de modélisation ainsi que leur validation expérimentale, à différentes échelles d'étude. Dans cette thèse, nous avons utilisé ce cadre pour prédire de nombreuses et hypothétiques associations de bactéries du microbiote racinaire d'*Arabidopsis thaliana* permettant le production de métabolites non productibles par des bactéries seules. Nous avons également modélisé l'impact des nutriments disponibles dans le milieu, et prédit que ces

contraintes nutritionnelles sont largement compensées par des comportements de cross-feeding. En parallèle, nous avons également utilisé des métriques simples pour corrélérer le potentiel de cross-feeding et de compétition avec le métabolisme des bactéries et leur distance phylogénétique. Ces métriques ont été testées expérimentalement avec un ensemble de petites communautés synthétiques. La compétition semblait dominer dans la plupart des communautés, mais a été déterminée comme étant plus faible dans les communautés abritant des souches à métabolisme différent. Nous avons identifié quelques bactéries profitant des exsudats d'autres espèces, particulièrement une *Achromobacter* sp. Finalement, les résultats suggèrent que le cross-feeding est courant mais potentiellement masqué par de la forte compétition lorsque des bactéries sont en contact.

Title : Genome-scale modelling, prediction, and experimental testing of cross-feeding interactions within the root microbiota of *Arabidopsis thaliana* : a microbial system ecology framework.

Keywords : Microbiota, cross-feeding, Microbial system ecology

Abstract : Microbial communities form complex, intricate networks of interacting organisms that highly participate in their structure and dynamics. Among these interactions, the exchange of metabolites between organisms, i.e. cross-feeding, is more and more considered as highly frequent and important. Cross-feeding behaviors can be investigated *in silico* with reconstructed metabolic networks, that allow modelling of organisms' metabolism, hence predictions of possibly exchanged metabolites. These modelling approaches take place in a particular framework, microbial system ecology, which organizes numerous modelling approaches and their experimental validation, from bottom-up to top-down scales. In this work, we used such a framework on a representative part of the root microbiota of *Arabidopsis thaliana* to carry out numerous and putative combinations of bacterial strains that would allow the production of

metabolites that were not producible by single strains. We also modelled the impact of available nutrients on strains metabolism, and predicted that cross-feeding can largely compensate severe nutritional constraints. In parallel, we used simple metrics correlating the capacity to exchange metabolites with bacteria's metabolism and phylogenetic distance. Such metrics were tested experimentally, with a number of small synthetic communities. Competition seemed to dominate most of the communities, but with a weaker effect in SynComs displaying metabolically divergent bacteria. We identified a few bacteria apparently able to feed on other's exudates; notably an *Achromobacter* sp. Finally, results suggest that cross-feeding is frequent, but potentially masked by strong, contact-dependent competition.