Université du Québec Institut National de la Recherche Scientifique Centre Institut Armand-Frappier

RÔLE DES TRAITS PHYSIOLOGIQUES, DE L'ABONDANCE RELATIVE ET DE LA TAXONOMIE DES ESPÈCES DANS LA COVARIATION DES BACTÉRIES

Par

Asma Jaba

Mémoire présenté pour l'obtention du grade de Maître ès sciences (M.Sc.) Microbiologie Appliquée

Jury d'évaluation

Président du jury et examinateur interne

Isabelle Dusfour INRS – Institut Armand-Frappier / Institut Pasteur

Timothée Poisot Université de Montréal

Philippe Constant INRS – Institut Armand-Frappier

Claude Guertin INRS – Institut Armand-Frappier

Examinateur externe

Directeur de recherche

Codirecteur de recherche

© Droits réservés de *(Asma Jaba)*, 2018

RÉSUMÉ

Une meilleure connaissance des règles impliquées dans le processus de covariation des communautés microbiennes pourrait mener vers de nouvelles solutions pour le secteur agroalimentaire en identifiant les facteurs de prédisposition à la contamination microbiologique. Dans cette optique, nous avons testé un modèle théorique qui prétend que la covariation des microorganismes est sous le contrôle de la taxonomie, l'abondance relative et les traits physiologiques de chaque espèce. Pour répondre à cette hypothèse, un séquençage de marqueurs taxonomiques fut appliqué à une série d'échantillons de chia pour répertorier les espèces de bactéries et de champignons présentes, ainsi que leur profil de covariation. Ces efforts de caractérisation ont été complémenté par l'isolement de 71 bactéries, dont cinq souches appartenant aux alpha- et aux gamma-Proteobacteria correspondaient à des espèces représentées dans le portrait moléculaire des communautés microbiennes. Le profil de tolérance à différents stress physiologiques et inhibiteurs de croissance fut obtenu pour établir un portrait des traits physiologiques propre à chacun de ces derniers isolats. L'application d'une série de régression linéaires a permis de démontrer que la divergence des traits physiologiques et l'abondance de chacune des paires d'espèces expliquaient 69 % de leur covariation observée dans les échantillons de chia. En validant ainsi l'hypothèse avancée, cette étude constitue une première initiative démontrant un potentiel à exploiter pour élaborer des nouveaux outils de diagnostic précoces de contamination microbiologique des aliments. L'application de l'approche utilisée dans le cadre de ce projet à d'autres aliments sera cependant nécessaire pour en démontrer la fiabilité et l'applicabilité.

Mots-clés : Covariation, taxonomie, abondance relative, traits physiologiques, régression linéaire

Étudiante

Directeur de recherche

ABSTRACT

Deciphering the rules defining microbial community assemblage is envisioned as a promising strategy to improve predictions of pathogens colonization and proliferation in food. Despite the increasing number of studies reporting microbial co-occurrence patterns, only a few attempts were made to challenge them in experimental or theoretical frameworks. Here, we tested the hypothesis that observed co-occurrence patterns could be explained by taxonomy, relative abundance and physiological traits of microbial species. PCR amplicon sequencing of taxonomic markers was first conducted to assess distribution and co-occurrence patterns of bacterial and fungal species found in 25 chia samples originating from eight different sources. The use of nutrient-rich and oligotrophic media enabled isolation of 71 strains encompassing 16 bacterial species, of which five corresponded to phylotypes represented in the molecular survey. Tolerance to different growth inhibitors and antibiotics was tested to assess physiological traits of these isolates. Multiple linear regression analyses demonstrated that divergence of physiological traits and relative abundance of each pair of species explained 69% of the co-occurrence profile displayed by cultivable bacterial phylotypes in chia. Despite the significance of this result for the conceptualization of microbial networks, application of the approach to more food products will be necessary to demonstrate reliability of the proposed theoretical framework to explain co-occurrence patterns in food.

Keywords: Microbial ecology, microbiome, microbial communities...

REMERCIEMENTS

Je consacre cette partie de mon travail à exprimer ma reconnaissance à toutes les personnes qui m'ont aidée de près ou de loin dans la réalisation de mon projet de recherche. D'ailleurs, j'ai beaucoup appris grâce à cette première expérience loin de mon pays. Je tiens tout d'abord à remercier mon directeur principal, le Pr. Philippe Constant, de m'avoir accueilli dans son laboratoire et pour tout ce qu'il m'a appris au cours de ma maîtrise.

Pr Philippe Constant, il n'y a pas de mots pour décrire ma gratitude envers toi, ton aide, ton appui, ton support aussi bien scientifique que moral, et pour le leadership que tu as montré tout au long de ma maîtrise sans oublier surtout ton sacré sens d'humour et ton extrême gentillesse. Sincèrement, je n'ai pas de mot pour t'exprimer ma profonde reconnaissance.

Pr Claude Guertin, je te suis profondément reconnaissante pour tes précieux conseils et pour l'énorme contribution que tu as apportée afin de mener à bien ce travail.

Ces remerciements s'adressent à notre belle équipe de laboratoire, Anne, Audrey-Anne, Josiann, Judith, mon ami Julien, Mondher, Narin, Sarah et Thibaut, Merci pour vos encouragements et pour votre collaboration. Je ferai un agréable devoir pour remercier Karla Vazquez pour son appui durant ma première période de manipulation au laboratoire, aussi je remercie toutes mes camarades de bureau, mon amie Valérie, Alexandra, Livie et Rose.

Quant aux membres de ma famille, merci d'avoir cru en moi en tout moment, je vous adore, Je remercie ainsi mes amis, sans citer vos noms, vous savez qui vous êtes (ma deuxième famille), je tiens à vous remercier pour votre soutien, et pour m'avoir fait passer des moments agréables.

Enfin, je tiens à remercier les membres du jury pour avoir consacrer une partie de leur temps à examiner et à juger ce travail, en me partageant une partie de leur savoir à travers leurs critiques.

iii

TABLE DES MATIÈRES

RÉSUMÉ	i
ABSTRACT	ii
REMERCIEM	IENTSiii
TABLE DES	MATIÈRESiv
LISTE DES 1	rableauxvi
LISTE DES F	FIGURES vii
LISTE DES É	ÉQUATIONS viii
	ABRÉVIATIONSix
SYNTHÈSE.	
CHAPITRE 1	l : Introduction2
1.1 Mise	e en contexte2
1.2 Port	trait des communautés microbiennes
1.3 Inte	ractions microbe-microbe6
1.3.1	Algorithme adaptés aux données compositionnelles
1.3.2	Limites des réseaux de covariations
1.4 Thé	orie expliquant la structure des réseaux écologiques
1.4.1	Covariation des espèces expliquée par la taxonomie
1.4.2	Covariation des espèces expliquée par la sélection de traits
1.5 Pres	ssions multifactorielles modulant la covariation des espèces
1.6 Éco	logie appliquée au secteur alimentaire12
1.6.1	Discrimination moléculaire des souches pathogènes et non pathogènes 13
1.6.2	Réseaux de covariance dans l'écologie alimentaire14
1.7 Нур	othèse15
1.7.1	Objectifs

ARTICL	.E		17				
СНАРІТ	RE 2	2 : Article	18				
2.1	Contribution à l'article						
2.2	Rés	Résumé19					
2.3	Abs	stract	20				
2.4	Intro	oduction	21				
2.5	Mat	erial and methods	23				
2.5	5.1	Samples	23				
2.5	5.2	Genomic DNA extraction	23				
2.5	5.3	Microbial community structure	24				
2.5	5.4	Isolation of bacteria	26				
2.5	5.5	Molecular identification and classification of bacterial isolates	26				
2.5	5.6	Physiological traits of bacterial isolates	27				
2.5	5.7	Statistical analyses	28				
2.6	Res	sults	30				
2.6	6.1	Microbial community structure and covariation between microbial spe 30	cies				
2.6	6.2	Isolation of bacteria associated with chia	34				
2.6	6.3	Model framework for co-occurrence of microbial species	38				
2.7	Dise	cussion	42				
2.8	Ack	nowledgements	45				
CONCL	USIC	ON ET PERSPECTIVES	46				
REFER	ENC	ES	48				

LISTE DES TABLEAUX

Table 1. General information related to molecular profile of bacterial and fungal	
communities associated with chia samples	31
Table 2. Multiple regressions showing the relationship of species co-occurrence $(A_{(i,j)})$	
with physiological traits ($\tau_{(i,j)}$) and relative abundance ($N_{(i,j)}$)	41

LISTE DES FIGURES

Figure 1.1 Les interactions biologiques possibles entre les membres de différentes	
espèces	7
Figure 1.2 (a) Réseau de covariance (association significative) entre les espèces	
microbiennes de microbiome de Brassica (b) zoom sur le réseau (c) Classification	
taxonomique de microbiome de Brassica1	5
Figure 1. Principal coordinate analysis of bacterial	2
Figure 2. Ecological network of ubiquitous bacterial and fungal OTU in chia	3
Figure 3. Maximum likelihood phylogenetic tree of 16S rRNA gene sequence of isolates	3
and OTU retrieved from the molecular survey	5
Figure 4. Evaluation of the isolation efforts and taxonomic distribution of the isolates3	6
Figure 5. Physiological traits of the five selected bacterial isolates	8
Figure 6. Theoretical framework to explain co-occurrence patterns of the five selected	
bacterial isolates in chia	9

LISTE DES ÉQUATIONS

$A(i,j) = \tau(i,j) + PD(i,j) + N(i,j)$	(equation 1)	29
$N(i, j) = n = 125xin = 125xj \times (n = 12)$	25xi + n = 125xj (e)	quation 2) 29

LISTE DES ABRÉVIATIONS

- H₂S sulfure d'hydrogène
- ARNr ARN ribosomique
- PCR Réaction en chaîne par polymérase
- ratio-log rapport de log
- UFC Unité Formant Colonie
- ITS Internal transcribed spacer
- ASV Amplicon sequence variant
- OTU Unité taxonomique opérationnelle (Operational Taxonomic Unit)
- RDP Ribosomal Database Project

SYNTHÈSE

CHAPITRE 1 : INTRODUCTION

1.1 Mise en contexte

Les microorganismes représentent le plus grand réservoir de biodiversité sur terre, avec 10¹² espèces de bactéries, archées et champignons microscopiques selon les derniers modèles mathématiques (Locey & Lennon, 2016). Dotés d'une flexibilité métabolique remarquable, ces microorganismes sont omniprésents dans tous les compartiments de la biosphère, occupant toutes les niches leur conférant les conditions nécessaires à leur survie et leur croissance. Par exemple les bactéries présentent une biomasse énorme estimée à 5X10³⁰ cellules sur la planète (Fuhrman, 2009; Whitman *et al.*, 1998). Les fonctionnalités assurées par le monde microbien modulent le cycle biogéochimique des éléments essentiels à la vie, supportant ainsi une pléiade de services écosystémiques (Dauga *et al.*, 2005).

Un des défis auxquels les microbiologistes actuels font face est de comprendre les aspects biologiques et physicochimiques du milieu qui gouvernent la colonisation et la prolifération des microorganismes (Blaser et al., 2016). Le succès de cette mission repose en grande partie sur l'écologie microbienne; une discipline qui vise une meilleure compréhension de la réponse des microorganismes (i.e. bactéries, archées et suite à à eucarvotes microscopiques) une exposition différents stimuli environnementaux, incluant les interactions interspécifiques et des mécanismes supportant le maintien de leur diversité (Xu, 2006). Au même titre que les déterminismes abiotiques tels que la température et le pH, il est de plus en plus évident que la composition des communautés microbiennes joue un rôle important définissant l'hospitalité des niches écologiques. L'assemblage des communautés microbiennes est effectivement un processus complexe impliquant entre autres l'échange de métabolites comme des précurseurs de vitamines, des acides aminés ou des enzymes ou des molécules signal (Foster & Bell, 2012; Little et al., 2008). Ces interactions peuvent être essentielles à la survie des espèces impliquées, tel qu'observé dans le cas des relations syntrophiques entre les bactéries fermentaires et les archées méthanogènes

(Schink, 2002) et dans le cas tapis microbiens où les bactéries réductrices des sulfates anaérobies transfèrent les produits intermédiaires réduits de leur métabolisme (H₂S) aux bactéries oxydant les sulfures qui maintiennent en échange une barrière contre une invasion de l'O₂ (Fukui et al., 1999). L'importance des interactions microbiennes se manifeste également dans les fonctions des microorganismes. Une étude récente a démontré que l'activité d'oxydation du méthane d'une bactérie méthanotrophe est stimulée par la présence d'autres bactéries incapables d'utiliser ce substrat (Ho et al., 2014b). Le mécanisme exact de cette coopération n'est toujours par défini, mais il impliquerait vraisemblablement un échange de molécules volatiles (Veraart et al., 2018). Il apparaît maintenant évident que la compréhension des règles d'assemblage des communautés microbiennes pourrait contribuer à déchiffrer le fonctionnement des écosystèmes, maintenir leur diversité et plus spécifiquement permettre une prise de contrôle du monde microbien dans le but d'actionner des métabolismes soutenant des services écosystémiques variés. Un exemple éloquent de ce champ d'application est en plein essor dans le domaine de l'agriculture où l'impact des pratiques de gestion des terres sur la réponse des microorganismes du sol est de plus en plus considéré comme critère de prise de décision. On cherche par exemple à encourager les pratiques qui supportent le maintien d'une forte biodiversité qui confèrent aux agroécosystèmes de meilleures performances environnementale et économigues (Hoy, 2015). Le progrès fulgurant des technologies de séquençage et des outils bioinformatiques constitue un levier important pour parvenir à ces fins, mais l'étude des communautés microbiennes en demeure pour le moins complexe en raison de la riche diversité, la redondance fonctionnelle et l'hétérogénéité intra- et interspécifique du paysage génomique du monde microbien.

1.2 Portrait des communautés microbiennes

Il existe deux principales stratégies pour étudier les communautés microbiennes, soient l'approche traditionnelle de mise en culture et les méthodes moléculaires. Dans le premier cas, l'isolement des microorganismes de l'environnement implique la formulation de milieux de culture synthétiques spécifiques permettant leur croissance.

En plus des besoins particuliers de certaines espèces, la croissance rapide des microorganismes copiotrophes (croissance rapide supportée par une forte capacité à utiliser les substrats biodisponibles abondants) limite la prolifération des oligotrophes (croissance lente favorisée lorsque les éléments nutritifs sont limitants vivant dans des milieux faibles en nutriments) qui peuvent alors être moins bien représentées dans les collections d'isolats (Koch, 2001). Les limites de ces approches traditionnelles sont constatées lorsque la diversité des collections de souches isolées est comparée à la diversité totale présente dans les habitats naturels complexes tel que le sol et l'eau (Dees *et al.*, 2014). Par exemple, les efforts d'isolement de microorganismes dans le sol a pu couvrir que 1 à 50% des espèces totales détectées par des approches moléculaires (Choi *et al.*, 2017). Cependant, il faut mentionner que l'utilisation de la microbiologie en fournissant des outils d'identification et de classification microbienne basés sur la détermination de caractéristiques phénotypiques et physiologiques.

Afin de contrer les limites de l'approche traditionnelle, de nouvelles technologies exploitant la biologie moléculaire émergent et sont de plus en plus puissantes d'une génération à une autre. Cette percée est en grande partie attribuable au professeur Norman Pace qui fut le premier à mettre en évidence la diversité des microorganismes de l'environnement sans les cultiver par une amplification du gène codant pour l'ARNr 5S par PCR (Stahl et al., 1985). En effet, suivant cette approche, l'ADN génomique ne provient pas d'un seul isolat microbien, mais plutôt de l'ADN total extrait à partir d'un échantillon environnemental représentant ainsi toute une communauté microbienne en fonction de la région ou du gène amplifié par PCR et séquencé. L'idée d'amplifier par PCR un gène présent chez tous les microorganismes et dont la séquence permettait d'inférer leur évolution et leur taxonomie demeure l'approche de choix pour étudier les communautés microbiennes de l'environnement. Actuellement, on remarque des progrès de cette approche qui se traduisent par une amélioration de la résolution des résultats. Les dernières innovations entourant le séquençage à haut débit permettent d'obtenir, à faible coût, un profil moléculaire relativement fidèle des communautés microbiennes. De ce fait, le séquençage à haut débit a permis aux recherches en écologie microbienne d'avancer et d'approfondir la connaissance quant à la diversité du

monde microscopique échappant aux efforts d'isolement. Le séquençage des marqueurs taxonomiques, comme l'ARNr 16S chez les bactéries, est maintenant intégré au processus de caractérisation des communautés et fait ressortir la présence de nouvelles souches de bactéries, d'archées et de champignons pour lesquelles il n'existe aucune correspondance dans les banques connues de référence des profils phénotypiques. L'accessibilité croissante du séquençage à haut débit provoque actuellement un véritable tsunami de profilage moléculaire donnant un recensement des microorganismes associés à divers organismes hôtes, incluant les plantes, les invertébrés et les vertébrés. Ceci relève la notion du « microbiome » qui représente l'ensemble de génomes de microorganismes occupant un environnement donné (Lederberg & McCray, 2001; Toju *et al.*, 2018). La structure du microbiome est dynamique puisqu'il est constitué de microorganismes ubiquitaires ou/et transitoires procurant une certaine diversité au sein des communautés microbiennes retrouvées dans des habitats homologues et hétérologues.

Bien que les avances technologiques moléculaires aient permis de détecter un profil des communautés microbiennes, ce dernier apporte peu d'informations quant au potentiel physiologique des espèces présentes dans le milieu étudié (Aßhauer et al., 2015). Ces limites tendent à être minimisées par l'émergence des approches métagénomiques où l'ADN total fragmenté, au lieu d'un produit de PCR, est séguencé et annoté. Suivant ces techniques, il est possible d'assembler des génomes partiels et d'inférer le potentiel métabolique des communautés microbiennes. Paradoxalement, le développement de ces technologies sophistiquées démontre l'importance de ne pas négliger les efforts de domestication et des microorganismes de l'environnement puisqu'une bonne proportion (environ 50 %) des séquences obtenues par métagénomique codent pour des gènes hypothétiques dont les fonctions ne sont pas connues dans les bases de données génomiques de référence. Dans une optique d'accéder à un profil plus représentatif des communautés microbiennes en termes de taxonomie et de potentiel métabolique, il devient primordial de combiner les efforts d'isolement aux technologies moléculaires. Dans un contexte de recherche des règles d'assemblage des communautés microbiennes, l'obtention de collections d'isolats permettra entre autres de tester si la convergence ou la divergence du profil de

distribution des espèces dans l'environnement sont gouvernées par des interactions qui pourraient être mises en évidence par des expériences en co-cultures.

1.3 Interactions microbe-microbe

Contrairement à la majorité des travaux dans le domaine de la microbiologie classique basés sur des observations à partir de cultures axéniques, l'étude des processus l'assemblage des communautés microbienne doit se faire en tenant compte que les microorganismes forment des réseaux d'interactions biologiques. Ces interactions peuvent avoir un impact positif, un impact négatif ou aucun impact sur les espèces impliquées dans le réseau écologique définissant la structure des communautés microbiennes (Faust & Raes, 2012) (Figure 1). La définition de modèles mathématiques expliquant les profils d'abondance des espèces a fait l'objet de plusieurs études classiques dans le domaine de l'écologie (Connor & Simberloff, 1979; Gotelli & McCabe, 2002; MacArthur & Levins, 1967). En 1975, Diamond a étudié la distribution des espèces d'oiseaux dans des îles tropicales et a définit les règles d'assemblage de ces communautés à travers des modèles où la distribution des espèces était exprimée par des données nominales de présence ou absence. Il a trouvé que la présence de certaines espèces empêche la colonisation d'autres oiseaux dans le même environnement, ce qui génère une exclusion due à la compétition entre les espèces (concurrentielle). Ceci a constitué un premier pas vers les analyses menées pour étudier la covariation des espèces qui recherchent leurs coprésences et/ou aussi l'exclusion mutuelle de certaines espèces (Horner-Devine et al., 2007).



Figure 1.1 Les interactions biologiques possibles entre les membres de différentes espèces (Faust & Raes, 2012).

Des réseaux d'assemblages microbiens ont été calculés pour diverses communautés microbiennes associées à des écosystèmes différents tel que le sol, les océans, le corps humain ainsi que les aliments. Ces réseaux sont obtenus sur la base de l'analyse de la covariation des espèces reposant sur des données issues de séquençage à haut débit de marqueurs taxonomiques. L'abondance relative des espèces permet le calcul d'une matrice de corrélation pour toutes les combinaisons de paires d'espèces possibles (Xu, 2006). Des seuils critiques doivent ensuite être attribués pour identifier les corrélations significatives (Faust & Raes, 2012). L'idée derrière ces calculs de corrélation positives ou négative entre le profil de distribution des individus.

1.3.1 Algorithme adaptés aux données compositionnelles

Malgré le fait que l'observation de la covariation des espèces d'une communauté soit possible par des analyses de corrélation, l'interprétation biologique et le traitement statistique de ces résultats constituent de réels défis en raison de la nature compositionnelle des données métagénomiques (Faust & Raes, 2012). En effet, la nature des données générées de l'analyse génomique, telles que celles obtenues à partir du séquençage du gène de l'ARNr 16S chez les bactéries, n'est pas en mode

d'abondance absolue mais plutôt en mode d'abondance relative de gènes ou d'espèces : chaque espèce représente une proportion de la communauté dont la somme des entités dans un échantillon est un. De ce fait, l'abondance relative des espèces constitues des variables dépendantes. Par exemple, si la proportion d'une espèce dominante augmente, la proportion d'autres espèces présentes dans l'échantillon doit obligatoirement diminuer en raison du fait qu'un nombre fini de séquences peut être obtenu dans une librairies de séquençage. Dans cet exemple, la diminution de l'abondance d'une espèce peut donc être artificielle et ne pas refléter la situation réelle au sein d'un écosystème étudié. D'autre part, la nature dépendante des données d'abondance relative des espèces rend ces dernières incompatibles avec les analyses de corrélation classiques de types Spearman, Pearson et Kendall. Il a d'ailleurs été démontré que les corrélations identifiées comme significatives entre certaines espèces du microbiome humain selon des tests de corrélation Spearman et de Pearson étaient fausses (Fang et al., 2015; Faust & Raes, 2016; Faust et al., 2012; Friedman & Alm, 2012). La nature compositionnelle des données de séguençage requière le développement d'algorithmes adaptés à la nature des données pour évaluer la covariation des espèces (Aitchison, 1982). La nature des données compositionnelle nécessite certaines manipulations et transformations pour avoir par la suite des corrélations robustes(Watts et al., 2018). Ceci est rendu possible en avant recours à la transformation des abondances des espèces en logarithme (le rapport de log ; ratiolog). Selon cette approche, chaque combinaison de paires d'espèces est représentée par un ratio-log; l'abondance relative de chaque paire (variables dépendantes) sont combinée en une nouvelle variable. Ce ratio est calculé pour chaque combinaison d'espèces, dans chaque échantillon, et un calcul de variance permet de déterminer si deux espèces tendent à co-exister, à être mutuellement exclusives ou à démontrer aucune covariation. Friedman et Alm (2012) ont proposé une méthode d'approximation robuste appelée SparCC pour déduire la matrice de corrélation basée sur une estimation approximative de la variance du ratio-log des espèces (Fang et al., 2015). Puisque la transformation en logarithme ne peut pas être appliquée aux zéros, les valeurs nulles sont remplacées par une valeur arbitraire (i.e. pseudo count). Cette approche suppose implicitement que tous les composants sont effectivement présents

dans un échantillon et que les valeurs nulles résultent d'une limite de détection. Une série d'itérations permet d'identifier les espèces dont les profils de distribution sont les plus fortement corrélés, minimisant ainsi le nombre de corrélations significatives obtenues avec des outils de corrélation classiques dont les résultats ne permettent pas de définir les covariations indirectes (e.g. présence de l'espèce A favorisée par l'espèce B si l'espèce C est également présente). Le seuil de signification statistique des corrélations SparCC est basé sur des permutations, générant ainsi des pseudo-valeurs de probabilités. Aussi, il existe des autres algorithmes traitant les données compositionnelles citant par exemple Spiec-Easi et Lasso (Fang et al., 2015; Tsilimigras & Fodor, 2016).

1.3.2 Limites des réseaux de covariations

Un réseau écologique est une représentation des interactions directes incluant le parasitisme, le commensalisme, le mutualisme, l'amensalisme et la compétition, et les interactions indirectes, qu'elles soient positives et /ou négatives. Les interactions positives sont le reflet de processus de coopération entre les organismes. Par exemple, des substances secrétées par une espèce peuvent être métabolisées par une autre. Pour leur part, les interactions négatives reflètent plutôt un effet d'antagonisme qui peut s'expliquer, entre autres, par une compétition pour les ressources limitées ou pour l'espace. Bien qu'un réseau de covariation soit basé sur l'analyse de corrélations, ces dernières ne reflètent pas nécessairement l'existence d'interactions écologiques réelles et ne doivent pas être considérés comme un portrait absolu du dynamisme d'une communauté microbienne (Xiao et al., 2017). Effectivement, certaines covariations observées par les analyses de corrélation peuvent s'expliquer par d'autres facteurs tels que l'adaptation des espèces qui coexistent aux conditions de l'environnement (Bell, 2005b). De ce fait, l'identification du type d'interaction est cruciale à la compréhension des mécanismes de régulation devant exister entre les microorganismes pour assurer et maintenir leurs survies.

Une approche prometteuse pour traduire la structure des profils de covariation des espèces est d'expliquer les résultats observés par l'application de théories écologiques. Dans ce contexte, la covariation des espèces est d'abord déterminée statistiquement

dans un cadre spatio-temporel bien défini. Dans un deuxième temps, la matrice de covariation obtenue est intégrée dans le but de conceptualiser et à chercher des règles d'assemblages expliquant les successions de covariations qui capturées au cours du temps. Dans un objectif de définir les facteurs pouvant expliquer la covariation entre les espèces, plusieurs hypothèses ont été mise à ce propos, y compris les abondances relatives des espèces, la taxonomie et la distribution locale de traits écologiques.

1.4 Théorie expliquant la structure des réseaux écologiques

Bien que les réseaux d'interactions aient été beaucoup étudiés en écologie, les réseaux d'interactions microbiennes restent moins étudiés en raison de leur énorme diversité et aussi en raison de l'inaccessibilité aux bactéries non cultivables. Aussi, les recherches en biodiversité microbienne accentuent l'effort sur la richesse et l'abondance des espèces, mais mettent moins l'emphase sur les interactions entre elles (Zhou *et al.*, 2010). Toutefois, une étude entre des profils de covariation des microorganismes et des organismes supérieurs a révélé que la structure des réseaux est conservée, peu importe le rang taxonomique des organismes impliqués. Effectivement, il est apparu que certaines espèces microbiennes coexistent moins souvent que prévu par hasard, tel qu'observé pour certains organismes supérieurs (Horner-Devine *et al.*, 2007).

1.4.1 Covariation des espèces expliquée par la taxonomie

Par analogie aux observations faite en écologie végétale, certaines études ont conclu que les bactéries reliées taxonomiquement tendent à coexister ensemble dans le même milieu en raison de leurs similarités écologiques (Cavender-Bares et al., 2006; Cavender-Bares et al., 2009; Kembel & Hubbell, 2006; Webb, 2000; Weiblen et al., 2006). Cependant, d'autres études avancent plutôt le contraire, avec un lien moins évident entre la covariation et la taxonomie, en raison de la haute fréquence des événements de transfert latéral de gènes permettant à des espèces du même groupe taxonomique d'acquérir de nouvelles fonctionnalités les distinguant d'un point de vue écologique (Manhart & Shakhnovich, 2018; Ochman et al., 2000). Par ailleurs, des études récentes sur les communautés microbiennes ont abouti au fait que le filtrage

exercé par l'environnement sélectionne la fonction convergente dans l'habitat, tout en permettant la variabilité taxonomique au sein de chaque classe fonctionnelle (Goldford et al., 2018). En outre, la coopération entre les espèces appartenant à différents groupes phylogénétiques éloignées taxonomiquement pourrait être due à la similitude métabolique réduite (Foster & Bell, 2012).

1.4.2 Covariation des espèces expliquée par la sélection de traits

Les traits écologiques représentent la réponse ou l'expression des espèces face à des conditions environnementales bien définie. Certains auteurs se sont intéressés à évaluer la réponse des espèces à des perturbations environnementales en leur attribuant les traits d'espèces tolérantes ou résistantes ou résiliente alors que d'autres auteurs préconisent l'utilisation de paramètres de croissance (e.g. affinité pour le substrat) pour étudier la covariation des espèces (Ho *et al.*, 2014b). De manière analogue aux processus de sélection naturelle, les habitats agissent comme un filtre pour les espèces induisant la cohabitation de celles qui sont mieux adaptés au détriment des moins adaptés (Bell, 2005b; Fisher *et al.*, 2017). D'autres études ont montré que la compétition entre les espèces qui va conduire à l'extinction de certaines de ces derniers, affecte la distribution des traits (Fisher *et al.*, 2017). Mesurer les traits écologiques représente un défi en écologie microbienne (Cornwell *et al.*, 2006). Une question fondamentale se rapporte à la sélection d'un ou de plusieurs traits pertinents pour identifier leur implication dans la définition des règles d'assemblage.

1.5 Pressions multifactorielles modulant la covariation des espèces

Les théories précédentes introduisent que les espèces cohabitent sans tenir compte des interactions potentielles entre les facteurs environnementaux. Cependant, l'existence d'une interaction ne se limite pas à la présence des espèces impliquées mais nécessite également des conditions environnementales particulières. En 2014, Poisot et ses collaborateurs ont proposé un modèle mathématique théorique expliquant la covariation des espèces en fonction de leur abondance relative et de la réponse aux traits écologiques. Suivant ce modèle, l'utilisation d'une matrice de covariation n'est pas

une solution optimale pour étudier les règles d'assemblage des communautés microbiennes, puisque les fluctuation environnementales ne sont pas prises en compte dans ce type d'analyse (Poisot *et al.*, 2015). L'abondance relative d'une espèce est la variable fondamentale pour qu'une covariation pourra exister entre deux espèces, notamment l'abondance d'une espèce influence les covariations qui pourront soit exister ou disparaitre, même l'augmentation ou la diminution d'une ressource dans le milieu peut favoriser l'augmentation de l'abondance de certaines espèces similaires ou la diminution de celle-ci. Ainsi, ces forces écologiques opposées agissent simultanément sur l'abondance des espèces pour façonner la composition d'une communauté.

1.6 Écologie appliquée au secteur alimentaire

Il est généralement admis que les aliments sont des milieux propices aux développement de certains microorganismes dont certains sont reliés à la détérioration de la qualité et la salubrité alimentaire ou aussi jouant un rôle essentiel dans la structure de certains aliments (Huang, 2017). La microbiologie alimentaire traditionnelle a touché surtout l'aspect de détection des microorganismes pathogènes, l'identification des mécanismes responsables de la contamination et le rôle des microorganismes bénéfiques pour la fermentation industrielle (Ercolini, 2013). Bien que ces approches soient toujours en place, la détection des pathogènes et le contrôle de salubrité ainsi que la maîtrise des contaminations alimentaires sont limités par le temps assez long de l'analyse dépassant des fois la durée de conservation de l'aliment lui-même. Par exemple les méthodes traditionnelles de dépistage nécessitent un enrichissement préalable pour pouvoir isoler les microorganismes en question sur géloses nutritives sélectives différentielles. La voie de recherche du secteur bioagroalimentaire est en pleine effervescence cherchant de nouvelles approches pour assurer la qualité de ses produits. Ces dernières dépassent la simple détection de pathogènes (limite de détection 1 UFC par portion d'échantillon : la présence des antimicrobiens) envisageant des approches plus sophistiquées pour identifier la source de contamination.

1.6.1 Discrimination moléculaire des souches pathogènes et non pathogènes

Un des progrès du secteur vise le développement d'outils discriminants pour distinguer les microorganismes pathogènes et non pathogènes. Évidemment, le séquençage du génome complet a permis de découvrir les facteurs de virulence, mais la disponibilité de ces bases de données a permis de mettre en relation la pathogénicité avec certains facteurs environnementaux qui prédisposent les aliments à une contamination microbiologique. En fait, il est désormais possible de prédire la susceptibilité antimicrobienne de la flore microbienne des aliments (déterminer la présence des gènes de résistance antimicrobienne) ainsi que de prédire leurs profils de virulence (Leekitcharoenphon et al., 2013; Parks et al., 2015; Tyson et al., 2015). Par exemple, la comparaison de génomes d'Escherichia coli à une base de données contenant 2500 gènes de résistance a permis de prédire les profils montrant une résistance antimicrobienne (Tyson et al., 2015). Cela contribuera à améliorer considérablement les systèmes de surveillance des antimicrobiens permettant ainsi un meilleur suivi des isolats résistants aux antimicrobiens dans l'approvisionnement alimentaire. Ainsi, la détection des pathogènes pourrait raffiner les approches traditionnelles de microbiologie prédictive reposant sur la détection d'indicateurs pour prédire la durée de vie des aliments afin d'assurer leurs salubrités.

La microbiologie prédictive consiste à employer des modèles mathématiques en vue de prédire la croissance et/ou la survie des microorganismes pathogènes en fonction de conditions physicochimiques (e.g. température, pH, l'activité de l'eau) (Koyama et al., 2017; Tamplin, 2018). Au préalable, des modèles empiriques ont été exploités pour décrire la croissance des microorganismes, mais récemment des modèles plus sophistiqués ont été appliqués sur les aliments. Par exemple, un modèle a été établi pour identifier la durée maximale de conservation de gâteaux de riz en fonction de la charge initiale en *Staphylococcus aureus* (Wang et al., 2017). Sachant que ces approches reposent sur la détection des pathogènes par des méthodes de culture de bactéries indicatrices de contamination (e.g. coliformes totaux), le développement de diagnostic moléculaire basé sur la détection des pathogènes (e.g. gènes de virulence) s'avère prometteur.

1.6.2 Réseaux de covariance dans l'écologie alimentaire

La disponibilité de l'eau et des nutriments, le pH, l'activité de l'eau font des aliments un environnement propice à sa colonisation par des microorganismes (Kim et al., 2018). L'étude du microbiome chez les aliments et les plantes a révélé l'existence d'un profil moléculaire distinct selon le type d'aliment. Les modèles de covariation des microorganismes associés aux aliments constituent un premier pas à l'étude des interactions potentielles entre les microorganismes (figure 3). Des outils de lutte biologique peuvent être élaborés en se basant sur la structure de microbiome. Par exemple, 4,1 % du microbiome de Brassica constitue des espèces antagonistes au pathogène Verticillium qui attaque les plantes (Wassermann et al., 2017). Aussi, des corrélations positives entre la présence des bactéries de genre Alkanindige de et la présence de Xanthomonas campestris pv. Vitians causant une maladie affectant les feuilles de laitue a été observé, ou aussi des corrélations positives entre l'abondance de certaines souches de *Pseudomonas* et la bonne santé de laitue a été révélé (Berg et al., 2014; Erlacher et al., 2014). Une perspective en microbiologie alimentaire est donc d'élaborer des techniques précoces d'identification de contamination alimentaire et d'agir en amont avant que la colonisation des pathogènes soit réussie en se basant sur le microbiome (Berg et al., 2014). Cela nécessitera la compréhension des processus régissant les réseaux d'interactions de communautés microbiennes associées aux aliments et aux plantes.



Figure 1.2 (a) Réseau de covariance (association significative) entre les espèces microbiennes de microbiome de Brassica (b) zoom sur le réseau (c) Classification taxonomique de microbiome de Brassica (Wassermann et al., 2017)

1.7 Hypothèse

En considérant l'importance de la connaissance des règles impliquées dans le processus de covariation des communautés microbiennes, nous proposons un modèle théorique qui prétend que la covariation des bactéries est sous le contrôle de la taxonomie, l'abondance relative et les traits physiologiques de chaque espèce.

1.7.1 Objectifs

- 1. Obtenir un portrait moléculaire des communautés microbiennes associé aux chia.
- 2. Etablir une collection d'isolats bactériens représentative des profils moléculaires en utilisant une combinaison de milieux de culture.
- Obtenir un portrait de traits physiologiques des isolats (un profil de tolérance à différents stress physiologiques et inhibiteurs de croissance).
- Tester le modèle mathématique proposé en appliquant une série de régressions linéaires.

Physiological traits and relative abundance of species as explanatory variables of co-occurrence pattern of cultivable bacteria associated with chia seeds

ARTICLE

CHAPITRE 2 : ARTICLE

<u>Titre</u>

Physiological traits and relative abundance of species as explanatory variables of cooccurrence pattern of cultivable bacteria associated with chia seeds

<u>Auteurs</u>

Asma Jaba, Fadi Dagher, Amir Hamidi, Claude Guertin and Philippe Constant

<u>Journal</u>

Canadian Journal of Microbiology

État actuel

Soumis

2.1 Contribution à l'article

Asma Jaba a réalisé les expériences en laboratoire, a compilé les données, a interprété les résultats et a participé à la rédaction de l'article sous la supervision de Pr Philippe Constant et Pr Claude Guertin.

2.2 Résumé

Déchiffrer les règles impliqués dans l'assemblage des communautés microbiennes constitue une stratégie prometteuse, notamment pour améliorer l'identification des facteurs de prédisposition à la contamination microbiologique. Malgré le nombre croissant d'études rapportant des modèles de covariation microbiennes, il existe très peu de tentatives expérimentales qui ont testé leurs fiabilités. Nous cherchons à vérifier l'hypothèse proposant que les modèles de covariation observés pourraient s'expliquer par la taxonomie, l'abondance relative et les traits physiologiques des espèces microbiennes. Le séquençage des marqueurs taxonomiques a permis d'évaluer les profils de distribution et d'élaborer un modèle de covariation d'espèces bactériennes et fongiques trouvées dans 25 échantillons de chia provenant de huit sources différentes. En utilisant une combinaison des milieux de cultures riches en nutriments et oligotrophe, 16 espèces bactériennes ont été isolés dont cinq correspondaient à des phylotypes représentés dans le profil moléculaire. La tolérance à différents inhibiteurs de croissance et aux antibiotiques a été testée pour évaluer les caractéristiques physiologiques de ces isolats. Des analyses de régression linéaire multiple ont démontré que la divergence des traits physiologiques et l'abondance relative de chaque paire d'espèces expliquaient 69% du profil de cooccurrence présenté par les phylotypes bactériens cultivables associé aux chia. Malgré l'importance de ce résultat pour la conceptualisation des réseaux microbiens, l'application de notre approche à des autres matrices alimentaires sera nécessaire pour démontrer la fiabilité du cadre théorique proposé pour expliquer les modèles de cooccurrence dans les aliments.

2.3 Abstract

Deciphering the rules defining microbial community assemblage is envisioned as a promising strategy to improve predictions of pathogens colonization and proliferation in food. Despite the increasing number of studies reporting microbial co-occurrence patterns, only a few attempts were made to challenge them in experimental or theoretical frameworks. Here, we tested the hypothesis that observed co-occurrence patterns could be explained by taxonomy, relative abundance and physiological traits of microbial species. PCR amplicon sequencing of taxonomic markers was first conducted to assess distribution and co-occurrence patterns of bacterial and fungal species found in 25 chia samples originating from eight different sources. The use of nutrient-rich and oligotrophic media enabled isolation of 71 strains encompassing 16 bacterial species, of which five corresponded to phylotypes represented in the molecular survey. Tolerance to different growth inhibitors and antibiotics was tested to assess physiological traits of these isolates. Multiple linear regression analyses demonstrated that divergence of physiological traits and relative abundance of each pair of species explained 69% of the co-occurrence profile displayed by cultivable bacterial phylotypes in chia. Despite the significance of this result for the conceptualization of microbial networks, application of the approach to more food products will be necessary to demonstrate reliability of the proposed theoretical framework to explain co-occurrence patterns in food.

Keywords: Microbial ecology, microbiome, microbial communities.

2.4 Introduction

Progress in sequencing and bioinformatics technologies experienced over the last decade is accompanied by an astonishing buildup of genomic databases of microbial community structure originating from almost any types of environments (Stephens *et al.*, 2015). These advances provide unprecedented opportunities to bridge knowledge gaps into processes governing microbial community structure driven by a combination of stochastic and selection mechanisms (Nemergut *et al.*, 2013; Vellend, 2010). The selection effect of food matrix is exemplified by a certain level of specificity for microbial communities associated with different fruits and vegetables (Leff & Fierer, 2013). However, foodborne pathogens or contaminants may arise following accidental exposure and infection of food during transport or storage. Persistence and proliferation of these allochthonous microorganisms are then driven by abiotic and biotic features characterizing food matrix.

The relevance of abiotic features of food produces in determining the structure of their microbiota is demonstrated by predictive microbiology approaches implementing sophisticated models incorporating environmental factors to predict the proliferation of microorganisms (Mejlholm et al., 2010; Tenenhaus-Aziza & Ellouze, 2015; Zoellner et al., 2018). On the other hand, the impact of biotic features including microbe-microbe interactions has received less attention, even though experimental evidence supports the notion that species co-occurrence patterns are non-random (Berg, 2015). A pioneering work was the application of the checkerboard principle, where species exclusions are competitive interactions induced by resource limitation, that unveiled non-random species co-occurrence patterns observed in mammals, birds, plants and invertebrates hold in microbial communities (Horner-Devine et al., 2007). The reliance of the checkerboard principle on presence/absence datasets complicates the implementation of the approach for microbial community profiles obtained by highthroughput sequencing technologies. Indeed, these portraits are compositional matrices represented by a finite number of sequences retrieved from environmental DNA resulting in inverse relationships between the relative abundance and the variance of sequence features and the occurrence of several challenging to handle zero values

(Kaul *et al.*, 2017). These specificities regarding microbial community data assembly were considered in the implementation of several algorithms designed to compute species co-occurrence relying on their relative abundance with positive (aggregation) and negative (avoidance) relationships (Friedman & Alm, 2012; Kurtz *et al.*, 2015). However, such potential microbe-microbe interactions are more suggestive than definitive owing to neutral processes selecting coexisting species from their fitness in the habitat (Bell, 2005a).

Despite the constraints imposed by molecular tools, implementation of microbial species correlation matrix into theoretical frameworks is expected to identify general principles or rules deciphering assembly and structure of microbial communities. An emerging strategy is to relate pairwise correlation score of microbial species to their physiological traits selected by habitat constraints. Indeed, incongruence between phylogenetic distance and physiological traits suggest the later are the most significant parameter to infer microbial species interactions with their environment (Ho *et al.*, 2013; Krause *et al.*, 2014). Under this framework, co-occurrence of microbial species can be computed from molecular profiles and the resulting pairwise correlation coefficients are then related to physiological traits of individual species expressed either as whole genome sequence similarity (Kamneva, 2017), predicted functions (Mandakovic *et al.*, 2018) or approaches relying on phenotype characterization of cultivable members of the communities, as proposed here.

In recent years, chia has acquired great interest among consumers mainly due to the nutritional benefits that are claimed for this crop (Salgado-Cruz *et al.*, 2013). They are often consumed raw, and some recalls were experienced due to contamination with foodborne pathogens (Tamber *et al.*, 2016). In this study, the hypothesis that phylogenetic distance, physiological traits, and relative abundance of species exert a filtering effect on microbial assemblages was challenged using a combination of molecular and cultivation methods. PCR-amplicon sequencing of taxonomic marker genes was first conducted to obtain a portrait bacterial and fungal community structure associated with chia and explore co-occurrence between individual species. Efforts were then invested in building a representative collection of bacterial isolates associated with chia. Finally, bacterial isolates for which the 16S rRNA gene sequence was

identical to phylotype detected in the molecular survey were characterized. Pairwise phylogenetic distance, distance based on physiological traits, and relative abundances were computed for all combinations of isolates to test the relevance of these variables to explain observed interspecific co-occurrences in the molecular survey.

2.5 Material and methods

2.5.1 Samples

Chia (*Salvia hispanica* L.) is an annual herbaceous plant that belongs to the Lamiaceae family, which is native from northern Guatemala and southern Mexico. The plant producing these brown to black or non-pigmented beige seeds is generally cultivated in tropical and subtropical regions (Ixtaina et al., 2008; Muñoz et al., 2012). Agri-Neo Inc. provided twenty-five samples of non-sprouted chia from Argentina (n = 7), Paraguay (n = 4), Bolivia (n = 5), Ecuador (n = 3), Nicaragua (n = 1), Mexico (n = 1), and a Canadian distributor (n=3). Another sample was obtained from Nativas[®] Organics (n = 1) a California based seed processor and distributor. Chia samples originating from the same site arise from different batches. These seeds were stored at room temperature in plastic bags for less than two weeks upon genomic DNA extraction.

2.5.2 Genomic DNA extraction

Total genomic DNA was extracted from chia samples using Fast DNA Spin Kit[®] (MP Biomedicals, Santa Ana, California, USA) according to specifications of the manufacturer, including two successive mechanical lysis steps performed with a FastPrep-24[®] homogenizer (MP Biomedicals, Santa Ana, California, USA). Based on preliminary tests of the extraction procedure on different amounts of chia (*i.e.* 100, 200 or 300 mg), samples were processed using 100 mg chia because seed mucilage impaired genomic DNA extraction from 200 and 300 mg samples. For bacterial isolates, the biomass of axenic culture was first washed from agar plate with 5 ml glycerol (20% v/v) followed by 10-minute centrifugation (20,000 *g*, 4°C). The chia sample or bacterial pellet was mixed with 100-200 mg silica beads (150-212 µm diameter), 1 ml TEN buffer (50 mM Tris-HCl, 10 mM EDTA, 150 mM NaCl, pH 8.0) and 20 µL SDS (20% w/v).

Mechanical cell lysis was performed in two successive times with FastPrep-24[®] homogenizer (MP Biomedicals, Santa Ana, California, USA) for 45 seconds at 6.5 m s⁻¹, with a 5-minute incubation on ice between cycles. The lysed cell mixture was centrifuged (20,000 *g* for 10 minutes) and the recovered aqueous phase was treated with RNase (20 μ g ml⁻¹) for 10 minutes at room temperature before adding 500 μ l phenol: chloroform: isoamyl alcohol solution (25: 24: 1, pH 7.0) to purify DNA. The aqueous phase obtained after centrifugation (20,000 *g* for 10 minutes) was mixed with 500 μ L chloroform: isoamyl alcohol solution (24: 1) and centrifuged at 20,000 *g* for 2 minutes at 4°C. After collecting the aqueous phase (approximatively 500 μ L), a 167 μ L volume of ammonium acetate (10 M) was added, incubated for 20 minutes on ice and centrifuged (20,000 *g* for 15 minutes). The supernatant was collected, and nucleic acids were precipitated with 1 ml ethanol (100%) at -20 °C. The pellet was recovered after centrifugation (20,000 *g* for 15 minutes at 4°C) and DNA was confirmed by 1% (w/v) agarose gel electrophoresis. Extracted DNA aliquots were stored at -20°C.

2.5.3 Microbial community structure

Extracted DNA samples were subjected to marker genes PCR amplification, library preparation and high-throughput sequencing on Illumina MiSeq PE-250 platform at the McGill University and Genome Quebec Innovation Center (Montreal, Canada). Primers B969F and BA1406R were used to PCR-amplify V6-V8 region of bacterial 16S rRNA gene (Comeau *et al.*, 2011), while the primers ITS3_KYO2F and ITS4_KYO3R were used for the second internal transcribed spacer (ITS2) flanking the 5.8S and 28S rRNA genes in fungi (Toju *et al.*, 2012). Raw sequencing reads were proceeded using the software Usearch version 10 (Edgar, 2010). Paired reads were assembled to a total length varying between 400 and 500 nucleotides for bacteria and 190 and 450 nucleotides for fungi. In all 2,691,709 bacterial and 1,583,794 fungal merged read sequences were subject to quality control. Maximum mismatch threshold in the overlapped region of assembly was set at 5 for bacteria, and 10 for fungi and primers were removed from each sequence. Sequences having less than one erroneous base were accepted for downstream quality control steps. Reads were then dereplicated,

singletons were discarded and denoised using Unoise3 (Edgar, 2016b). Sequences shorter than 366 and 150 nucleotides for bacteria and fungi were discarded, respectively. Classification of the resulting filtered sequences was undertaken using two successive schemes. In the first approach, single sequences were clustered into amplicon sequence variant (ASV) corresponding to a classification at the 100% sequence identity threshold (Callahan et al., 2017). That clustering approach was proven efficient to relate ASV sequence retrieved from the molecular survey to 16S rRNA gene sequence of bacterial isolates. Indeed, unambiguous assignment of each isolate to a single ASV was mandatory for the theoretical framework of microbial species co-occurrence. ASV representing less than 0.005% of read counts were removed before the second clustering approach grouping ASV sequences into Operational Taxonomic Units (OTU) using a 97% identity level cutoff. This scheme was necessary to increase the prevalence of phylotypes in chia samples in order to fulfill standard requirements of ecological network analyses. Indeed, the OTU clustering method reduced the number of null values in bacterial and fungal relative frequency datasets. The clustering procedures led to the identification of 187 and 115 ASV grouped into 91 and 90 OTU for bacteria and fungi (Table S1), respectively. Taxonomic affiliation of ASV and OTU was predicted by k-mer similarity of ASV representative sequences to the RDP (Ribosomal Database Project) version 16 training set of 16S rRNA gene for bacteria (Cole et al., 2014) and RDP Warcup training set version 2 for fungi (Deshpande et al., 2016), altogether considering taxa agreed upon by over 80% of bootstrap replications using the SINTAX algorithm (Edgar, 2016a). Reads assigned to (i) chloroplasts and (ii) the spurious bacterial OTU36 and fungal OTU58, OTU 60, OTU61, OTU73, OTU93, OTU131, OTU138, OTU141, OTU142 identified as unknown bacterium and fungus for which Blast hits against the National Center for Biotechnology Information (NCBI) genomic database (<u>http://www.ncbi.nlm.nih.gov/</u>) were nonspecific, were removed from the dataset. Raw sequence reads were deposited in the Sequence Read Archive of the National Center for Biotechnology Information under the Bioproject PRJNA485274.

2.5.4 Isolation of bacteria

Two attempts were conducted to isolate bacteria associated with chia, using five chia samples randomly selected for each isolation procedure. For the first approach, 100 mg of seeds were transferred into 900 µL saline solution (0.85% NaCl). Seeds were thoroughly mixed for one minute and 100 µL of the mixture was inoculated on oligotrophic R2A agar plate containing (in g L⁻¹) proteose peptone (3.0), sodium pyruvate (0.30), yeast extract (0.50), dibasic potassium phosphate (0.30), casein acid hydrolyzate (0.50), magnesium sulfate heptahydrate (0.05), glucose (0.50), soluble starch (0.50), and agar (15) with a final pH of 7.0. The second approach was aimed at isolating copiotrophic bacteria, using tryptone soya medium containing (in g L⁻¹) Bacto[®] Tryptone (17.0), Bacto[®] Soytone (3.0), glucose (2.5), sodium chloride (5.0), dipotassium hydrogen phosphate (2.5), and agar (15). Pre-incubation of 100 mg chia seeds was done in 5 ml tryptone soya broth supplemented with the antifungal cycloheximide (100 mg/ml) for three days at 25°C under 150 rpm agitation from which 100 µL was spread on TSB agar plate. Incubation of R2A and TSB agar plates was performed at 30°C. Individual colonies were isolated and purified through three successive transfers on agar plates. Aliquots of axenic cultures were collected for storage at -80°C in 20% glycerol and genomic DNA extraction.

2.5.5 Molecular identification and classification of bacterial isolates

Genomic DNA extracted from axenic cultures was subjected to PCR amplification of the 16S rRNA gene using the primers 27F and 1492R (Lane, 1991; Turner *et al.*, 1999). PCR products were shipped to McGill University and Genome Quebec Innovation Center (Montreal, Canada) for Sanger sequencing using the primer 1492R. Gene sequences of the isolates were thoroughly examined and ambiguous bases were manually corrected on Chromas[®] version 2.6.5 (Technelysium Pty Ltd, South Brisbane, Australia). The sequences from isolates and representative sequence of ASV detected in the molecular survey were aligned with Muscle algorithm (Edgar, 2004). The alignment of sequences obtained counted 700 positions. The software Bioedit version 7.0.3 (Hall, 1999) was used to generate a pairwise identity matrix to cluster isolates into

species (97% identity cutoff between 16S rRNA gene sequence of bacterial isolates) and find which bacterial isolate represent ASV sequences (100% identity cutoff between ASV and isolate 16S rRNA gene sequences). Basic Local Alignment Search Tool was utilized to retrieve 16S rRNA sequences from the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/) similar to those of bacterial isolates. Only genomic information from type material was considered, and three entries from different genera showing the highest identity score to query were kept for phylogenetic analyses. Phylogenetic analyses were conducted using 16S rRNA sequences retrieved from the molecular survey (OTU), bacterial isolates, and NCBI. The resulting 147 sequences were imported in the software Mega 6.0 (Tamura *et al.*, 2013) and aligned using Muscle algorithm (Edgar, 2004), resulting in 302 nucleic acid positions in the final dataset. The phylogenetic tree was computed by using the Maximum Likelihood method based on the Tamura-Nei model (Tamura & Nei, 1993).

2.5.6 Physiological traits of bacterial isolates

Physiological traits of bacterial isolates were evaluated using GEN III MicroPlate[®] (Biolog Inc, Hayward, California, USA) comprising 94 phenotypic reactions, with 71 corresponding to carbon sources and 23 corresponding to sensitivity tests to chemical inhibitors. The use of carbon substrates or resistance to chemical inhibitors by isolates results in the appearance of a purple color. This purple coloration is produced by reducing the indicator, tetrazolium, to a colored formazan compound. Qualitative reading of the plate was conducted by the assignation of 0 (no color) or 1 (purple color) binary score to each well after 24-36 hours incubation at 30°C. For each plate, wells were inoculated with 100 µL bacterial suspension consisting of 3-mm diameter colony forming unit harvested from R2A agar in 15 ml inoculation fluid. Selection of inoculation fluid (protocol A or B) was based on the response of negative control of each strain, with protocol B utilized in the case of appearance of a purple color in the negative control well included in GEN III MicroPlate[®] with protocol A. Physiological traits of bacterial strains were defined by their response to 23 chemical sensitivity assays encompassing tolerance to acidic pH (pH 5 and 6), salt (NaCl 1, 4, and 8%), growth inhibitors (Niaproof 4[®], guanidine hydrochloride, sodium lactate, D-serine, lithium chloride, potassium

tellurite, sodium butyrate, sodium bromate, tetrazolium violet, and tetrazolium blue), and antibiotics (troleandomycin, rifampicin, minocycline, lincomycin, vancomycin, nalidixic acid, aztreonam, and fusilic acid).

2.5.7 Statistical analyses

Statistical analyses were performed with the software R version 3.3.0 (Team, 2008) using the Vegan package (Oksanen et al., 2012) unless otherwise stated. The potential relationship between the geographic origin of chia samples and microbial community structure was first examined following two complementary approaches. Firstly, alpha diversity indices and species richness estimator of bacterial and fungal community structure were computed, and Kruskal-Wallis tests were executed with the "stats" package (Team, 2008). Secondly, the contribution of geographic origin of chia samples to partition Bray-Curtis distance matrix of bacterial and fungal OTU relative abundance data were computed by permutational multivariate analysis of variance (PERMANOVA) and visualized with principal coordinate analysis (PCoA) using the "Phyloseg" package (McMurdie & Holmes, 2013). Co-occurrence network of bacterial and fungal OTU was examined using SparCC coefficient computed in the "SpiecEasi" package (Kurtz et al., 2015) parameterized with a threshold of 0.6 and a maximum number of iterations of 20 (Friedman & Alm, 2012). The significance of pairwise SparCC coefficients was determined through 1000 bootstraps. Although correlation matrix derived from a subset of bacterial and fungal species detected in a minimal number of samples (e.g., 40-50%) is a current practice to avoid analyses involving rare species whose counts are uncertain, that approach leads to substantial loss of data. Using that approach, 86% of bacterial and fungal OTU were excluded and not relevant for subsequent isolation efforts. If one further considers the small proportion of cultivable microorganisms in the environment, these constraints leave little room to obtain experimental evidence supporting co-occurrence patterns. Therefore, a second approach was used to compute microbial species co-occurrence using the whole bacterial OTU datasets to achieve a trade-off between confidence regarding computed pairwise correlations and the probability to obtain relevant isolates for the proposed co-occurrence model framework. Statistical analyses related to bacterial isolates first included the elaboration of

rarefaction curves to assess the isolation effort of the bacteria on R2A and TSB cultivation media. Comparison of isolates based on physiological traits was performed using a Jaccard distance matrix generated on the binary dataset.

Multiple linear regression analyses based on the theoretical framework of Poisot et al. (Poisot *et al.*, 2015) were computed to evaluate the contribution of abundance, physiological traits, and taxonomy in explaining co-occurrence pattern of bacterial species observed in the 25 chia samples:

$$A_{(i,j)} = \tau_{(i,j)} + PD_{(i,j)} + N_{(i,j)}$$
 (equation 1)

where bacterial species are isolates *i* and *j* assigned to OTU *i* and *j* in the molecular survey, $A_{i,j}$ is the pairwise SparCC coefficient between the OTU *I*, and *j* computed using the whole OTU *table*, $\tau_{(i,j)}$ is the Jaccard distance of physiological traits between isolates *i* and *j*, $PD_{(i,j)}$ is the phylogenetic distance computed as the difference score (score of 0 corresponds to 100% identity) between 16S rRNA partial gene sequences of isolates *i* and *j* after pairwise alignment and $N_{(i,j)}$ is a term representing the relative abundance of OTU *i* and *j*:

$$N_{(i,j)} = \left(\frac{\sum_{n=1}^{25} x_i}{\sum_{n=1}^{25} x_j}\right) \times \left(\sum_{n=1}^{25} x_i + \sum_{n=1}^{25} x_j\right)$$
(equation 2)

where $\sum_{n=1}^{25} xi$ and $\sum_{n=1}^{25} xj$ correspond to the sum of the relative abundance of x_i and x_j in the 25 chia samples, respectively. By convention, the smallest value between $\sum_{n=1}^{25} xi$ and $\sum_{n=1}^{25} xj$ is used as numerator for the first term of the equation. This calculation was elaborated to adjust the ratio of the relative abundance of both species with the sum of their relative abundance with the rationale that the probability for two bacterial species to be in close proximity and interact in chia increases with their absolute abundance (Cardinale *et al.*, 2015; Malakar *et al.*, 2003).

2.6 Results

2.6.1 Microbial community structure and covariation between microbial species

A total of 25 chia samples originating from eight different sources were obtained for this study. According to species richness estimator (Chao1), the marker gene PCR-amplicon sequencing effort was sufficient to recover the whole diversity of microbial communities (Table 1). Bacterial communities were dominated by the phyla *Proteobacteria* (91.8%), *Firmicutes* (3.5%) and *Chlamydiae* (1.7%), *Actinobacteria* (1.2%), *Bacteroidetes* (1.2%), and *Chloroflexi* (0.6%), while most of the fungi were represented by *Ascomycota* (99.9%). The bacterial OTU B6 and B17 affiliated to Burkholderiaceae and the fungal OTU 1 affiliated to Pleosporaceae comprised a core microbiome detected in all samples. Species richness and evenness of microbial communities could not be discriminated by origin (Table 1).

Table 1. General information related to molecular profile of bacterial and fungal communities associated with chia samples. None of the estimated parameters showed significant difference between samples of different sources (Kruskal-Wallis test; p > 0.05).

Source	n	Bacteria				Fungi					
		n _{reads}	n _{otu} †	Shannon [†]	Simpson [†]	Chao [†]	n _{reads}	n _{otu} ⁺	Shannon ⁺	Simpson [†]	Chao⁺
Paraguay	4	5541	13 ± 5	1.5 ± 0.5	0.66± 0.18	16 ± 7	266 467	26 ± 9	1.3 ± 0.6	0.52± 0.26	28 ± 10
Argentina	7	6076	17 ± 8	2.1 ± 0.5	0.81± 0.09	2 ± 1	503 894	20 ± 10	0.7 ± 0.6	0.33± 0.29	26 ± 10
Nicaragua	1	674	13	2.0	0.85	13	46 492	14	1.4	0.74	16
Bolivia	5	5061	18 ± 13	1.8 ± 0.7	0.73± 0.20	23 ± 13	328 465	22 ± 6	1.1 ± 0.5	0.56± 0.25	26 ± 11
Ecuador	3	2848	15 ± 7	2.0 ± 0.6	0.80± 0.14	16 ± 9	47 564	16 ± 4	1.2 ± 0.1	0.57± 0.11	19 ± 8
Mexico	1	356	7	1.0	0.46	8	3398	10	0.08	0.02	11
Canada	3	1919	11 ± 4	1.7 ± 0.3	0.74 ±0.89	16 ± 9	216 744	18 ± 6	0.26 ± 0.26	0.11± 0.12	23 ±6
USA	1	31 140	22	1.79	0.75	7	109 226	25	1.77	0.77	28

[†]For chia sources where n > 1, mean and standard deviation are provided.

The potential impact of chia sample origin on the beta diversity of microorganisms was further explored by computing PCoA, with the first two axes explaining, respectively, 32 and 41% of the variation in bacterial (Figure 1A) and fungal (Figure 1B) community profiles. Dispersion of microbial community profiles in the reduced space of the PCoA showed no unambiguous patterns explained by sample origin. The relationship between sample origin and microbial community profiles was further examined using Permanova analyses, considering the whole community instead of the previous reduced space, computed on chia seed samples represented by at least three independent samples. The results indicated no significant contribution of sample origin in explaining composition of bacterial communities ($R^2 = 0.23$; P = 0.08) whereas a significant contribution was observed for fungi ($R^2 = 0.28$; P = 0.02).



Figure 1. Principal coordinate analysis of bacterial (A) and fungal (B) communities in chia samples.

Co-occurrence of microbial species was firstly explored using a restricted list of ubiquitous OTUs detected in more than 11 (44%) chia samples (Figure 2). The relative abundance of these ubiquitous species varied between 0.05-34% and 0.01-42% for bacteria and fungi, respectively. Network structure was fragmented, with the occurrence of three modules comprising between two and five members (Figure 2). The fungal OTU F132 showed the highest connectivity, with positive co-occurrence with fungal OTU 49 and negative co-occurrence with fungal OTU-F6 and OTU-F9. In the second approach, microbial species co-occurrence was analyzed using the whole bacterial and fungal OTU datasets to achieve a trade-off between confidence regarding computed pairwise correlations and the probability to obtain relevant isolates for the proposed co-occurrence model framework ($A_{(i,j)}$; Table S2).



Figure 2. Ecological network of ubiquitous bacterial and fungal OTU in chia. The OTU are represented by the nodes and edges illustrate significant pairwise covariation (SparCC, pseudo P value < 0.05). Nodes are colored according to taxonomic affiliation of the OTU, the numbers appearing next to the nodes refer to OTU identifier and the size of the nodes is scaled according to the logarithm of OTU sequence read count in the 25 chia samples. Color of the edges indicates positive (green) or negative (red) correlations.

2.6.2 Isolation of bacteria associated with chia

Individual colony forming units propagated on R2A and TSA cultivation media were isolated for downstream DNA extraction and 16S rRNA gene sequencing. All isolates encompassed the Firmicutes and Proteobacteria (Figure 3).



Figure 3. Maximum likelihood phylogenetic tree of 16S rRNA gene sequence of isolates and OTU retrieved from the molecular survey.

An overview of the taxonomic classification of 16S rRNA gene sequence encompassing three phyla is presented in the first panel (A) with magnification of (B) alpha- and (C) gamma-Proteobacteria clusters in the secondary panels. Bootstrap values (%) are represented in red characters for the nodes that are supported by at least 50% iterations. The five isolates whose 16S rRNA sequences was 100% identical to sequences retrieved from the molecular survey are shown in bold characters.

In contrast to conventional approaches where differences in phenotypic traits are used to guide isolation efforts, isolation and sequencing of 16S rRNA of all colonies enabled reliable quantification (Table S3). Indeed, clusterization of 16S rRNA sequence of isolates at the 97% identity cut-off was used to compare the number of individual species with theoretical estimates of the whole diversity of cultivable bacteria (Figure 4A). For TSB medium, 14 isolates were clustered into five different species mainly represented by *Stenotrophomonas* sp., *Enterobacter* sp. and *Bacillus* sp. (Figure 4B). Additional isolation efforts were not intended since the number of retrieved species corresponds to 100% species richness estimators (Figure 4A). A broader diversity was observed on R2A agar, with 57 isolates were clustered into 14 different species, expected to represent 85% cultivable representatives according to Chao1 species richness estimator (Figure 4A). As observed with TSB medium, there was a cultivation bias for *Bacillus* spp. and *Enterobacter* spp. in R2A medium (Figure 4B).



Figure 4. Evaluation of the isolation efforts and taxonomic distribution of the isolates. (A) Rarefaction curves were computed for quantitative analysis of isolation efforts on TSB-A and R2A media. Number of observed species (n) and Chao1 species richness estimator are presented. (B) Proportion of bacterial genera represented by isolates encompassing Alpha-Proteobacteria, Firmicutes and Gamma-Proteobacteria.

Since this work seeks to examine the contribution of phylogenetic distance, abundance, and physiological traits in shaping co-occurrence noticed in molecular profile, an exact concordance between the 16S rRNA gene sequence of isolates and PCR amplicons was imposed to bridge cultivation-dependent and cultivation-independent datasets. In all, the 16S rRNA gene sequence of five isolates encompassing alpha- and gamma-Proteobacteria was 100% identical to sequences retrieved from the molecular survey, namely Sphingomonas sp. AJ3, Enterobacter sp. AJ7, Rhizobium sp. AJ32, Sphingomonas sp. AJ28, and Methylobacterium sp. AJ8. None of these isolates but Sphingomonas sp. AJ28 corresponded to members of ubiquitous species. The other isolates whose 16S rRNA sequence was more than 97% identical to PCR amplicons classified into more than one OTU were prone to ambiguous associations and thus not considered for further analyses.

Physiological traits of the five selected bacterial isolates were defined by their response to 23 chemical sensitivity assays encompassing tolerance to acidic pH, salt, growth inhibitors, and antibiotics. The results of individual assays were integrated to compute a Jaccard distance matrix of physiological traits between isolates (τ (i,j)). Sphingomonas sp. AJ3 was the most sensitive and Enterobacter sp. AJ7 was the most resistant to chemicals. Clusterization of the phenotype profiles was not explained by the taxonomic distance of isolates since Sphingomonas sp. AJ28 demonstrated a more similar profile with Rhizobium sp. AJ32 than Sphingomonas sp. AJ3 (Figure 5).



Figure 5. Physiological traits of the five selected bacterial isolates. The heatmap shows positive (1) and negative (0) results for chemical sensitivity assays encompassing tolerance to acidic pH, salt, growth inhibitors and antibiotics. The UPGMA agglomerative clustering of bacterial isolates is based on a Jaccard distance matrix calculated with results of the chemical sensitivity assays.

2.6.3 Model framework for co-occurrence of microbial species

The five selected isolates and their OTU counterparts in the molecular survey were included in the theoretical framework aimed at testing whether abundance, physiological traits and taxonomy explain the co-occurrence pattern of bacteria species associated with chia (Table S4). For the analysis, pairwise SparCC coefficients were retrieved from the co-occurrence analysis computed using the whole bacterial database (Table S2). None of the three independent variables alone explained observed variations in pairwise correlations between the five isolates (Figure 6A). Application of forward stepwise multiple regression analyses showed that model including physiological traits and abundance terms offered the best performance, explaining 69% variations of observed pairwise SparCC correlation coefficients (Figure 6B).



Figure 6. Theoretical framework to explain co-occurrence patterns of the five selected bacterial isolates in chia. (A) Single and multiple linear regressions were computed using PD, N and T as independent variables to explain variation in SparCC pairwise coefficient observed between each pair of OTUs represented by the eight isolates. Multiple R-square value is presented for each regression analysis. The symbols ** denote p-values lower than 0.05. (B) Linear regression between observed SparCC coefficients and predicted coefficients with physiological traits and relative abundance terms (see Table 2 for equation parameters).

According to model parameters, co-occurrence on chia sample is expected for two species displaying dissimilar traits, but this relationship is lowered when the abundance term ($N_{(i,j)}$) of both species is elevated (Table 2).

Table 2. Multiple regressions showing the relationship of species co-occurrence $(A_{(i,j)})$ with physiological traits $(r_{(i,j)})$ and relative abundance $(N_{(i,j)})$. Equations were derived with observations of five selected bacterial isolates represented in the molecular survey.

Equations	Multiple R ² (p-value)	Residual
$A_{(i,j)} = 0.49(\pm 0.15)\tau_{(i,j)} - 0.080(\pm 0.027)N_{(i,j)} - 0.27(\pm 0.10)$	0.69 (0.02)	0.08
$A_{(i,j)} = 0.49(\pm 0.16)\tau_{(i,j)} - 0.27(\pm 1.32)PD_{(i,j)} - 0.09(\pm 0.06)N_{(i,j)} - 0.23(\pm 0.21)$	0.69 (0.06)	0.08
$A_{(i,j)} = 0.44(\pm 0.17)\tau_{(i,j)} + 1.4(\pm 0.70)PD_{(i,j)} - 0.46(\pm 0.16)$	0.57 (0.05)	0.09

2.7 Discussion

The low cost and convenience of high-throughput sequencing technologies have contributed to facilitating our ability to characterize microbial communities in the environment. It is expected that investigations into the biodiversity and interactions among the members of the microbial communities associated with food will lead to novel advanced biocontrol technologies to establish beneficial microorganisms protecting food against ripping and pathogens (Berg, 2015; Teplitski et al., 2011). This approach has been widely applied to study microbial community dynamics in food fermentation and food processing environment, offering the potential to identify biomarkers for product quality (Bokulich et al., 2016). In contrast, very few attempts were made to characterize the microbiome of fresh food. A recent survey showed that fresh fruits and vegetables harbor distinct microbial community structures (Leff & Fierer, 2013). The dominance of Enterobacteriaceae lineage was noticed in sprouts, lettuce, spinach, peppers, and strawberry, distinguishing the bacterial community profile of these product types from those observed in apples, peaches, grapes, and mushrooms (Leff & Fierer, 2013). Prevalence of Enterobacteriaceae in fresh lettuce and tomatoes also was noticed in two independent studies (Jackson et al., 2013; Ottesen et al., 2013). From the best of our knowledge, this is the first report on the composition of microbial communities associated with ready-to-eat dry products. The number of OTU was in the same order of magnitude than fresh fruits and vegetables (Leff & Fierer, 2013; Wassermann et al., 2017), with bacterial and fungal OTU ranging between 7-40 and 10-40 per chia sample, respectively. Although only one survey of fungi is available, it showed the dominance of the fungal lineage Ascomycota in tomatoes (Ottesen et al., 2013). Our results are in agreement with this survey to suggest that species richness of fungal communities associated with chia is of the same magnitude than species richness of bacteria.

Even though there is an increasing number of reports on microbial species cooccurrence patterns in food (Chaillou *et al.*, 2014), only a few attempts were made to disentangle their mechanisms. Experimental validation of species interactions

suggested by co-occurrence patterns is complicated by the fact that functioning and metabolism of species examined in axenic culture can be significantly altered in the presence of other species (Ho *et al.*, 2014a). Nevertheless, both molecular survey and *in vitro* experiments demonstrated that *Paracoccus aminovorans* to promote the growth of *Vibrio cholerae* in human gut (Midani *et al.*, 2018), supporting the interest of co-cultures to validate potential interactions inferred through co-occurrence analyses. In this study, only five bacterial isolates were unambiguously assigned to OTU retrieved from the molecular survey. Except for *Enterobacter* sp. AJ7 representative of ubiquitous OTU 24 in the molecular survey, the isolates represented non-abundant OTU, precluding the application of standard arbitrary sample prevalence cutoffs to select a subset of OTU in co-occurrence profiles computing efforts. SparCC correlation matrix was thus computed using all OTU represented by a minimum of 8 reads as specified in paired-end sequence processing (see method section).

In contrast to approaches aimed at challenging potential microbe interaction in cocultures, we proposed a theoretical framework to conceptualize co-occurrence patterns by taking into account characteristics of isolates. The results support the hypothesis that relative abundance, phylogenetic distance and physiological traits explain observed cooccurrence patterns of microbial communities associated with chia seeds. The relevance of the proposed model is further supported through an application of the theoretical framework to co-occurrence pattern of cultivable bacterial OTU expressed using Spearman correlation coefficient instead of SparCC coefficient (Table S5). According to the model, dissimilarity of physiological traits increases the strength of cooccurrence. As expected owing to functional redundancy in bacteria (Ho et al., 2013; Krause et al., 2014), phylogenetic distance term with species abundance or in conjunction with both species abundance and physiological traits lowered model performances. Interpretation of the negative influence of abundance $(N_{(i,i)})$ on cooccurrence in the model is shaded by the compositional nature of OTU datasets, which do not allow inferring the concept of a relationship between species abundance and their interactions.

This study is the first attempt to test the relevance of physiological traits, phylogenetic distance, and relative abundance of bacterial species to explain their co-occurrence

patterns in molecular survey. At this stage, the results are suggestive owing to limited number and diversity of isolates representative of microbial OTUs detected in the molecular survey. Cultivation bias that led to the isolation of Proteobacteria and Firmicutes strains combined with the absence of fungi isolate impairs a sound evaluation of the relevance of phylogenetic distance in explaining species cooccurrence. Positive correlations were frequently observed for phylogenetically-close microbial species expected to share similar ecological traits in different habitats encompassing soil, lettuce and human gut (Barberan et al., 2012; Cardinale et al., 2015; Faust et al., 2012). The structure of microbial communities reported in this study represent the legacy of microbial successions that took place along the whole production chain of chia. Based on previous investigations on dairy products and meats, these successions are expected to be in part driven by filtering effects selecting species whose metabolism is compatible with physicochemical conditions prevailing in chia seeds such as low water activity, in addition to stochastic dissemination of environmental species across the different post-harvest stages including processing and distribution (Chaillou et al., 2014; Guidone et al., 2016). The former mechanism is supported by the observation of a core microbiome represented by two bacterial species affiliated to Burkholderiaceae and one fungus species affiliated Pleosporaceae in the 25 chia samples originated from different sources. Nevertheless, stochastic dissemination of microbial species is supported by a weak relationship between species distribution and chia origin as well as low connectivity of co-occurrence patterns which is a particularity distinguishing food from soil or host environments (Parente et al., 2016; Parente et al., 2018). Finding the exact origin of detected bacterial and fungi species is beyond the scope of this study, impairing a sound evaluation of spatial and temporal variations of co-occurrence patterns in chia. As a consequence, it remains unclear whether the role of physiological traits and abundance in deciphering observed bacterial species co-occurrence was the result of a filtering effect taking place in chia or in the environmental reservoir of microbes that contaminated chia during the whole supply chain. Despite these limitations, the approach we used should be considered in future investigations aimed at conceptualizing microbial species co-occurrence patterns.

2.8 Acknowledgements

This work has been supported by a Natural Sciences and Engineering Research Council of Canada Engage grant (493409-16) and a Natural Sciences and Engineering Research Council of Canada Engage Plus grant (508707-17) to PC. The authors are grateful to Sarah Piché-Choquette who introduced AJ to the bioinformatics tools utilized in this study and submitted raw sequence reads to the Sequence Read Archive repository (National Center for Biotechnology Information). The authors wish to acknowledge the contribution of the McGill University and Génome Québec Innovation Centre (Montréal, Canada) for PCR amplicon library preparation and sequencing services.

CONCLUSION ET PERSPECTIVES

Notre hypothèse testée pour la première fois en microbiologie consistait à démontrer que la covariation des espèces pourrait été expliquée par l'abondance relative, la taxonomie et les traits physiologiques de chaque espèce. Cette étude a permis de constater que l'abondance relative et la divergence des traits physiologiques de chaque phylotype bactérien cultivable justifiaient 69 % de la co-variation observées dans des profils moléculaires, ce qui implique l'importance de ces deux variables dans la cohabitation des espèces.

La combinaison entre l'approche moléculaire et l'isolement des phylotypes bactériens cultivables a permis de générer des données moléculaires représentant les communautés microbiennes associées au chia et des données expérimentales correspondant à chaque paire de phylotype bactérien cultivable (caractéristique physiologique). Ensuite, ces données ont été utilisées pour alimenter notre modèle mathématique de covariation. Le fait que la distance phylogénétique n'est pas ressortie comme un facteur décisif expliquant la covariation des espèces doit être pris avec une certaine réserve. En effet, cela pourrait être dû à notre collection d'isolats peu diversifié, dominé par des protéobactéries.

Un autre fait surprenant est l'alimentation du modèle mathématique de covariance par des comparaisons impliquant des espèces non dominantes, retrouvées dans moins de 40% des échantillons, qui sont habituellement écartées des analyses en raison des incertitudes entourant leur détection ou non détection. Puisque le modèle théorique élaboré à partir de ces valeurs de corrélations SparCC atypiques ait été validé, il remet en quelque sorte le choix arbitraire des seuils utilisés pour étudier les réseaux écologiques. Selon nos résultats, l'implication des espèces peu abondantes dans l'étude des réseaux de covariation devrait être revisitée.

Finalement, il serait intéressant dans l'avenir de tester ce même modèle pour les champignons, en vue de comparer les règles impliquées dans la covariance aux celles impliquées pour les bactéries. Ceci pourrait être réaliser en accentuant l'effort de l'isolement entrepris dans le cadre de cette maîtrise. Ce volet n'a pas été abordé dans

le mémoire, mais des essais d'isolement de champignons ont permis d'obtenir à peine deux espèces représentatives de champignons, dont aucune ne correspondait aux phylotypes détectés dans le profilage moléculaire. Par ailleurs, il sera toujours nécessaire de tester ce même modèle pour des autres aliments. La compréhension des mécanismes de covariation des espèces constitue un potentiel à exploiter pour minimiser les risques de contamination microbiologique des aliments.

REFERENCES

- Aitchison J (1982) The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B (Methodological)* :139-177.
- Aßhauer KP, Wemheuer B, Daniel R & Meinicke P (2015) Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics* 31(17):2882-2884.
- Barberan A, Bates ST, Casamayor EO & Fierer N (2012) Using network analysis to explore co-occurrence patterns in soil microbial communities. *The International Society for Microbial Ecology* 6(2):343-351.
- Bell G (2005a) The co-distribution of species in relation to the neutral theory of community ecology. *Ecology* 86(7):1757-1770.
- Bell G (2005b) The co-distribution of species in relation to the neutral theory of community ecology. *Ecology* 86(7):1757-1770.
- Berg G (2015) Beyond borders: investigating microbiome interactivity and diversity for advanced biocontrol technologies. *Microbial Biotechnology* 8(1):5-7.
- Berg G, Grube M, Schloter M & Smalla K (2014) Unraveling the plant microbiome: looking back and future perspectives. *Frontiers in Microbiology* 5:148.
- Blaser MJ, Cardon ZG, Cho MK, Dangl JL, Donohue TJ, Green JL, Knight R, Maxon ME, Northen TR & Pollard KS (2016) Toward a predictive understanding of Earth's microbiomes to address 21st century challenges. *The American Society for Microbiology*.
- Bokulich NA, Lewis ZT, Boundy-Mills K & Mills DA (2016) A new perspective on microbial landscapes within food production. *Current Opinion in Biotechnology* 37:182-189.
- Callahan BJ, McMurdie PJ & Holmes SP (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The International Society for Microbial Ecology*. Doi: 10.1038/ismej.2017.119.
- Cardinale M, Grube M, Erlacher A, Quehenberger J & Berg G (2015) Bacterial networks and co-occurrence relationships in the lettuce root microbiota. *Environmental Microbiology* 17(1):239-252.
- Cavender-Bares J, Keen A & Miles B (2006) Phylogenetic structure of Floridian plant communities depends on taxonomic and spatial scale. *Ecology* 87(sp7):S109-S122.
- Cavender-Bares J, Kozak KH, Fine PV & Kembel SW (2009) The merging of community ecology and phylogenetic biology. *Ecology Letters* 12(7):693-715.
- Chaillou S, Chaulot-Talmon A, Caekebeke H, Cardinal M, Christieans S, Denis C, Hélène Desmonts M, Dousset X, Feurer C, Hamon E, Joffraud J-J, La Carbona S, Leroi F, Leroy S, Lorre S, Macé S, Pilet M-F, Prévost H, Rivollier M, Roux D, Talon R, Zagorec M & Champomier-Vergès M-C (2014) Origin and ecological

selection of core and food-specific bacterial communities associated with meat and seafood spoilage. *The International Society for Microbial Ecology* 9:1105.

- Choi J, Yang F, Stepanauskas R, Cardenas E, Garoutte A, Williams R, Flater J, Tiedje JM, Hofmockel KS & Gelder B (2017) Strategies to improve reference databases for soil microbiomes. *The International Society for Microbial Ecology* 11(4):829.
- Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR & Tiedje JM (2014) Ribosomal database project: data and tools for high throughput rRNA analysis. *Nucleic Acids Research* 42(D1):D633-D642.
- Comeau AM, Li WKW, Tremblay J-É, Carmack EC & Lovejoy C (2011) Arctic ocean microbial community structure before and after the 2007 record sea ice minimum. *PLOS ONE* 6(11):e27492.
- Connor EF & Simberloff D (1979) The assembly of species communities: chance or competition? *Ecology* 60(6):1132-1140.
- Cornwell WK, Schwilk DW & Ackerly DD (2006) A trait-based test for habitat filtering: convex hull volume. *Ecology* 87(6):1465-1471.
- Dauga C, Doré J & Sghir A (2005) La diversité insoupçonnée du monde microbien. *M/S: Médecine Sciences* 21(3):290-296.
- Dees MW, Lysøe E, Nordskog B & Brurberg MB (2014) Bacterial communities associated with surfaces of leafy greens: shift in composition and decrease in richness over time. *Applied and Environmental Microbiology* :03470-03414.
- Deshpande V, Wang Q, Greenfield P, Charleston M, Porras-Alfaro A, Kuske CR, Cole JR, Midgley DJ & Tran-Dinh N (2016) Fungal identification using a Bayesian classifier and the Warcup training set of internal transcribed spacer sequences. *Mycologia* 108(1):1-5.
- Edgar R (2016a) SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS sequences. *BioRxiv*. Doi: 10.1101/074161.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32(5):1792-1797.
- Edgar RC (2010) Search and clustering orders of magnitude faster than blast. *Bioinformatics* 26(19):2460-2461.
- Edgar RC (2016b) UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *BioRxiv*. Doi: 10.1101/081257.
- Ercolini D (2013) High-throughput sequencing and metagenomics: steps ahead in the culture-independent analysis of food microbial ecology. *Applied and Environmental Microbiology* :00256-00213.
- Erlacher A, Cardinale M, Grosch R, Grube M & Berg G (2014) The impact of the pathogen Rhizoctonia solani and its beneficial counterpart Bacillus amyloliquefaciens on the indigenous lettuce microbiome. *Frontiers in Microbiology* 5:175.

- Fang H, Huang C, Zhao H & Deng M (2015) CCLasso: correlation inference for compositional data through Lasso. *Bioinformatics* 31(19):3172-3180.
- Faust K & Raes J (2012) Microbial interactions: from networks to models. *Nature Reviews Microbiology* 10(8):538.
- Faust K & Raes J (2016) CoNet app: inference of biological association networks using Cytoscape. *F1000Research* 5.
- Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J & Huttenhower C (2012) Microbial co-occurrence relationships in the human microbiome. *PLOS Computational Biology* 8(7):e1002606.
- Fisher CK, Mora T & Walczak AM (2017) Variable habitat conditions drive species covariation in the human microbiota. *PLOS Computational Biology* 13(4):e1005435.
- Foster KR & Bell T (2012) Competition, not cooperation, dominates interactions among culturable microbial species. *Current biology* 22(19):1845-1850.
- Friedman J & Alm EJ (2012) Inferring correlation networks from genomic survey data. *PLoS Comp. Biol.* 8(9):e1002687.
- Fuhrman JA (2009) Microbial community structure and its functional implications. *Nature* 459(7244):193.
- Fukui M, Teske A, Aßmus B, Muyzer G & Widdel F (1999) Physiology, phylogenetic relationships, and ecology of filamentous sulfate-reducing bacteria (genus Desulfonema). *Archives of Microbiology* 172(4):193-203.
- Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, Segrè D, Mehta P & Sanchez A (2018) Emergent simplicity in microbial community assembly. *Science* 361(6401):469-474.
- Gotelli NJ & McCabe DJ (2002) Species co-occurrence: a meta-analysis of JM Diamond's assembly rules model. *Ecology* 83(8):2091-2096.
- Guidone A, Zotta T, Matera A, Ricciardi A, De Filippis F, Ercolini D & Parente E (2016) The microbiota of high-moisture mozzarella cheese produced with different acidification methods. *International Journal of Food Microbiology* 216:9-17.
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series*. [London]: Information Retrieval Ltd., c1979-c2000., p 95-98.

Ho A, De Roy K, Thas O, De Neve J, Hoefman S, Vandamme P, Heylen K & Boon N (2014b) The more, the merrier: heterotroph richness stimulates methanotrophic activity. *The International Society for Microbial Ecology* 8(9):1945.

Ho A, Kerckhof F-M, Luke C, Reim A, Krause S, Boon N & Bodelier PLE (2013) Conceptualizing functional traits and ecological characteristics of methaneoxidizing bacteria as life strategies. *Environmental Microbiology Reports* 5(3):335-345.

- Horner-Devine MC, Silver JM, Leibold MA, Bohannan BJ, Colwell RK, Fuhrman JA, Green JL, Kuske CR, Martiny JB & Muyzer G (2007) A comparison of taxon cooccurrence patterns for macro-and microorganisms. *Ecology* 88(6):1345-1353.
- Hoy CW (2015) Agroecosystem health, agroecosystem resilience, and food security. *Journal of Environmental Studies and Sciences* 5(4):623-635.
- Huang L (2017) Dynamic identification of growth and survival kinetic parameters of microorganisms in foods. *Current Opinion in Food Science* 14:85-92.
- Jackson CR, Randolph KC, Osborn SL & Tyler HL (2013) Culture dependent and independent analysis of bacterial communities associated with commercial salad leaf vegetables. *BMC Microbiology* 13(1):274.
- Kamneva OK (2017) Genome composition and phylogeny of microbes predict their cooccurrence in the environment. *PLOS Computational Biology* 13(2):e1005366.
- Kaul A, Mandal S, Davidov O & Peddada SD (2017) Analysis of microbiome data in the presence of excess zeros. *Frontiers in Microbiology* 8(2114).
- Kembel SW & Hubbell SP (2006) The phylogenetic structure of a neotropical forest tree community. *Ecology* 87(sp7):S86-S99.
- Kim M-S, Bae J-W & Park E-J (2018) Geographic and host-associated variations in bacterial communities on the floret surfaces of field-grown broccoli. *Applied and Environmental Microbiology* : 02837-02817.
- Koch AL (2001) Oligotrophs versus copiotrophs. *Bioessays* 23(7):657-661.
- Koyama K, Hokunan H, Hasegawa M, Kawamura S & Koseki S (2017) Modeling stochastic variability in the numbers of surviving Salmonella enterica, enterohemorrhagic Escherichia coli, and Listeria monocytogenes cells at the single-cell level in a desiccated environment. *Applied and Environmental Microbiology* 83(4):e02974-02916.
- Krause S, Le Roux X, Niklaus PA, Van Bodegom PM, Lennon JT, Bertilsson S, Grossart H-P, Philippot L & Bodelier PLE (2014) Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. *Frontiers in Microbiology* 5(251).
- Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ & Bonneau RA (2015) Sparse and compositionally robust inference of microbial ecological networks. *PLOS Computational Biology* 11(5):e1004226.
- Lane D (1991) 16S/23S rRNA sequencing. John Wiley & Sons,
- Lederberg J & McCray AT (2001) Ome sweet omics-A genealogical treasury of words. *The Scientist* 15(7):8-8.
- Leekitcharoenphon P, Friis C, Zankari E, Svendsen CA, Price LB, Rahmani M, Herrero-Fresno A, Fashae K, Vandenberg O & Aarestrup FM (2013) Genomics of an emerging clone of salmonella serovar typhimurium ST313 from Nigeria and the Democratic Republic of Congo. *The Journal of Infection in Developing Countries* 7(10):696-706.

- Leff JW & Fierer N (2013) Bacterial Communities Associated with the Surfaces of Fresh Fruits and Vegetables. *PLOS ONE* 8(3):e59310.
- Little AE, Robinson CJ, Peterson SB, Raffa KF & Handelsman J (2008) Rules of engagement: interspecies interactions that regulate microbial communities. *Annual Review of Microbiology* 62:375-401.
- Locey KJ & Lennon JT (2016) Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences* 113(21):5970-5975.
- MacArthur R & Levins R (1967) The limiting similarity, convergence, and divergence of coexisting species. *The American Naturalist* 101(921):377-385.
- Malakar PK, Barker GC, Zwietering MH & van't Riet K (2003) Relevance of microbial interactions to predictive microbiology. *International Journal of Food Microbiology* 84(3):263-272.
- Mandakovic D, Rojas C, Maldonado J, Latorre M, Travisany D, Delage E, Bihouée A, Jean G, Díaz FP, Fernández-Gómez B, Cabrera P, Gaete A, Latorre C, Gutiérrez RA, Maass A, Cambiazo V, Navarrete SA, Eveillard D & González M (2018) Structure and co-occurrence patterns in microbial communities under acute environmental stress reveal ecological factors fostering resilience. *Scientific Reports* 8(1):5875.
- Manhart M & Shakhnovich E (2018) Growth tradeoffs produce complex microbial communities on a single limiting resource. *BioRxiv* :266569.
- McMurdie PJ & Holmes S (2013) Phyloseq: An R package for reproducible Interactive analysis and graphics of microbiome census data. *PLOS ONE* 8(4):e61217.
- Mejlholm O, Gunvig A, Borggaard C, Blom-Hanssen J, Mellefont L, Ross T, Leroi F, Else T, Visser D & Dalgaard P (2010) Predicting growth rates and growth boundary of Listeria monocytogenes — An international validation study with focus on processed and ready-to-eat meat and seafood. *International Journal of Food Microbiology* 141(3):137-150.
- Midani FS, Weil AA, Chowdhury F, Begum YA, Khan AI, Debela MD, Durand HK, Reese AT, Nimmagadda SN, Silverman JD, Ellis CN, Ryan ET, Calderwood SB, Harris JB, Qadri F, David LA & LaRocque RC (2018) Human gut microbiota predicts susceptibility to Vibrio cholerae infection. *The Journal of Infectious Diseases* 218(4):645-653.
- Nemergut DR, Schmidt SK, Fukami T, O'Neill SP, Bilinski TM, Stanish LF, Knelman JE, Darcy JL, Lynch RC, Wickey P & Ferrenberg S (2013) Patterns and processes of microbial community assembly. *Microbiology and Molecular Biology Reviews* 77(3):342-356.
- Ochman H, Lawrence JG & Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* 405(6784):299.
- Oksanen J, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, Simpson G, Solymos P, Henry M, Stevens H & Wagner H (2012) Vegan: community ecology package. R package version 2.0-4. <u>http://cran.r-project.org/package=vegan</u>

- Ottesen AR, González Peña A, White JR, Pettengill JB, Li C, Allard S, Rideout S, Allard M, Hill T, Evans P, Strain E, Musser S, Knight R & Brown E (2013) Baseline survey of the anatomical microbial ecology of an important food plant: Solanum lycopersicum (tomato). *BMC Microbiology* 13(1):114.
- Parente E, Cocolin L, De Filippis F, Zotta T, Ferrocino I, O'Sullivan O, Neviani E, De Angelis M, Cotter PD & Ercolini D (2016) FoodMicrobionet: A database for the visualisation and exploration of food bacterial communities based on network analysis. *International Journal of Food Microbiology* 219:28-37.
- Parente E, Zotta T, Faust K, De Filippis F & Ercolini D (2018) Structure of association networks in food bacterial communities. *Food Microbiology* 73:49-60.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P & Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research* :gr. 186072.186114.
- Poisot T, Stouffer DB & Gravel D (2015) Beyond species: why ecological interaction networks vary through space and time. *Oikos* 124(3):243-251.
- Salgado-Cruz MdIP, Calderón-Domínguez G, Chanona-Pérez J, Farrera-Rebollo RR, Méndez-Méndez JV & Díaz-Ramírez M (2013) Chia (*Salvia hispanica* L.) seed mucilage release characterisation. A microstructural and image analysis study. *Industrial Crops and Products* 51:453-462.
- Schink B (2002) Synergistic interactions in the microbial world. *Antonie Van Leeuwenhoek* 81(1-4):257-261.
- Stahl DA, Lane DJ, Olsen GJ & Pace NR (1985) Characterization of a Yellowstone hot spring microbial community by 5S rRNA sequences. *Applied and Environmental Microbiology* 49(6):1379-1384.
- Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, Iyer R, Schatz MC, Sinha S & Robinson GE (2015) Big data: astronomical or genomical? *PLOS Biology* 13(7):e1002195.
- Tamber S, Swist E & Oudit D (2016) Physicochemical and bacteriological characteristics of organic sprouted chia and flax seed powders implicated in a foodborne salmonellosis outbreak. *Journal of Food Protection*® 79(5):703-709.
- Tamplin ML (2018) Integrating predictive models and sensors to manage food stability in supply chains. *Food Microbiology* 75:90-94.
- Tamura K & Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* 10(3):512-526.
- Tamura K, Stecher G, Peterson D, Filipski A & Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30(12):2725-2729.
- Team RDC (2008) *R*: A language and environment for statistical computing. *R* Foundation for Statistical Computing (Consulté e Data Accessed)

- Tenenhaus-Aziza F & Ellouze M (2015) Software for predictive microbiology and risk assessment: a description and comparison of tools presented at the ICPMF8 Software Fair. *Food Microbiology* 45:290-299.
- Teplitski M, Warriner K, Bartz J & Schneider KR (2011) Untangling metabolic and communication networks: interactions of enterics with phytobacteria and their implications in produce safety. *Trends in Microbiology* 19(3):121-127.
- Toju H, Peay KG, Yamamichi M, Narisawa K, Hiruma K, Naito K, Fukuda S, Ushio M, Nakaoka S & Onoda Y (2018) Core microbiomes for sustainable agroecosystems. *Nature Plants* 4:247-257.
- Toju H, Tanabe AS, Yamamoto S & Sato H (2012) High-coverage ITS primers for the DNA-based identification of Ascomycetes and Basidiomycetes in environmental samples. *PLOS ONE* 7(7):e40863.
- Tsilimigras MC & Fodor AA (2016) Compositional data analysis of the microbiome: fundamentals, tools, and challenges. *Annals of epidemiology* 26(5):330-335.
- Turner S, Pryer KM, Miao VP & Palmer JD (1999) Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *Journal of Eukaryotic Microbiology* 46(4):327-338.
- Tyson GH, McDermott PF, Li C, Chen Y, Tadesse DA, Mukherjee S, Bodeis-Jones S, Kabera C, Gaines SA & Loneragan GH (2015) WGS accurately predicts antimicrobial resistance in Escherichia coli. *Journal of Antimicrobial Chemotherapy* 70(10):2763-2769.
- Vellend M (2010) Conceptual synthesis in community ecology. *The Quarterly Review of Biology* 85(2):183-206.
- Veraart A, Garbeva P, Beersum Fv, Ho A, Hordijk C, Meima-Franke M, Zweers A & Bodelier P (2018) Living apart together—bacterial volatiles influence methanotrophic growth and activity. *The International Society for Microbial Ecology* :1.
- Wang J, Koseki S, Chung M-J & Oh D-H (2017) A novel approach to predict the growth of Staphylococcus aureus on Rice Cake. *Frontiers in Microbiology* 8:1140.
- Wassermann B, Rybakova D, Müller C & Berg G (2017) Harnessing the microbiomes of Brassica vegetables for health issues. *Scientific Reports* 7(1):17649.
- Watts SC, Ritchie SC, Inouye M & Holt KE (2018) FastSpar: rapid and scalable correlation estimation for compositional data. *BioRxiv* :272583.
- Webb CO (2000) Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. *The American Naturalist* 156(2):145-155.
- Weiblen GD, Webb CO, Novotny V, Basset Y & Miller SE (2006) Phylogenetic dispersion of host use in a tropical insect herbivore community. *Ecology* 87(sp7):S62-S75.
- Whitman WB, Coleman DC & Wiebe WJ (1998) Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences* 95(12):6578-6583.

- Xiao Y, Angulo MT, Friedman J, Waldor MK, Weiss ST & Liu Y-Y (2017) Mapping the ecological networks of microbial communities. *Nature Communications* 8(1):2042.
- Xu J (2006) Invited review: microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Molecular Ecology* 15(7):1713-1731.
- Zhou J, Deng Y, Luo F, He Z, Tu Q & Zhi X (2010) Functional molecular ecological networks. *MBIO* 1(4):e00169-00110.
- Zoellner C, Al-Mamun MA, Grohn Y, Jackson P & Worobo R (2018) Post-harvest supply chain with microbial travelers: a novel farm-to-retail microbial simulation and visualization framework. *Applied and Environmental Microbiology*. Doi: 10.1128/aem.00813-18.