UNIVERSITÉ DU QUÉBEC À MONTRÉAL

METHYLOBACTERIUM DANS LA PHYLLOSPHÈRE : L'IMPORTANCE ÉCOLOGIQUE ET ÉVOLUTIVE DES INTERACTIONS ENTRE BACTÉRIES ET PLANTES HÔTES

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN BIOLOGIE

PAR

JOCELYN LAUZON

MARS 2025

REMERCIEMENTS

Je remercie chaleureusement mon directeur de recherche et mentor, le professeur Steven W. Kembel, pour sa présence, sa disponibilité et son ouverture d'esprit, pour la confiance qu'il m'a accordée, pour sa guidance et sa clarté de réflexion qui m'ont permis de naviguer relativement droit à travers mes nombreuses élucubrations, et pour son soutien indéfectible envers mon cheminement académique. Apprendre à devenir chercheur auprès du Pr Kembel a été une expérience fort enrichissante et je suis immensément reconnaissant d'avoir eu cette opportunité. Je remercie grandement mon codirecteur, Jean-Baptiste Leducq, professeur à l'Université Laval, pour m'avoir transmis ses connaissances et sa passion pour *Methylobacterium*, pour son implication dans mon projet de recherche, ainsi que pour ses nombreux conseils et enseignements sur la microbiologie et les approches omiques. Je remercie aussi les professeures Karine Dufresne (UQAM) et Marie Filteau (ULaval), membres de mon comité d'évaluation, pour leurs suggestions qui ont contribué à élever la qualité du présent manuscrit. Je tiens également à remercier tous.tes les membres du laboratoire Kembel que j'ai côtoyé.e.s durant mon (trop bref) passage, tout particulièrement Zihui Wang, Sarah Ishak, Sarah Piché-Choquette, David Ross, Élanore Favron et Geneviève Bourret. Merci pour votre écoute attentionnée, votre support indispensable en laboratoire, vos suggestions constructives sur mes analyses et votre amitié.

Je remercie la Fondation de l'UQAM et l'ensemble des donateurs et donatrices – particulièrement Mme D'Amour et M. Bredin, ainsi que la Fondation J.A. DeSève –, pour le précieux appui financier, moral et professionnel apporté par leurs généreuses bourses. Je reconnais également le soutien financier du Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) et du Fonds de recherche du Québec secteur Nature et technologies (FRQNT) pour la réalisation de mon projet de maîtrise. J'aimerais aussi remercier chaleureusement tous.tes les professeur.e.s et chargé.e.s de cours du Département des sciences biologiques de l'UQAM pour le partage de leurs savoirs et de leurs passions pour la biologie et l'écologie. Un grand merci aussi à Chantal Oligny, conseillère d'orientation à l'UQAM, pour son écoute et son empathie, et pour m'avoir aider à prendre d'importantes décisions.

Je remercie enfin mes parents, Julie et Sylvain, pour votre amour inconditionnel et pour avoir instillé en moi la confiance personnelle en mes capacités, confiance qui m'aura été essentielle pour entreprendre des études universitaires dans un domaine qui m'était, il y a six ans, totalement inconnu. De plus, sans votre présence et votre soutien, notamment en tant que super-mamie et

super-papi, je n'aurais jamais pu compléter cette maîtrise en deux ans. Je remercie également mes beaux-parents, Monique et Richard, pour leur support et pour les discussions enflammées sur la science et sa place dans l'expérience humaine.

Joëlle, sans toi, ce mémoire n'aurait jamais été écrit. En m'offrant mes premières jumelles, tu as ouvert mes yeux, ma tête et mon cœur sur la biodiversité qui m'entourait. Tu m'as aidé à focaliser ma soif inépuisable d'apprendre, de comprendre et de découvrir. Tu m'as écouté discourir sur les bactéries et les plantes tout comme sur mes angoisses existentielles, et tu m'as convaincu de poursuivre le chemin que je me traçais. Tu continues de me nourrir quotidiennement de ta propre curiosité pour le monde, juste ce qu'il faut de candide pour continuer de rêver tout en avançant. Merci infiniment. Flore, tu es ma source ultime de joie, de fous rires, de créativité et d'amour. Merci d'avoir ponctué mes analyses et ma rédaction par tes jeux de personnages et tes chansons saugrenues, ainsi que pour toutes les délicieuses crèmes glacées mangées sur des licornes dans les nuages toutes sortes de couleurs avec toi.

DÉDICACE

À Flore et Joëlle

Ils s'agrippaient aux branches des arbres
comme s'ils étaient en danger,
comme si les arbres étaient leurs mères.
Et les arbres, ne pouvant s'empêcher d'être apprivoisés,
se prenaient à désirer tenir des enfants dans leurs bras.

– Heather O'Neill, Perdre la tête
(traduction de Dominique Fortier)

AVANT-PROPOS

Mon mémoire porte sur l'écologie et l'évolution microbiennes. Mes travaux ont pour écosystème d'étude le microbiome de la phyllosphère – plus précisément, les bactéries vivant à la surface des feuilles des plantes. J'ai axé ma recherche sur *Methylobacterium*, un genre bactérien clé de la phyllosphère qui stimule la croissance végétale, protège ses hôtes contre les pathogènes, et joue un rôle fondamental dans le fonctionnement des écosystèmes terrestres. Le cœur de ce mémoire est composé de deux articles scientifiques, chacun faisant l'objet d'un chapitre distinct (chapitres 1 et 2), qui abordent de manière complémentaire les relations écologiques et évolutives entre la biodiversité de *Methylobacterium* au sein de la phyllosphère et ses plantes hôtes. Pour chacun de ces articles, j'ai réalisé une revue de la littérature, élaboré les objectifs et les hypothèses de recherche, conçu le plan expérimental, et effectué l'échantillonnage, le travail de laboratoire, les analyses bio-informatiques et statistiques ainsi que la rédaction en tant que premier auteur.

La première section consiste en une introduction générale qui met en contexte ma recherche dans le cadre théorique actuel en présentant l'état des connaissances scientifiques sur les principaux thèmes abordés dans mon mémoire. Les objectifs et les hypothèses de mes deux articles y sont énoncées.

Le chapitre 1 consiste en un premier article intitulé *Host tree species drive Methylobacterium ecology and evolution in the northern temperate forest*. Ce chapitre vise à comprendre le rôle des espèces d'arbres et d'arbustes hôtes dans l'écologie et l'évolution du genre *Methylobacterium* au niveau des espèces et de leurs gènes, ainsi qu'à révéler les bases génétiques des adaptations à l'hôte tout en évaluant le rôle du pangénome accessoire dans ces adaptations. Pour ce faire, j'ai combiné l'étude du pangénome du genre *Methylobacterium* avec l'analyse des métagénomes environnementaux de la phyllosphère. Steven W. Kembel et Jean-Baptiste Leducq, respectivement mes directeur et codirecteur de maîtrise, ont collaboré à la conception du projet de recherche et sont co-auteurs sur cet article en préparation.

Le chapitre 2 consiste en un second article intitulé *Richness and composition of phyllosphere Methylobacterium communities cause variation in Arabidopsis thaliana growth*. Ce chapitre vise à comprendre le rôle de la biodiversité microbienne dans la productivité des plantes hôtes. Pour ce faire, j'ai réalisé une expérience en laboratoire dans laquelle j'ai assemblé puis inoculé sur la plante modèle *Arabidopsis thaliana* des communautés bactériennes synthétiques composées de

diverses souches de *Methylobacterium.* L'article compte quatre co-auteurs. Jérémie Pelletier, un collègue étudiant à la maîtrise en biologie, a contribué aux analyses statistiques et l'interprétation écologique des résultats. Élanore Favron, une stagiaire de 1er cycle, et Dr ZIhui Wang ont participé à l'expérience en laboratoire. Steven W. Kembel a collaboré à la conception du projet de recherche et à l'édition du manuscrit ici présenté. Cet article a été soumis au journal *Oikos.*

Le chapitre 3 consiste en une conclusion générale dans laquelle je synthétise et mets en relation les résultats et leurs interprétations tirées des deux articles. Les limites de mon étude y sont indiquées et des pistes de recherche y sont proposées pour orienter les futurs travaux portant sur *Methylobacterium* et la phyllosphère.

TABLE DES MATIÈRES

LISTE DES FIGURES

LISTE DES TABLEAUX

LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

ABBA	*Abies balsamea*
ACSA	*Acer saccharum*
ADN	Acide désoxyribonucléique
ANOVA	Analysis of variance
ARNr	Acide ribonucléique ribosomal
BAM	Binary Alignment Map
C1	Composé à un atome de carbone
COCO	*Corylus cornuta*
COG	Clusters of Orthologous Groups
DAGC	Differentially abundant gene cluster
DNA	Deoxyribonucleic acid
FAGR	*Fagus grandifolia*
KEGG	Kyoto Encyclopedia of Genes and Genomes
HGT	Horizontal gene transfer
IAA	Acide indole 3-acétique
LDMC	Leaf dry matter content
MAG	Metagenome-assembled genome
MPD	Mean pairwise distance
MSH	Mont-Saint-Hilaire
PAA	Photosynthèse anoxygénique en aérobiose
PCoA	Principal coordinates analysis
PERMANOVA	Permutational analysis of variance
pN	non-synonymous polymorphism
pS	synonymous polymorphism
SBL	Station de biologie des Laurentides
SCV	Single codon variant
SES_{MPD}	Standardized effect size of mean pairwise distance
THOC	*Thuja occidentalis*

LISTE DES SYMBOLES ET DES UNITÉS

°C	degree Celsius
bp	base pairs
K_2HPO_4	dipotassium phosphate
KH_2PO_4	monopotassium phosphate
M	mol per litre
mm	millimetre
ng	nanogram
rpm	rotation per minute
µL	microlitre

RÉSUMÉ

La phyllosphère est un des plus vastes habitats terrestres pour les microorganismes; elle consiste en l'ensemble des parties aériennes des plantes. Les surfaces des feuilles abritent des communautés microbiennes très diversifiées, principalement dominées par les bactéries. Celles-ci jouent des rôles écologiques clés au sein des écosystèmes. Elles peuvent causer des maladies tout comme promouvoir la croissance et la santé de leur hôte, et elles participent aux cycles biogéochimiques et à la régulation du climat. Un lien positif a même été tracé entre la diversité bactérienne foliaire et la productivité des écosystèmes forestiers, bien que la relation de causalité n'ait pas été démontrée jusqu'ici. La phyllosphère a été étudié chez tous les types de plantes et de biomes, incluant chez les arbres de la forêt tempérée, qui couvre près de 7 millions km^2 de la surface terrestre. *Methylobacterium* est un genre bactérien omniprésent de la phyllosphère, qui compte plus d'une centaine d'espèces. Ces bactéries méthylotrophes sont notamment reconnues pour stimuler la croissance de leurs plantes hôtes par la sécrétion de phytohormones, ce qui fait d'elles des taxons clés dans le fonctionnement des écosystèmes. L'analyse du pangénome de *Methylobacterium* – l'ensemble des gènes de toutes les espèces – a mis en lumière l'histoire évolutive du genre, ainsi que ses principales capacités métaboliques. Au sein de la phyllosphère, la feuille de chaque espèce végétale possède des caractéristiques morphologiques et phytochimiques uniques qui agissent comme un filtre environnemental pour les bactéries. L'espèce de plante hôte est en effet un des facteurs écologiques les plus influents sur la structure des communautés de *Methylobacterium*. Or, on ne connaît toujours pas les gènes bactériens qui sous-tendent les adaptations à différentes espèces d'hôtes.

Le but de mon mémoire était, d'une part, de mieux comprendre comment l'espèce hôte façonne l'écologie et l'évolution de *Methylobacterium* au sein de la phyllosphère de la forêt tempérée, et d'autre part, d'investiguer la relation causale entre la biodiversité bactérienne de la phyllosphère et la croissance des plantes hôtes. Plus précisément, les objectifs de mon premier chapitre consistaient à quantifier l'influence de l'espèce hôte sur les structures taxonomique et génétique des communautés de *Methylobacterium*; identifier les bases géniques des adaptations de *Methylobacterium* aux espèces hôtes; et évaluer la nature adaptative du pangénome accessoire de *Methylobacterium*. Les objectifs de mon deuxième chapitre étaient de clarifier le lien causal entre richesse bactérienne foliaire et productivité de l'hôte; de révéler le rôle des souches bactériennes dans la relation diversité-productivité; et d'évaluer la contribution relative des différents mécanismes écologiques sous-tendant cette relation dans la phyllosphère.

Dans le cadre de mon premier chapitre, en combinant l'analyse des métagénomes des communautés foliaires d'une forêt tempérée du Québec méridional avec l'analyse du pangénome de *Methylobacterium*, j'ai révélé que l'espèce de plante hôte façonnait grandement la structure des communautés bactériennes et de *Methylobacterium*. Les associations dévoilées entre gènes et espèces hôtes suggèrent que la lumière du soleil et la phénologie des feuilles exercent des pressions de sélection fortes qui entraînent d'importantes divergences dans la composition taxonomique et génétique des communautés résidant respectivement sur les arbres et les arbustes, et sur les conifères et les feuillus. Une partie du pangénome accessoire de *Methylobacterium* semble jouer un rôle dans les adaptations à l'hôte. Dans le cadre de mon deuxième chapitre, à l'aide de communautés synthétiques composées de souches de *Methylobacterium*, j'ai révélé qu'une augmentation de la richesse bactérienne de la phyllosphère entraînait globalement une augmentation de la croissance de l'hôte, que les souches pouvaient avoir un effet positif, neutre ou négatif sur l'hôte et que la composition des communautés façonnait la relation diversité-productivité, en raison de la complémentarité des niches.

Mon mémoire met en lumière l'importance des interactions entre *Methylobacterium* et ses plantes hôtes. En liant les gènes à l'écologie de *Methylobacterium*, mes travaux soulignent l'apport fondamental de la biodiversité floristique dans l'écologie et l'évolution des bactéries, tout comme le rôle clé de leurs communautés phyllosphériques dans le fonctionnement des écosystèmes terrestres. Les hypothèses écologiques qui ont émergé de mon étude serviront à guider la recherche future sur *Methylobacterium* et la phyllosphère, un écosystème microbien au fort potentiel pour augmenter la résilience des cultures et des forêts face aux changements globaux.

Mots clés : écologie microbienne, phyllosphère, *Methylobacterium*, métagénome, pangénome, gènes, adaptation, biodiversité-productivité, communautés synthétiques.

ABSTRACT

The phyllosphere is among the largest terrestrial habitats for microorganisms; it consists of all the aerial parts of plants. Foliar surfaces harbour diverse microbial communities, mostly dominated by bacteria. These play key ecological roles within ecosystems. Bacteria can cause diseases as well as promote host growth and health, and they participate in biogeochemical cycles and in climate regulation. A positive correlation has even been established between foliar bacterial diversity and forest ecosystem productivity, although the causal relationship has not yet been demonstrated. The phyllosphere has been studied in all types of plants and biomes, including in trees of the temperate forest, which covers almost 7 million km^2 of the globe's surface. *Methylobacterium* is a ubiquitous bacterial genus of the phyllosphere, and is composed of over a hundred species. These methylotroph bacteria are notably known to stimulate the growth of their host plants by secreting phytohormones, which makes them key taxa regarding ecosystem functioning. The analysis of *Methylobacterium's* pangenome – the totality of genes in all species – has shed light on the evolutionary history of the genus, and on its principal metabolic capacities. In the phyllosphere, the leaf of each host plant possesses unique morphological and phytochemical characteristics which exert an ecological filter on bacteria. Host species is indeed one of the most influential factors on *Methylobacterium* community structure. However, we still do not know which bacterial genes underlie the adaptations to different host species.

The aim of this master's thesis was, on the one hand, to gain a better understanding of how host species shape *Methylobacterium*'s ecology and evolution in the phyllosphere, and on the other hand, to investigate the causal relation between phyllosphere bacterial biodiversity and host growth. More precisely, the objectives of my first chapter were to quantify the influence of host species on taxonomic and genetic structures of *Methylobacterium* communities; to identify the genetic basis of *Methylobacterium*'s host species adaptations; and to evaluate the adaptive nature of the accessory pangenome of *Methylobacterium*. The objectives of my second chapter were to clarify the causal relation between foliar bacterial diversity and host productivity; to reveal the role of bacterial strains in the biodiversity-productivity relationship; and to evaluate the relative contribution of different ecological mechanisms underlying this relationship in the phyllosphere.

In my first chapter, by combining the analysis of metagenomes from phyllosphere communities of a temperate forest in southern Quebec with the analysis of *Methylobacterium*'s pangenome, I revealed that host species extensively shaped the structure of bacterial and *Methylobacterium* communities. The associations found between genes and host species suggest that sunlight and leaf phenology exert strong selection pressures that generate important divergence patterns in the taxonomic and genetic composition of communities residing respectively on trees and shrubs, and on conifer and broadleaf species. Part of the accessory pangenome of *Methylobacterium* appeared to play a role in host species adaptations. In my second chapter, by designing and inoculating synthetic communities composed of *Methylobacterium* strains, I revealed that an increase in phyllosphere bacterial richness ultimately lead to an increase in host growth, that strains can have a positive, neutral or negative effect on host, and that community composition shaped the diversity-productivity relationship, due to niche complementarity.

My master's thesis sheds light on the importance of interactions between *Methylobacterium* and its host plants. By linking *Methylobacterium*'s genes and ecology, my research emphasizes the fundamental contribution of floristic biodiversity in bacterial ecology and evolution, as well as the key role of phyllosphere communities in terrestrial ecosystem functioning. The ecological

hypotheses that emerged from my study will guide future research efforts on *Methylobacterium* and the phyllosphere, a promising microbial ecosystem to increase the resilience of crops and forests in the face of global changes.

Keywords : microbial ecology, phyllosphere, *Methylobacterium*, metagenome, pangenome, genes, adaptation, biodiversity-productivity, synthetic communities.

INTRODUCTION

L'objectif global de mon mémoire est d'enrichir la compréhension des implications écologiques et évolutives des associations symbiotiques entre *Methylobacterium* et les plantes hôtes au sein de la phyllosphère. Cette introduction expose sous forme de revue de la littérature les connaissances scientifiques actuelles sur les principales thématiques liées à mon écosystème d'étude : la phyllosphère; le genre *Methylobacterium*; la pangénomique; la relation entre la biodiversité et le fonctionnement des écosystèmes; ainsi que les communautés synthétiques. La présentation de ce cadre théorique sert à contextualiser les travaux faisant l'objet des chapitres 1 et 2, dont les questions, objectifs et hypothèses de recherche sont énoncées à la fin du présent chapitre.

0.1 La phyllosphère

0.1.1 Un habitat microbien

La phyllosphère est un type d'habitat pour les microorganismes qui consiste en la partie aérienne des plantes, principalement les feuilles (Last, 1955; Ruinen, 1956). Bien que les levures et champignons filamenteux y forment parfois de riches assemblages (Jager *et al.*, 2001; Kembel et Mueller, 2014; Osono, 2006), les communautés microbiennes de la phyllosphère sont principalement dominées par les bactéries (Vorholt, 2012). À l'échelle planétaire, la surface foliaire des végétaux représente une superficie totale estimée à 640 millions km^2, qui serait habitée par environ 10^{26} cellules bactériennes (Lindow et Brandl, 2003; Morris et Kinkel, 2002).

0.1.2 Fonctions écologiques des bactéries foliaires

Les bactéries de la phyllosphère soutiennent plusieurs fonctions écologiques et écosystémiques. Bien que certaines espèces bactériennes entraînent des maladies chez les plantes (Baker *et al.*, 2010; Xin *et al.*, 2018), d'autres espèces promeuvent la croissance végétale via la sécrétion de phytohormones (Pattnaik *et al.*, 2017) ou l'acquisition de nutriments (Batool *et al.*, 2016). De plus, certaines espèces bactériennes protègent leur hôte contre des pathogènes fongiques ou bactériens via la compétition pour les ressources nutritionnelles (Innerebner *et al.*, 2011) et l'espace (Ritpitakphong *et al.*, 2016), la sécrétion de molécules antibiotiques ou l'activation des mécanismes de défense de la plante (De Mandal et Jeon, 2023). Enfin, certaines bactéries augmentent la tolérance des plantes aux stress abiotiques (Wang *et al.*, 2023b). À l'échelle des écosystèmes, les bactéries foliaires jouent un rôle clé dans les cycles biogéochimiques du carbone et de l'azote (Fürnkranz *et al.*, 2008; Moreira *et al.*, 2021; Yurimoto et Sakai, 2023), dans

la purification de l'atmosphère (Sandhu *et al.*, 2007), et dans la régulation du climat (Bringel et Couée, 2015). Elles participent ainsi potentiellement à moduler la réponse des écosystèmes terrestres aux changements globaux (Perreault et Laforest-Lapointe, 2022).

0.1.3 Facteurs structurant les communautés

Les communautés microbiennes de la phyllosphère ont été étudiées dans pratiquement toutes les formes de plantes et de biomes terrestres pour tenter de comprendre les dynamiques spatio-temporelles ainsi que les facteurs abiotiques et biotiques qui influencent leur diversité. Des études ont été menées sur des herbacées indigènes de l'Antarctique (Araya *et al.*, 2020), jusqu'aux arbres des forêts tropicales (Kembel *et al.*, 2014; Kim *et al.*, 2012; Lambais *et al.*, 2006), en passant par les arbustes xérophytes du désert (Liu *et al.*, 2023), les arbres de la forêt tempérée (DeBellis *et al.*, 2019; Laforest-Lapointe *et al.*, 2016a; Leducq *et al.*, 2022a), les arbres et les mousses de la forêt boréale (Ishak *et al.*, 2024; Rodríguez-Rodríguez *et al.*, 2023) et les plantes de culture dans les agroécosystèmes (Xiong et Lu, 2022).

Globalement, les principaux phyla bactériens qui dominent la phyllosphère sont *Pseudomonadota*, *Actinomycetota*, *Bacteriodota* et *Bacillota* (Bashir *et al.*, 2022; Sohrabi *et al.*, 2023). Au sein du phylum *Pseudomonadota*, parmi les genres les plus représentés on retrouve *Sphingomonas*, *Methylobacterium*, *Pseudomonas* et *Pantoea*. Au sein du phylum *Actinomycetota*, on retrouve entre autres les genres *Mycobacterium* et *Streptomyces* (Lindow et Brandl, 2003; Sohrabi *et al.*, 2023; Vorholt, 2012). La composition taxonomique des communautés bactériennes est influencée par divers facteurs, dont les plus documentés sont l'espèce hôte, le site et la saison (Kim *et al.*, 2012; Knief *et al.*, 2010; Laforest-Lapointe *et al.*, 2016a; Leducq *et al.*, 2022a; Liu *et al.*, 2023; Redford *et al.*, 2010; Redford et Fierer, 2009). Le génotype (Bodenhausen *et al.*, 2014) et les traits fonctionnels (Kembel *et al.*, 2014) de l'hôte, ainsi que la phylogénie de l'hôte et de ses symbiontes (Lajoie et Kembel, 2021) sont aussi des facteurs importants qui influencent la composition des communautés. D'autre part, des études ont révélé qu'une large portion de la diversité microbienne, tant d'un point de vue taxonomique (Kembel *et al.*, 2014; Laforest-Lapointe *et al.*, 2016a) que fonctionnel (Lajoie *et al.*, 2020), est similaire entre les feuilles des différentes espèces d'arbres. La recherche actuelle s'intéresse particulièrement à l'importance relative des processus de sélection et de dispersion dans la structuration des communautés (Wang *et al.*, 2023a), aux interactions entre microorganismes et aux taxons clés (Agler *et al.*, 2016; Carlström *et al.*, 2019), ainsi qu'à la relation entre la diversité bactérienne et la productivité des écosystèmes (Laforest-Lapointe *et al.*, 2017a).

0.1.4 Adaptations des bactéries épiphytes

Les bactéries épiphytes, celles vivant à la surface des feuilles, forment généralement des communautés plus diversifiées et abondantes que les bactéries endophytes qui vivent à l'intérieur des tissus foliaires. Ces dernières prolifèrent dans un habitat relativement plus stable et sont plus fortement soumises aux mécanismes défensifs de la plante (Sohrabi *et al.*, 2023). La surface des feuilles est un environnement globalement oligotrophe, caractérisé par une grande exposition au rayonnement ultraviolet (UV) et aux conditions météorologiques; la température et le taux d'humidité y fluctuent rapidement (Lindow et Brandl, 2003; Vorholt, 2012). Ces conditions de stress abiotiques exigent des bactéries épiphytes qu'elles possèdent des adaptations moléculaires et métaboliques particulières, notamment pour éviter la dessiccation ou le stress osmotique, acquérir des nutriments de la plante et se protéger du rayonnement UV (Sohrabi *et al.*, 2023; Vorholt, 2012). Pour éviter la dessiccation ou le stress osmotique, plusieurs bactéries phyllosphériques peuvent former des agrégats en produisant des polysaccharides extracellulaires qui les englobent dans un biofilm (Beattie, 2011), synthétiser ou importer des solutés compatibles (Freeman *et al.*, 2010) ou sécréter des biosurfactants qui augmentent la perméabilité des cuticules foliaires – participant ainsi également à la libération de nutriments (Schreiber *et al.*, 2005). Une stratégie additionnelle des épiphytes pour acquérir des substrats provenant de l'intérieur des tissus foliaires consisterait en la synthèse d'acide indole 3-acétique (IAA; une auxine) (Brandl *et al.*, 1996) – un des effets de cette phytohormone étant de desserrer les parois cellulaires (Fry, 1989). D'autre part, les bactéries peuvent se protéger des UVA en produisant des pigments (ex. caroténoïdes) qui absorbent leur énergie, et des enzymes antioxydantes (catalase et superoxide dismutase) qui détoxifient les espèces réactives de l'oxygène (ROS) générées par ce type de rayonnement (Jacobs *et al.*, 2005). Les dommages créés par les UVB peuvent être contrés par divers mécanismes de réparation de l'ADN (Gunasekera et Sundin, 2006).

0.1.5 Interactions biotiques

Les bactéries épiphytes de la phyllosphère sont en interaction constante avec les autres microorganismes et leur hôte. La surface foliaire est le lieu de relations aussi bien commensales, mutualistes qu'antagonistes (Faust et Raes, 2012; Vorholt, 2012). Les bactéries compétitionnent entre elles pour l'espace et les ressources. Cette compétition peut se dérouler sous forme d'exploitation (Innerebner *et al.*, 2011; Ji et Wilson, 2002; Ritpitakphong *et al.*, 2016; Wilson et Lindow, 1993) ou d'interférence, notamment par la production de molécules antimicrobiennes (Helfrich *et al.*, 2018; Stockwell *et al.*, 2002), par l'interférence des signaux de communication (*quorum quenching*; Ma *et al.*, 2013; Morohoshi *et al.*, 2009) ou par l'induction des réponses de

défense systémiques de l'hôte (Vogel *et al.*, 2016). Les interactions bactéries-bactéries au sein du microbiome de la plante sont également d'importants mécanismes de protection contre les pathogènes (Legein *et al.*, 2020).

En ce qui concerne les relations plantes-bactéries, les végétaux sécrètent par leurs feuilles une variété de métabolites secondaires (ex. molécules antimicrobiennes, composés carbonés) restreignant ou favorisant la croissance de différentes populations bactériennes (Karamanoli *et al.*, 2005; Shakir *et al.*, 2021). Parmi ces métabolites secondaires, les composés organiques volatils (COV) jouent un rôle central dans la biochimie de la phyllosphère (Farré-Armengol *et al.*, 2016). Certains COV émis par les plantes, comme les terpénoïdes, agissent comme des antimicrobiens (Gao *et al.*, 2005). D'autres COV, en particulier le méthanol, fournissent une importante source de carbone pour les bactéries méthylotrophes, comme *Methylobacterium* (Abanda-Nkpwatt *et al.*, 2006). Chez les plantes, le méthanol est formé lors de la croissance des cellules végétales ou en réponse à un dommage ou à un stress. Il résulte de la déméthylestérification des pectines (de type homogalacturonan) nouvellement acheminées à la paroi cellulaire, par les pectines méthylestérases (Dorokhov *et al.*, 2018). Le méthanol est par la suite excrété par les stomates (Nemecek-Marshall *et al.*, 1995). Par ailleurs, certaines bactéries stimulent la croissance des plantes en synthétisant puis en sécrétant des phytohormones (Sohrabi *et al.*, 2023), tandis que d'autres bactéries pathogènes réussissent à s'introduire à l'intérieur des tissus foliaires, souvent via les stomates, et à supprimer les défenses immunitaires de la plante et à s'établir dans son hôte, causant ainsi des maladies (Baker *et al.*, 2010; Roper, 2011; Xin *et al.*, 2018).

0.1.6 Une diversité de niches écologiques

La phyllosphère est un habitat hétérogène et éphémère – même les plantes sempervirentes (à feuillage persistent) renouvellent leurs feuilles à intervalle régulier. Les opportunités de colonisation de niches nouvellement disponibles, ainsi que morphologiquement et chimiquement variées, sont nombreuses et fréquentes pour les microorganismes (Vorholt, 2012). En effet, la feuille de chaque espèce de plante possède un ensemble de caractéristiques anatomiques et métaboliques qui lui est propre (Gargallo-Garriga *et al.*, 2020; Mercier et Lindow, 2000; Schweiger *et al.*, 2021) et qui reflète l'histoire évolutive de l'espèce et l'adaptation à son environnement (de la Riva *et al.*, 2017; Givnish, 1987; Nicotra *et al.*, 2011). Chaque espèce végétale hôte fournit donc à ses bactéries symbiontes un habitat caractérisé par une phytochimie distincte résultant de la sécrétion de différents métabolites (ex. sucres, huiles essentielles, COV). Ces niches

phytochimiques expliquent en partie la composition des communautés bactériennes épiphytes (Gaube *et al.*, 2023; Karamanoli *et al.*, 2005; Yadav *et al.*, 2005). Certains taxons sont même associés à des métabolites particuliers (Gaube *et al.*, 2023).

D'autre part, la phytochimie de la phyllosphère varie durant le développement de la feuille – et donc au cours des saisons dans la forêt tempérée. En effet, les feuilles matures et sénescentes exsudent plus de sucres et de nutriments inorganiques (Tukey, 1970). De plus, les émissions de COV changent avec l'âge de la feuille. Les émissions d'isoprène – utilisé comme substrat par de nombreux genres bactériens (Carrión *et al.*, 2020) – augmentent chez les feuilles matures, tandis que celles de méthanol sont plus élevées chez les jeunes feuilles (Dorokhov *et al.*, 2018; Eller *et al.*, 2012). Les variations saisonnières dans la phytochimie foliaire expliquent potentiellement certains changements temporels observés dans les communautés microbiennes au cours de la saison de croissance (Leducq *et al.*, 2022a; Redford et Fierer, 2009).

Certains traits des feuilles varient aussi en fonction de la latitude (Tian *et al.*, 2016) et de la position relative dans la canopée (Dörken et Lepetit, 2018). Toutefois, Laforest-Lapointe *et al.* (2016b) ont démontré que la structure des communautés bactériennes au sein d'un même individu hôte variait relativement peu. Enfin, chaque feuille individuelle présente elle-même aux microorganismes une multitude de microhabitats – stomates, trichomes, hydathodes, surface épidermique – physiquement et chimiquement différents (Remus-Emsermann et Schlechter, 2018; Schönherr et Baur, 1996), pour lesquels les espèces bactériennes démontrent des préférences (Esser *et al.*, 2015; Peredo et Simmons, 2018).

0.2 *Methylobacterium*

0.2.1 Phylogénie

Le genre *Methylobacterium* a été décrit pour la première fois par Patt *et al.* (1976). Sa phylogénie est sujette à débat. Selon Hördt *et al.* (2020), *Methylobacterium* fait partie de la famille *Methylobacteriaceae*, au sein de l'ordre *Hyphomicrobiales* et de la classe *Alphaproteobacteria*. Or, le *Genome Taxonomic Database* (GTDB) positionne plutôt le genre au sein de la famille *Beijerinckiaceae*, de l'ordre *Rhizobiales*. D'autre part, un genre qui avait été nouvellement proposé pour un sous-groupe d'espèce, *Methylorubrum* (Green et Ardley, 2018), s'est avéré être un clade au sein du genre *Methylobacterium* (Alessa *et al.*, 2021; Leducq *et al.*, 2022b).

Le genre *Methylobacterium* est taxonomiquement très diversifié. Plus de soixante espèces ont été décrites, mais une récente étude phylogénomique a estimé le nombre d'espèces existantes à 104, regroupées au sein de quatre clades évolutivement distincts (A, B, C, D). Le clade C serait le groupe sœur de A/B/D, et le clade B serait le groupe sœur de A/D (Leducq *et al.*, 2022b). Le transfert horizontal de gènes (HGT) et le tri de lignage incomplet ont joué un rôle important dans l'évolution du genre, la divergence des clades et la distribution des gènes parmi les espèces (Leducq *et al.*, 2022b).

Bien que des souches de tous les clades aient été isolées d'une grande diversité d'habitats, la majorité des membres des clades A et D (et une grande proportion du clade B) sont associés à la phyllosphère, tandis que ceux du clade C, très rares dans la phyllosphère, sont principalement retrouvés dans l'eau, les sols et la rhizosphère (au sein des nodules). Une hypothèse sur l'histoire évolutive de *Methylobacterium* énonce que l'ancêtre commun de toutes les espèces du genre aurait habité le sol. Les associations occasionnelles de certains taxons avec les plantes au sein de la rhizosphère auraient entraîné des adaptations à la vie symbiotique avec un hôte végétal, et éventuellement à la vie à la surface des feuilles (Leducq *et al.*, 2022b). Cette hypothèse est soutenue par de nombreuses caractéristiques génomiques, notamment la plus petite taille des génomes du groupe A/B/D (Leducq *et al.*, 2022b). L'histoire évolutive de *Methylobacterium* a une influence directe sur l'écologie de ses populations : au sein des arbres de la forêt tempérée, les communautés de *Methylobacterium* sont phylogénétiquement structurées, c'est-à-dire que les espèces qui cohabitent sont évolutivement rapprochées les unes des autres (Leducq *et al.*, 2022a).

0.2.2 Étude par culture et gène marqueur

Toutes les souches de *Methylobacterium* sont des bacilles flagellés à Gram négatif ou Gram variable, strictement aérobies, principalement chimioorganotrophes, et méthylotrophes facultatives (Green, 2006). Les souches peuvent être sélectivement isolées sur un milieu solide minimum de sels minéraux (MMS) contenant 0.1-0.2% méthanol, à pH neutre, à des températures optimales situées entre 20°C et 30°C (Green, 2006; Leducq *et al.*, 2022a). Pour étudier le genre dans son environnement sans avoir recours à la culture, Leducq *et al.* (2022a) ont conçu des amorces ciblant spécifiquement le gène *rpoB* chez la famille *Methylobacteriaceae*. Ce gène est très polymorphe et, contrairement au gène ARNr 16S largement utilisé en écologie microbienne, est présent en une seule copie dans les génomes bactériens. Le gène *rpoB* permet ainsi une meilleure résolution que le gène ARNr 16S, notamment pour l'identification taxonomique précise des souches, via le séquençage amplicon (Vos *et al.*, 2012) – une technique consistant d'abord à

amplifier (augmenter exponentiellement le nombre de copies) un gène d'intérêt via une réaction en chaîne par polymérase, puis à séquencer les copies du gène ainsi amplifié.

0.2.3 Méthylotrophie

Les bactéries méthylotrophes comme *Methylobacterium* peuvent croître en se nourrissant de composés à un atome de carbone (C1). Toutes les souches de *Methylobacterium* peuvent cataboliser le méthanol (leur principal substrat C1), le formaldéhyde et le formate, tandis que certaines souches assimilent aussi la méthylamine ou le méthane chloré (Alessa *et al.*, 2021; Green, 2006). Chez les méthylotrophes, plusieurs lanthanides (Ln; des éléments chimiques appartenant aux métaux terres rares) agissent comme cofacteurs d'enzymes impliquées dans le catabolisme du méthanol (Skovran *et al.*, 2019), c'est-à-dire qu'en culture avec méthanol, la présence de Ln stimule le taux de croissance des bactéries (Good *et al.*, 2019). Deux méthanol-déshydrogénases oxydent le méthanol en formaldéhyde, soit la XoxF Ln^{3+}-dépendante (Nakagawa *et al.*, 2012) et la MxaFI Ca^{2+}-dépendante. Leur expression différentielle est régulée par la concentration de Ln (Vu *et al.*, 2016). En plus d'oxyder quelques composés à plusieurs atomes de carbones (Good *et al.*, 2016), l'enzyme ExaF Ln^{3+}-dépendante (une éthanol déshydrogénase) oxyde le formaldéhyde en formate, mitigeant potentiellement l'effet toxique de l'accumulation de ce métabolite intermédiaire et central au métabolisme méthylotrophe (Good *et al.*, 2019).

0.2.4 La phyllosphère comme réservoir de diversité

Étant méthylotrophe, *Methylobacterium* est omniprésent dans la phyllosphère (Corpe et Rheem, 1989), où il se nourrit du méthanol libéré par les feuilles (Abanda-Nkpwatt *et al.*, 2006). La phyllosphère est en effet l'habitat contenant la plus grande diversité de *Methylobacterium* (Leducq *et al.*, 2022a); plusieurs espèces ont d'ailleurs été originalement isolées de cet habitat (voir par ex. Knief *et al.*, 2012b; Madhaiyan *et al.*, 2009, 2012; Madhaiyan et Poonguzhali, 2014; Wellner *et al.*, 2012). Une large part des protéines exprimées par *Methylobacterium* dans la phyllosphère est liée au métabolisme du méthanol et de ses dérivés; son protéome reflète aussi l'exposition aux nombreux stress abiotiques caractéristiques de la phyllosphère (Delmotte *et al.*, 2009). La protéine PhyR est centrale aux mécanismes d'adaptation à la vie à la surface des feuilles, car elle régule la réponse aux stress chez *Methylobacterium* (Gourion *et al.*, 2006). Cette protéine à deux domaines fait partie de la cascade signalétique PhyR–NepR–σEcfG qui contrôle la réponse générale aux stress chez *Alphaproteobacteria* en régulant la transcription de gènes codant pour des protéines impliquées dans la résistance contre la chaleur, la dessiccation, le stress oxydatif

et le rayonnement ultraviolet (Francez-Charlot *et al*., 2009; Gourion *et al*., 2006, 2008; Vorholt, 2012).

0.2.5 Autres habitats naturels et anthropiques

En plus de la phyllosphère, *Methylobacterium* prolifère aussi dans une multitude d'habitats naturels et anthropiques (Green, 2006). De nombreuses souches ont été isolées, souvent pour la première fois, de lacs et d'autres milieux dulcicoles (Chen *et al.*, 2019; Cho *et al.*, 2020; Patt *et al.*, 1974, 1976; Salka *et al.*, 2011; Sapp *et al.*, 2018), de systèmes d'aqueduc (Gallego *et al.*, 2005a, 2005b, 2005c), de sédiments (Tian *et al.*, 2018), de sols (Bousfield et Green, 1985; Cao *et al.*, 2011; Maeng *et al.*, 2021; Urakami et Komagata, 1984) et de la rhizosphère (Grossi *et al.*, 2020; Jourand *et al.*, 2004). D'autres souches ont été prélevées dans des milieux anthropiques tels que des équipements hospitaliers (Kovaleva *et al.*, 2014) et des climatiseurs de voitures (Park *et al.*, 2020). Enfin, des souches de *Methylobacterium* ont aussi été isolées au sein d'environnements pouvant être considérés comme étant extrêmes en raison de leurs conditions abiotiques, tels que des sols industriels (Doronina *et al.*, 1996; Ventorino *et al.*, 2014), le sol en Antarctique (Tahon et Willems, 2017), des croûtes de sol biologiques (Csotonyi *et al.*, 2010), ainsi que dans les profondeurs océaniques (Qiu *et al.*, 2019) où elles absorbent possiblement la faible lumière provenant de la radiation géothermique des cheminées hydrothermales (Beatty *et al.*, 2005; Reynolds et Lutz, 2001).

0.2.6 Photosynthèse

De nombreuses souches de *Methylobacterium* possèdent la capacité de photosynthèse anoxygénique en aérobiose (PAA) (Nissinen *et al.*, 2023; Stiefel *et al.*, 2013; Zervas *et al.*, 2019). Zervas *et al.* (2019) ont émis l'hypothèse que les gènes impliqués dans ce type de photosynthèse auraient potentiellement été incorporés par transfert horizontal de gènes dans le génome d'une souche ancestrale. Les bactéries capables de PAA sont des photohétérotrophes pouvant produire de l'ATP par absorption de la lumière par le pigment bactériochlorophylle *a* (BChl *a*) intégré dans des centres réactionnels et des antennes collectrices, sans fixer de carbone (absence de Rubisco) ni produire de dioxygène. Les pigments caroténoïdes, abondants chez les bactéries PAA, servent de pigments accessoires pour élargir le spectre d'absorption de la lumière, en plus de protéger contre le stress oxydatif (Koblížek, 2015; Yurkov et Hughes, 2017). Toutes les espèces de *Methylobacterium* sauf deux – *M. nodulans* et *M. jeotgali* (Alessa *et al.*, 2021) – synthétisent des pigments caroténoïdes, ce qui confère à leurs colonies une coloration rosée qui leur a valu l'appellation *pink-pigmented facultative methylotrophs* (Corpe et Rheem, 1989; Green, 2006).

Puisque la lumière stimule le taux de croissance des bactéries PAA, la photohétérotrophie confère aux bactéries un avantage compétitif dans les environnements lumineux, mais pauvres en carbone (Ferrera *et al.*, 2017; Hauruseu et Koblížek, 2012), comme la phyllosphère.

0.2.7 Fonctions écologiques dans la phyllosphère

Au sein de la phyllosphère, *Methylobacterium* est reconnu pour favoriser la croissance (Abanda-Nkpwatt *et al.*, 2006) et la santé de ses plantes hôtes (Oeum *et al.*, 2024), avec lesquelles il forme des associations mutualistes. Toutes les souches de *Methylobacterium* synthétisent et sécrète l'auxine acide indole 3-acétique (IAA) (Alessa *et al.*, 2021; Ivanova *et al.*, 2001; Omer *et al.*, 2004; Pattnaik *et al.*, 2017) et plusieurs, des cytokinines – zéatine, *trans*-zéatine et *trans*-zéatine riboside (Ivanova *et al.*, 2000; Koenig *et al.*, 2002; Lee *et al.*, 2006). Ces phytohormones stimulent la croissance végétale : l'IAA desserre les parois cellulaires (Fry, 1989), tandis que les cytokinines promeuvent la division et l'élongation cellulaires (Wu *et al.*, 2021). Puisque le méthanol est sécrété des feuilles lors de leur croissance (Dorokhov *et al.*, 2018), les phytohormones bactériennes augmentent par le fait même la libération de méthanol, entraînant une boucle de rétroaction positive bénéfique aux partenaires mutualistes (Sohrabi *et al.*, 2023).

D'autre part, Madhaiyan *et al.* (2004, 2006) ont démontré que certaines souches pouvaient protéger leurs hôtes contre des pathogènes fongiques en induisant chez la plante une réponse de défense systémique (Romera *et al.*, 2019), se traduisant par une augmentation de l'activité de protéines liées à la pathogenèse – telles que la phénylalanine ammonia-lyase, la *β*-1,3-glucanase, et la peroxidase – et de la concentration en phénols. Bien que le genre *Methylobacterium* contienne une grande diversité et soit reconnu pour favoriser la croissance de ses hôtes végétaux, on ne sait toujours pas si sa diversité en soi joue un rôle bénéfique pour l'hôte. Au niveau des écosystèmes, *Methylobacterium* participe au cycle biogéochimique du carbone en oxydant le méthanol sécrété par les feuilles lors de la croissance de leurs tissus, mais aussi en augmentant indirectement la quantité de CO_2 fixé par les plantes hôtes (Yurimoto et Sakai, 2023).

0.2.8 Assemblage des communautés phyllosphériques

Dans la forêt tempérée, les communautés de *Methylobacterium* fluctuent au cours de la saison de croissance. Lors du débourrement printanier, la colonisation de la surface des feuilles est stochastique et les communautés sont taxonomiquement et écologiquement très diversifiées. Au fil du temps, la composition de ces communautés converge. Celles-ci deviennent de plus en plus homogènes (Leducq *et al.*, 2022a), possiblement en réponse à la sélection imposée par l'évolution

des conditions climatiques, de la phytochimie foliaire – notamment au niveau des quantités de méthanol sécrété (Dorokhov *et al.*, 2018) – et des interactions biotiques. La saison estivale favorise l'établissement de souches à haut rendement, tandis que l'automne profite aux souches à croissance rapide (Leducq *et al.*, 2022a).

Comme pour l'ensemble des bactéries de la phyllosphère, hormis la variation due aux saisons, la composition taxonomique des communautés de *Methylobacterium* diffère principalement en fonction de l'espèce hôte et du site (Knief *et al.*, 2010; Leducq *et al.*, 2022a; Tani *et al.*, 2015). Cela est possiblement dû à la sélection des taxons les plus adaptés aux caractéristiques morphologiques et phytochimiques spécifiques à l'espèce végétale ainsi qu'aux conditions foliaires modulées par l'environnement. Certains taxons sont plus généralistes et d'autres plus spécialistes en ce qui concerne leurs espèces hôtes (Dourado *et al.*, 2012). Par ailleurs, certaines associations entre souches de *Methylobacterium* et métabolites végétaux ont été révélées. Par exemple, la présence de *Methylobacterium* est négativement corrélée à l'abondance relative de monoterpénoïdes et de sucres (certaines souches ne possèdent pas les gènes impliqués dans le catabolisme du fructose-6-phosphate en fructose-1,6-bisphospate). Or, sa présence est positivement corrélée à la concentration de certaines substances volatiles des feuilles vertes (Matsui, 2006), telles que le *cis*-3-hexèn-1-ol, l'acétate de *cis*-hex-3-ènyle, le *trans*-2-hexèn-1-ol, et le *trans*-3-hexèn-1-ol (Gaube *et al.*, 2023).

Enfin, on ne connaît toujours pas les gènes clés expliquant les associations, découlant potentiellement d'une évolution adaptative, entre les espèces de *Methylobacterium* et les différentes espèces végétales hôtes dont elles colonisent la surface foliaire.

0.3 Le pangénome bactérien

0.3.1 Structure du pangénome

La première mention du terme « pangénome » est attribuée à Tettelin *et al.* (2005), qui constatent que le nombre de gènes identifiés chez la bactérie pathogène *Streptococcus agalactiae* augmente à chaque fois que le génome d'une nouvelle souche est séquencé – similairement à ce qu'avaient remarqué Perna *et al.* (2001) et Welch *et al.* (2002) en étudiant les génomes de différentes souches de *Escherichia coli.* Le pangénome d'un taxon – souvent une espèce ou un genre – est constitué de la totalité des gènes contenus dans les génomes de ses espèces et ses souches. Il comprend des gènes *core*, présents dans toutes les souches, et des gènes accessoires, partagés

par une partie des souches ou uniques à une souche (Tettelin *et al.*, 2005; Tettelin et Medini, 2020). De manière générale, le génome *core* est principalement composé de gènes rattachés aux fonctions et aux métabolismes cellulaires de base ainsi que de gènes ribosomaux. Le génome accessoire comporte surtout des gènes impliqués dans les métabolismes secondaires et dans la composition biochimique et structurelle de la surface cellulaire qui sont associés à la communication, à la compétition, à la défense et à la pathogénicité (Bach *et al.*, 2022; Boucher *et al.*, 2011; Hyun *et al.*, 2022; Polz *et al.*, 2013). Un pangénome est considéré « ouvert » lorsque le nombre de gènes augmente significativement avec chaque nouveau génome séquencé, tandis que dans un pangénome « fermé », le nombre de gènes maximal retrouvé chez le taxon est rapidement atteint (Medini *et al.*, 2005). Comparativement à un pangénome fermé, un pangénome ouvert a une plus grande taille, contient une plus faible proportion de gènes *core* et possède un contenu génétique plus varié (Brockhurst *et al.*, 2019; Domingo-Sananes et McInerney, 2021). Des corrélations positives ont été démontrées entre la taille des populations et un mode de vie généraliste, la taille de leurs génomes, et la taille du pangénome qu'elles forment (Bobay et Ochman, 2018).

0.3.2 Transfert horizontal de gènes et composition génique

La composition du pangénome accessoire résulte à la fois du gain de gènes – via le transfert horizontal de gènes (HGT) – et de la perte de gènes dans les génomes des différentes populations. Le HGT consiste en l'insertion (principalement par conjugaison, transformation ou transduction) d'un fragment d'ADN exogène dans un génome via recombinaison homologue ou non (Ochman *et al.*, 2000; Shapiro *et al.*, 2016). Les expériences montrent que le HGT peut être délétère, neutre ou adaptatif (Gogarten et Townsend, 2005; Knöppel *et al.*, 2014; Slomka *et al.*, 2020). Son effet et ses coûts associés dépendent de l'environnement génétique et écologique du récepteur (Baltrus, 2013; Domingo-Sananes et McInerney, 2021). Certains gènes transférés dont l'effet immédiat est neutre persistent à une faible fréquence dans la population réceptrice, car la sélection n'agit pas sur eux; ces gènes fournissent toutefois un potentiel d'adaptation à de nouvelles niches (Woods *et al.*, 2020). Parfois, un gène acquis par une souche lui confère un avantage immédiat dans l'environnement où elle se trouve. Ce transfert horizontal adaptatif peut, en certains cas, mener à la diversification écologique de populations (Niehus *et al.*, 2015; Wiedenbeck et Cohan, 2011) et éventuellement, si cette diversification les isole et entrave leur recombinaison, à la spéciation (Polz *et al.*, 2013; Shapiro et Polz, 2015).

En raison de la prévalence du HGT, l'ADN chez les procaryotes peut être considéré comme un « bien public », c'est-à-dire une ressource partagée entre populations d'une même espèce ou de différentes espèces (McInerney *et al.*, 2020). Une hypothèse avancée énonce que chaque niche (ex. une espèce hôte) pourrait contenir son réservoir local de gènes au potentiel adaptatif, formant un réseau de transfert horizontal efficace et écologiquement structuré (Nowell *et al.*, 2014; Polz *et al.*, 2013). En poussant l'idée plus loin, Shapiro (2014) introduit le modèle d'*écologie des gènes*, qui suppose que ce sont les gènes qui occupent les niches écologiques et qui sont soumis à la sélection, indépendamment des espèces bactériennes hôtes au sein desquels génomes ils séjournent (une sélection génique, *sensu* Douglas et Shapiro [2021]).

0.3.3 Évolution du pangénome

La sélection adaptative (processus déterministe) ainsi que la mutation et la dérive (deux processus neutres) agissent sur les variations génomiques et font évoluer le pangénome (Andreani *et al.*, 2017; Brockhurst *et al.*, 2019; Domingo-Sananes et McInerney, 2021; McInerney *et al.*, 2017). Toutefois, il y a une incertitude quant à l'importance relative de ces deux types de processus – agissant potentiellement en parallèle – dans le maintien de la diversité génétique au sein du pangénome accessoire (Shapiro, 2017). Andreani *et al.*, (2017) soutiennent l'idée que la plus grande diversité génétique d'une grande population résulterait principalement d'une évolution neutre (Kimura, 1983). Or, la corrélation entre la taille d'une population et la taille de son pangénome pourrait être due au fait qu'une population de grande taille a la possibilité d'occuper des niches écologiques variées dont les conditions environnementales (incluant les interactions biotiques) sélectionnent des gènes diversifiés, maintenant ainsi une forte diversité génétique et une grande taille de pangénome. Selon cette logique, le pangénome serait principalement adaptatif (Bobay et Ochman, 2018; McInerney *et al.*, 2017), d'autant plus que la sélection est un processus évolutif relativement plus important que la dérive dans les populations de plus grande taille (Charlesworth, 2009). L'importance du génome accessoire dans le *fitness* – la capacité à survivre et se reproduire – des individus et dans les adaptations des populations aux différentes niches doit toutefois être clarifiée (Domingo-Sananes et McInerney, 2021). Qu'il entraîne une évolution adaptative ou neutre, le HGT demeure un processus clé dans le façonnement et l'évolution du pangénome bactérien.

0.3.4 Le pangénome de *Methylobacterium*

Le pangénome du genre *Methylobacterium* a été caractérisé dans le cadre d'une étude génomique comparative et phénotypique, afin de révéler les bases génétiques des différents

phénotypes (Alessa *et al.*, 2021). Des similarités et des différences entre les clades et les souches y ont été révélées, notamment au niveau du métabolisme méthylotrophe, de la résistance aux stress abiotiques et des traits liés à la stimulation de la croissance des plantes. De plus, le pangénome de l'espèce *M. extorquens* a été analysé par Lee *et al.* (2022) afin de comparer plus finement le métabolisme du carbone entre les souches. Dans l'étude phylogénomique de Leducq *et al.* (2022b), les auteurs ont aussi défini le pangénome du genre *Methylobacterium*, cette fois-ci pour reconstruire sa phylogénie à partir du génome *core*. L'utilisation d'approches complémentaires – basées sur la concaténation ou la coalescence des gènes *core* – a permis de révéler le rôle du HGT et du tri de lignage incomplet dans l'évolution du genre et de proposer une phylogénie plus robuste que celles des études précédentes (Alessa *et al.*, 2021; Green et Ardley, 2018). Toujours selon l'étude de Leducq *et al*. (2022b), le pangénome de *Methylobacterium* semble ouvert.

Bien que ces études pangénomiques soient exhaustives et permettent la comparaison génomique et fonctionnelle ainsi que la détermination de la relation phylogénétique des espèces ou des souches de *Methylobacterium*, elles ne mettent pas en relation les caractéristiques génomiques avec l'écologie spécifique aux populations naturelles. Or, en présentant le contenu génique de toute espèce ou toute souche d'un taxon, le pangénome fournit un canevas propice pour analyser les associations entre gènes et niches écologiques.

0.4 Lier les gènes à l'écologie

0.4.1 Gènes marqueurs

Une grande partie des études sur les communautés bactériennes de la phyllosphère fait appel au séquençage amplicon du gène ARNr 16S, un gène marqueur permettant notamment d'identifier les taxons et d'estimer leur abondance relative dans un échantillon environnemental. Or, le gène marqueur 16S a une faible résolution taxonomique – sa précision se limite au genre – et un faible pouvoir de prédiction fonctionnelle (Laforest-Lapointe et Whitaker, 2019). En effet, le gène 16S peut être fortement conservé entre plusieurs populations, alors que d'autres loci de leur génome sont variables et qu'elles occupent différentes niches écologiques (Jaspers et Overmann, 2004; Kettler *et al.*, 2007; Shapiro *et al.*, 2012). Ce gène est donc utile pour exposer des patrons globaux dans la variation de la structure des communautés, mais ne permet pas l'exploration fine des mécanismes d'adaptations des populations bactériennes à différentes niches écologiques.

Chez *Methylobacterium*, Leducq *et al.* (2022a) ont révélé la précision supérieure du gène marqueur *rpoB* pour identifier les espèces et les souches. Or, parce qu'il peut y avoir une variation intraspécifique dans le contenu génique en raison d'un transfert horizontal, d'une duplication, ou d'une perte de gènes, il est difficile d'inférer la présence ou l'absence des gènes dans une population donnée en se basant seulement sur la détection d'un gène marqueur, aussi précis soit-il pour distinguer les souches. Deux approches utilisant le pangénome permettent de faire le lien entre les gènes d'un taxon et son écologie : la construction d'un pangénome à partir de souches isolées de l'environnement d'étude, et la juxtaposition des métagénomes – les séquences d'ADN de communautés naturelles – à un pangénome de référence.

0.4.2 Pangénome de souches cultivées

En premier lieu, lorsque le pangénome est construit à partir de souches isolées de populations habitant différents environnements, ou différentes niches écologiques d'intérêt, il est possible de faire le lien entre la génétique et différentes conditions environnementales. Par exemple, pour révéler les signatures géniques des adaptations du microbiome du sol face au réchauffement climatique, Choudoir *et al.* (2023) ont collecté, isolé et séquencé de multiples souches appartenant à six genres bactériens à partir d'échantillons de sol forestier provenant de deux traitements expérimentaux de chaleur (+ 5°C et contrôle). En construisant leur pangénome de cette manière, les auteurs ont pu révéler pour chacun des six genres les voies métaboliques enrichies et celles appauvries dans les populations soumises à une plus grande chaleur. De leur côté, en analysant les gènes *core* et accessoires de souches isolées de deux régions océaniques spatialement distinctes, Coleman et Chisholm (2010) ont révélé qu'un ensemble de gènes liés à l'acquisition et au métabolisme du phosphore étaient responsables de la diversification des populations atlantique (enrichies en gènes liés au phosphore) et pacifique des genres bactériens *Prochlorococcus* et *Pelagibacter.*

0.4.3 Métapangénomique : combiner métagénomes et pangénome

La deuxième approche ne fait pas directement appel à la culture des bactéries provenant des sites ou des environnements étudiés, mais plutôt au séquençage de leurs métagénomes. La métagénomique peut révéler à la fois l'identité taxonomique et le contenu génomique des populations de microorganismes formant une communauté naturelle dans un environnement échantillonné, sans mise en culture, pour ainsi identifier les fonctions potentielles de la communauté dans cet environnement (Quince *et al.*, 2017; Ranjan *et al.*, 2016). En combinant les approches métagénomique et pangénomique, soit en juxtaposant les séquences d'ADN de

métagénomes environnementaux sur les génomes de référence formant le pangénome d'un taxon d'intérêt, il est possible de révéler précisément la structure taxonomique et fonctionnelle des communautés naturelles. Cette approche gagne en justesse lorsque les génomes de référence proviennent de souches isolées plutôt que de génomes assemblés à partir de métagénomes (*Metagenome-assembled genomes*; MAGs), et lorsqu'ils sont représentatifs de la diversité génétique de l'espèce ou de la souche.

De cette façon, les populations d'espèces bactériennes et leurs gènes peuvent être associés à différentes variables écologiques – telles que l'habitat (Utter *et al.*, 2020), la distribution géographique (Delmont et Eren, 2018) ou un gradient physico-chimique (Boeuf *et al.*, 2021) – pour ainsi exposer les bases géniques des adaptations aux niches écologiques. Par exemple, Delmont et Eren (2018) ont utilisé cette approche dite « métapangénomique » pour étudier la distribution environnementale et la variation génomique des populations du genre *Prochlorococcus* dans les océans. Dans une étude du microbiome oral, cette même approche a révélé, chez l'espèce *Haemophilus parainfluenzae* et le genre *Rothia*, la divergence génomique et écologique de sous-groupes présentant des adaptations fonctionnelles spécifiques à leur habitat préférentiel (Utter *et al.*, 2020). D'autre part, chez le clade bactérien océanique SAR324, Boeuf *et al.* (2021) ont circonscrit quatre écotypes habitant des niches spatiotemporellement distinctes définies par la profondeur et les saisons et possédant différents métabolismes adaptatifs clés liés à l'acquisition d'énergie.

L'approche métapangénomique comporte plusieurs avantages. Contrairement à l'analyse pangénomique de souches isolées des différents environnements étudiés (Choudoir *et al.*, 2023; Coleman et Chisholm, 2010; Martino *et al.*, 2016; Simon *et al.*, 2017), la métapangénomique permet la détection des gènes de chacune des populations naturelles d'un taxon d'intérêt dans un échantillon sans avoir recours directement à la culture. Cela évite des biais sélectifs (Youseif *et al.*, 2021) et permet d'obtenir une représentation plus précise de la structure des communautés, notamment par rapport à la richesse, à l'identité des espèces et à l'abondance relative de leurs populations. D'autre part, la métapangénomique permet d'étudier adéquatement la variabilité génomique au sein des populations, en évitant les biais encourus par les MAGs. Ceux-ci contiennent souvent une faible représentation des gènes accessoires (Meziti *et al.*, 2021) et parfois une forte abondance des séquences d'ADN provenant de l'hôte, surtout lorsqu'il n'est pas possible d'exclure les séquences provenant d'un hôte dont le génome n'a jamais été séquencé, comme c'est le cas pour la majorité des espèces arborescentes de la forêt tempérée. La

métapangénomique permet d'évaluer le rôle du pangénome accessoire dans l'écologie d'un taxon. Toutefois, puisque la métapangénomique est basée sur la juxtaposition des métagénomes sur un pangénome, il y a un risque que certains gènes accessoires présents dans les métagénomes ne soient pas pris en compte si le pangénome ne les contient pas (Delmont et Eren, 2018).

0.4.4 Détection du processus de sélection : ratio *pN*/*pS*

Enfin, la combinaison des approches métagénomique et pangénomique fournit l'occasion d'identifier les signaux de sélection naturelle – positive ou négative – au sein de chacun des gènes de chaque population formant les communautés bactériennes naturelles, et de lier ces signaux aux différentes niches écologiques. Le type et la force de sélection peuvent être évalués en analysant le polymorphisme au sein des gènes codants des protéines. Une des méthodes consiste à calculer le ratio *pN*/*pS* – soit le taux de polymorphisme non synonyme divisé par le taux de polymorphisme synonyme. Le ratio *pN*/*pS* est similaire au ratio *dN*/*dS*, mais évalue la variation au sein d'une population plutôt qu'entre différentes espèces (Schloissnig *et al.*, 2013). Le ratio *pN*/*pS* d'un gène (ou pour l'ensemble des gènes d'un génome) peut être calculé à partir de la variation au niveau de ses codons ou de ses acides aminés (Kiefl *et al.*, 2023). Lorsque la sélection est positive, le ratio $pN/pS > 1$, lorsque la sélection est négative (une sélection dite « purifiante »), le ratio $pN/pS < 1$, et lorsque la sélection est faible ou nulle, le ratio $pN/pS \simeq 1$. La variation du ratio *pN*/*pS* peut être associée à des variables ou des gradients environnementaux (Dong *et al.*, 2023; Shenhav et Zeevi, 2020) et être comparée entre gènes (Kiefl *et al.*, 2023), entre génomes (Schloissnig *et al.*, 2013) et entre habitats (Li *et al.*, 2024a). L'analyse du polymorphisme permet donc de mettre en lumière l'influence des niches écologiques sur l'évolution des taxons bactériens et de leur pangénome.

Finalement, au meilleur de mes connaissances, malgré son fort potentiel pour mieux comprendre le lien entre les gènes et l'écologie, l'approche métapangénomique n'a jamais été utilisée pour étudier les populations microbiennes d'un taxon associé à la phyllosphère.

0.5 La biodiversité et le fonctionnement des écosystèmes

0.5.1 Un cadre d'étude classique en écologie

En écologie, un des paradigmes les plus étudiés est la relation entre la biodiversité et le fonctionnement des écosystèmes (Cardinale *et al.*, 2012; Hooper *et al.*, 2005; Tilman *et al.*, 2014). Cette relation a d'abord été mesurée puis décortiquée chez les plantes, où il a été démontré

qu'une plus grande richesse d'espèces floristiques entraînait une plus grande stabilité de l'écosystème végétal (Tilman, 1996), une plus grande résistance aux espèces envahissantes (Naeem *et al.*, 2000) et une plus grande productivité de l'écosystème (Hector *et al.*, 1999; Tilman *et al.*, 2001, 2012). Deux principaux mécanismes ont été proposés pour expliquer la relation entre la biodiversité et le fonctionnement des écosystèmes : l'effet de la complémentarité et l'effet de la sélection, qui stipulent respectivement que l'influence de la biodiversité sur les processus écosystémiques relève d'une exploitation complémentaire des ressources ou d'une facilitation entre espèces, ou alternativement d'une sélection des espèces les plus performantes (Loreau et Hector, 2001).

0.5.2 La biodiversité bactérienne de la phyllosphère

Les effets de la diversité sur les écosystèmes ont aussi été étudiés chez les bactéries, notamment au niveau de la phyllosphère. Laforest-Lapointe *et al.* (2017a) ont révélé une corrélation positive entre la diversité bactérienne foliaire et la productivité des communautés végétales arborescentes, ainsi qu'une absence d'effet de la composition en espèces bactériennes sur la productivité végétale, suggérant selon les auteurs un effet de complémentarité des espèces plutôt qu'un effet de sélection. Il est à noter que la corrélation entre la diversité bactérienne et la croissance des arbres hôtes individuels n'a pas été testée. Bien que cette étude apporte des preuves solides du lien entre la biodiversité des bactéries et la productivité de l'écosystème, elle ne démontre pas de manière irréfutable la relation de cause à effet entre les deux variables, car aucune manipulation expérimentale du microbiome foliaire n'a été effectuée. Une question reste donc ouverte : est-ce que la plus grande diversité bactérienne pourrait en fait être causée par la plus grande productivité de l'écosystème, qui offrirait par exemple plus de ressources nutritionnelles et d'opportunités de colonisation de par sa plus grande surface foliaire?

0.5.3 Communautés synthétiques

Les bactéries s'avèrent être des organismes fort pertinents pour tester des hypothèses ou mettre à l'épreuve des théories écologiques (Jessup *et al.*, 2004), notamment par la création et l'utilisation de communautés synthétiques dans des microcosmes expérimentaux. Les communautés synthétiques sont des assemblages artificiels composés d'une diversité de souches isolées et cultivées en laboratoire, et forment un système modèle pour étudier des processus écologiques complexes dans des conditions expérimentales simplifiées et contrôlées. Elles permettent notamment de surmonter la difficulté de manipuler des communautés bactériennes en nature et d'éviter les facteurs confondants (ex. effets de sélection et de dispersion,

échelles spatio-temporelles) pouvant rendre l'interprétation des résultats difficiles (Großkopf et Soyer, 2014). Surtout, l'utilisation de communautés synthétiques dans des microcosmes simples permet d'établir des liens de causalité entre les caractéristiques des communautés et l'effet observé au niveau de l'écosystème (Vorholt *et al.*, 2017). En ce qui concerne la relation biodiversité-fonctionnement des écosystèmes, plusieurs études *in vitro* ont démontré qu'une augmentation de la richesse en espèces bactériennes dans les communautés synthétiques causait une hausse de l'activité métabolique – un proxy pour la croissance de la communauté –, et que la composition en espèces influençait aussi cette activité (Bell *et al.*, 2005; Langenheder *et al.*, 2010). Ces résultats suggèrent que les interactions interspécifiques modulent potentiellement le fonctionnement des écosystèmes bactériens. Jousset *et al.* (2014) ont démontré un effet similaire de la richesse et de la composition sur l'activité antifongique des communautés bactériennes.

Durant la dernière décennie, les communautés synthétiques ont aussi été abondamment employées dans le cadre d'études *in planta* de la phyllosphère pour étudier le lien entre les dynamiques des communautés bactériennes des feuilles et la plante hôte (Bodenhausen *et al.*, 2014; Carlström *et al.*, 2019; Innerebner *et al.*, 2011). L'intérêt pour les microcosmes foliaires s'explique notamment en raison de l'importance du microbiome phyllosphérique pour les écosystèmes terrestres – dont les agroécosystèmes –, de la facilité à manipuler les bactéries sur les feuilles et de l'opportunité d'investiguer les processus écologiques à différentes échelles spatio-temporelles (Meyer et Leveau, 2012). Dans les expériences basées sur les communautés synthétiques, la plante modèle *Arabidopsis thaliana* a joué un important rôle comme plante hôte, servant à mieux comprendre les mécanismes d'interactions entre souches bactériennes et entre bactéries et hôte, ainsi que l'impact de ces symbioses sur le fonctionnement de l'écosystème microbien. Par exemple, Innerebner *et al.* (2011) ont démontré l'effet protecteur de plusieurs souches de *Sphingomonas* isolées de la phyllosphère contre le pathogène *Pseudomonas syringae* chez *A. thaliana.* Carlström *et al.* (2019) ont démontré l'importance cruciale de l'effet de priorité – un phénomène écologique largement étudié (Stroud *et al.*, 2024) – et identifié de potentielles espèces clés de voute – un enjeu éminent en écologie microbienne (Banerjee *et al.*, 2018; Röttjers et Faust, 2018) – sur l'assemblage des communautés phyllosphériques. Bodenhausen *et al.* (2014) ont révélé que le génotype de la plante hôte influençait des changements dans la structure des communautés bactériennes. Toutefois, peu d'études – voire aucune – employant des communautés synthétiques se sont penchées directement sur l'effet de la biodiversité sur la productivité de l'hôte dans le but d'évaluer le lien de cause à effet.

En plus de contribuer à parfaire la théorie écologique (Loreau, 2010), comprendre la mécanique sous-jacente aux effets de la biodiversité végétale et phyllosphérique sur les écosystèmes présente un intérêt majeur pour les domaines des sciences appliquées telles que la foresterie (Li *et al.*, 2024b; Mori *et al.*, 2017), l'agriculture (Brooker *et al.*, 2021; de Souza *et al.*, 2020) et la conservation (Busby *et al.*, 2022; Kollmann *et al.*, 2016), particulièrement dans le contexte actuel de la crise de la biodiversité et des changements globaux (Hisano *et al.*, 2018; Oliver *et al.*, 2015; Perreault et Laforest-Lapointe, 2022; Pires *et al.*, 2018).

0.6 Chapitre 1 : Influence de l'espèce hôte sur *Methylobacterium*

En continuité avec les nombreuses études soulignant le rôle de l'espèce de plante hôte sur la structure des communautés bactériennes foliaires, ce premier chapitre interroge plus en profondeur l'influence de l'espèce hôte arborescente ou arbustive sur l'écologie et l'évolution du genre *Methylobacterium* au sein de la phyllosphère de la forêt tempérée nordique. Mon étude débute à l'échelle des communautés bactériennes et progresse en résolution jusqu'aux gènes des populations de *Methylobacterium*. Je tente ultimement de mettre en lumière les bases génétiques de l'adaptation à l'hôte.

Objectifs

Obj.1 Quantifier l'influence de l'espèce hôte sur la composition en genres des communautés bactériennes, et particulièrement sur la prévalence du genre *Methylobacterium.*

Obj.2 Quantifier l'influence de l'espèce hôte sur la richesse et la composition en espèces et en gènes des communautés de *Methylobacterium.*

Obj.3 Identifier les gènes potentiellement responsables des adaptations de *Methylobacterium* à différentes espèces hôtes, via leur abondance différentielle ou la variation dans leurs séquences génomiques.

Obj.4 Évaluer le rôle du pangénome accessoire de *Methylobacterium* dans les adaptations à ses hôtes.

Hypothèses

Considérant que la feuille de chaque espèce hôte possède des caractéristiques morphologiques et phytochimiques qui lui sont propres et que, conséquemment, elle fournit une niche écologique distincte qui exige ou entraîne des adaptations chez les bactéries de la phyllosphère, j'ai émis les hypothèses suivantes :

Hyp.1 L'espèce hôte sélectionne différents genres bactériens, menant à des communautés de composition distinctes.

Hyp.2 L'espèce hôte sélectionne différentes espèces pour former les communautés de *Methylobacterium*, influençant ainsi la structure taxonomique des communautés.

Hyp.3 L'espèce hôte exerce une pression de sélection qui lui est propre et qui détermine en partie la composition génique des communautés de *Methylobacterium.*

Hyp.4 Certains gènes conférant des adaptations sont enrichis chez des espèces hôtes particulières.

Hyp.5 Certains gènes évoluent sous l'effet d'une sélection positive chez des espèces hôtes particulières.

Considérant que les niches écologiques possèdent possiblement un réservoir local de gènes adaptatifs auquel les populations bactériennes ont accès via le transfert horizontal de gènes, j'ai émis l'hypothèse suivante :

Hyp.6 Le pangénome accessoire est partiellement adaptatif et sera donc bien représenté parmi les gènes enrichis et parmi les gènes soumis à la sélection positive chez différentes espèces hôtes.

0.7 Chapitre 2 : Diversité de *Methylobacterium* et croissance de l'hôte

Faisant suite à la démonstration de la corrélation positive entre la biodiversité des bactéries de la phyllosphère et la productivité des écosystèmes arborescents, ce chapitre tente de mettre en lumière les liens de causalité entre la richesse et la composition des communautés bactériennes foliaires et la productivité de la plante hôte.

Objectifs

Obj.1 Tester le lien de causalité et analyser le patron de corrélation entre la richesse bactérienne foliaire et la productivité de l'hôte.

Obj.2 Révéler le rôle des taxons bactériens et des interactions au sein de leurs communautés dans la relation entre diversité et productivité au sein de la phyllosphère.

Obj.3 Investiguer, sans hypothèse *a priori*, la contribution relative de deux mécanismes écologiques, la sélection et la complémentarité, au patron observé dans la relation entre diversité et productivité au sein de la phyllosphère.

Hypothèses

Hyp.1 Considérant la corrélation positive observée entre la diversité bactérienne foliaire et la productivité des communautés végétales, ainsi que l'effet bénéfique de l'inoculation foliaire de souches de *Methylobacterium* sur la croissance végétale, une plus grande richesse taxonomique de *Methylobacterium* dans la phyllosphère cause une augmentation de la croissance de la plante hôte.

Hyp.2.1 Considérant que chaque souche microbienne possède des caractéristiques génomiques, phénotypiques et écologiques distinctes qui peuvent influencer ses interactions avec l'hôte et les autres bactéries de sa communauté, les souches de *Methylobacterium* se différencient dans leur effet sur la croissance de la plante hôte.

Hyp.2.2 Considérant l'importance potentielle des taxons individuels et de leurs interactions au sein des communautés bactériennes sur le fonctionnement des

écosystèmes, le patron décrivant la relation entre la richesse bactérienne et la croissance de l'hôte varie en fonction des taxons présents dans les communautés phyllosphériques.

CHAPITRE 1

HOST TREE SPECIES DRIVE *METHYLOBACTERIUM* ECOLOGY AND EVOLUTION IN THE NORTHERN TEMPERATE FOREST

Article in preparation

Jocelyn Lauzon[1], Jean-Baptiste Leducq[2] & Steven W. Kembel[1]

[1] *Département des sciences biologiques, Université du Québec à Montréal*
[2] *Département de phytologie, Université Laval*

1.1 Résumé

Les feuilles des plantes forment un habitat microbien vaste et hétérogène, la phyllosphère. Les bactéries foliaires jouent d'importants rôles écologiques au niveau de la santé des plantes et du fonctionnement des écosystèmes. *Methylobacterium* est un genre bactérien méthylotrophe et omniprésent dans la phyllosphère, où il stimule la croissance de ses hôtes. En offrant des niches foliaires à la morphologie et à la phytochimie unique, différentes espèces de plantes hôtes abritent des communautés de *Methylobacterium* distinctes. Or, on ne connaît pas les gènes bactériens responsables des adaptations aux espèces hôtes et servant de bases aux associations symbiotiques. Nos objectifs consistaient à quantifier l'influence de l'espèce hôte sur la structure des communautés des genres bactériens, ainsi que des assemblages taxonomiques et génétiques de *Methylobacterium*; à identifier les gènes potentiellement responsables des adaptations de *Methylobacterium* à différentes espèces hôtes; et à évaluer la contribution du pangénome accessoire de *Methylobacterium* dans les adaptations aux hôtes. Nous avons collecté puis séquencé les métagénomes d'un total de 25 communautés foliaires provenant de cinq espèces hôtes représentatives d'une forêt tempérée du Québec méridional (Canada) : deux conifères (*Abies balsamea* et *Thuja occidentalis*), deux arbres feuillus (*Acer saccharum* et *Fagus grandifolia*) et un arbuste feuillu (*Corylus cornuta*). Nous avons effectué l'annotation taxonomique des séquences d'ADN obtenues afin d'analyser les communautés bactériennes au niveau du genre. Pour les analyses des communautés d'espèces et de gènes de *Methylobacterium*, nous avons juxtaposé les métagénomes obtenus sur le pangénome de *Methylobacterium* afin de mesurer la couverture et la composition de chaque nucléotide pour chacune des populations résidant sur chacun des hôtes individuels. Nous avons révélé d'importantes divergences dans la structure des communautés taxonomiques et génétiques entre les espèces d'arbres et d'arbuste, ainsi qu'entre les espèces de conifères et de feuillus. La lumière du soleil semble jouer un rôle clé dans la différenciation des communautés de la phyllosphère des arbres et des arbustes en forêt fermée, comme le démontrait l'enrichissement en gènes impliqués dans la photosynthèse, dans la phosphorylation oxydative et dans la réponse au stress chez les communautés arborescentes. D'autre part, la phénologie des feuilles semble façonner les communautés de la phyllosphère. Les communautés associées aux conifères étaient enrichies en gènes impliqués dans le métabolisme des lipides, suggérant une plus grande efficacité d'acclimatation aux fluctuations de température, tandis que les communautés sur feuillus étaient enrichies en gènes impliqués dans la défense, la membrane cellulaire et la transcription, suggérant une meilleure habileté à faire face aux variations annuelles dans les interactions biotiques résultant d'événements de colonisation stochastiques. Nos résultats démontrent que de nombreux gènes accessoires de *Methylobacterium* sont des éléments clés dans les adaptations aux espèces hôtes. En liant les gènes aux niches environnementales, notre étude a souligné le rôle crucial joué par la biodiversité floristique dans l'écologie et l'évolution microbiennes.

Mots clés : phyllosphère, *Methylobacterium*, métapangénomique, pangénome, métagénome, gènes, adaptation, rayonnement solaire, phénologie des feuilles.

1.2 Abstract

Plant leaves form a vast and heterogenous habitat for microorganisms, the phyllosphere. Foliar bacteria play important ecological roles in plant health and ecosystem functioning. *Methylobacterium* is a ubiquitous, methylotrophic, and host growth-promoting bacterial genus of the phyllosphere. Different host plant species, by offering unique morphological and phytochemical leaf niches, harbour distinct *Methylobacterium* communities. However, we do not know which bacterial genes confer adaptations to different host species and underlie specific symbiotic associations. Our objectives were to quantify the influence of host species on the community structure of bacterial genera, and *Methylobacterium* taxonomic and genetic assemblages; identify genes potentially responsible for *Methylobacterium* adaptations to different host species; and evaluate the contribution of *Methylobacterium*'s accessory pangenome in host species adaptations. We collected and sequenced the metagenomes of 25 phyllosphere communities spanning five host species in a temperate forest of southern Quebec (Canada), including conifers (*Abies balsamea* and *Thuja occidentalis*), two deciduous trees (*Acer saccharum* and *Fagus grandifolia*) and one deciduous shrub (*Corylus cornuta*). We performed a taxonomic annotation of the obtained DNA sequences to analyze bacterial communities at the genus level. For *Methylobacterium* species and gene-level analyses, we mapped the metagenomes onto *Methylobacterium*'s pangenome in order to obtain nucleotide-level coverage and composition for each population residing on each individual host. We revealed important divergences in the taxonomic and genetic community structures between trees and shrub species, as well as between conifers and broadleaf species. Sunlight seems to play a key role in differentiating phyllosphere communities of trees and shrubs in closed-canopy forests, as reflected in the tree-enriched genes involved in photosynthesis, oxidative phosphorylation, and stress response. In addition, leaf phenology seems to shape phyllosphere communities. Conifer *Methylobacterium* communities were enriched in genes involved in lipid metabolism, suggesting an enhanced ability for acclimating to fluctuating temperature, whereas broadleaf communities were enriched in genes involved in defence, cell membrane and transcription, suggesting an enhanced ability to face diverse biotic interactions resulting from yearly stochastic colonization events. We demonstrated that numerous *Methylobacterium*'s accessory genes are key elements of host species adaptation. By linking genes to environmental niches, our study highlighted the crucial role played by floristic biodiversity in shaping microbial ecology and evolution.

Keywords: phyllosphere, *Methylobacterium*, metapangenomics, pangenome, metagenome, genes, adaptation, sunlight, leaf phenology.

1.3 Introduction

The phyllosphere, the aerial parts of plants, harbours diverse microorganisms among which bacteria are the most prevalent (Vorholt, 2012). This habitat is characterized by harsh environmental conditions, such as nutrient scarcity, high UV exposition and rapid shifts in temperature and humidity, calling for particular biochemical adaptations in epiphytic bacteria (Beattie, 2011; Jacobs *et al.*, 2005; Lindow and Brandl, 2003; Schreiber *et al.*, 2005; Sohrabi *et al.*, 2023; Vorholt, 2012). The leaf bacterial microbiome supports key ecological functions both for host health and ecosystem functioning (Bashir *et al.*, 2022; Laforest-Lapointe *et al.*, 2017a). While some epiphytic bacteria are commensal and others are pathogens (Baker *et al.*, 2010; Xin *et al.*, 2018), many mutualist taxa can promote host growth and protection through secretion of phytohormones (Pattnaik *et al.*, 2017), nutrient acquisition (Batool *et al.*, 2016), resource competition (Innerebner *et al.*, 2011; Ritpitakphong *et al.*, 2016), and production of antibiotics or activation of plant defence mechanisms (De Mandal and Jeon, 2023). Phyllosphere bacteria also take part in global carbon and nitrogen cycles (Fürnkranz *et al.*, 2008; Moreira *et al.*, 2021; Yurimoto and Sakai, 2023), as well as in climate regulation (Bringel and Couée, 2015).

The temperate forest covers 6.7 million km^2 of land mainly distributed throughout America, Europe and Western Asia, as well as in some coastal regions of the southern hemisphere's continents (FAO, 2020). The vegetation is generally dominated by broadleaf deciduous trees, and the climate is characterized by strong seasonality with freezing temperatures in winter and precipitations distributed along the year (Frelich *et al.*, 2015; Kuennecke, 2008). The landscape has been extensively disturbed by human populations for residential and industrial development, forestry and agriculture (Wade *et al.*, 2003). Phyllosphere bacterial communities of the temperate forest trees have been explored in numerous studies, with a focus on exploring the effects of host species and phylogeny, spatial location and seasonality (Copeland *et al.*, 2015; Laforest-Lapointe *et al.*, 2016a; Lajoie and Kembel, 2021; Leducq *et al.*, 2022a), leaf functional traits (Yuan *et al.*, 2023), rhizosphere microbiome (Siegenthaler *et al.*, 2024), urbanization (Laforest-Lapointe *et al.*, 2017b) and latitude (De Bellis *et al.*, 2022; Wang *et al.*, 2023a) on community composition and assembly processes.

Methylobacterium (*Methylobacteriaceae*, *Hyphomicrobiales*, *Alphaproteobacteria*, *Pseudomonadota*; Hördt *et al.*, 2020) is a ubiquitous bacterial genus of the phyllosphere that is consistently detected across plant forms in all terrestrial biomes (Ishak *et al.*, 2024; Kembel *et al.*, 2014; Leducq *et al.*, 2022a; Sohrabi *et al.*, 2023; Vorholt, 2012). These bacteria are facultative methylotrophs

(Alessa *et al.*, 2021; Green, 2006); in the phyllosphere, they consume methanol secreted by leaves during their growth (Dorokhov *et al.*, 2018) as one of their main carbon and energy sources. Almost all species have the ability to perform aerobic anoxygenic photosynthesis (Alessa *et al.*, 2021; Zervas *et al.*, 2019): they can produce ATP but cannot fix carbon dioxide. Their photoheterotrophy confers them an advantage in a low chemical-energy environment such as leaf surfaces (Ferrera *et al.*, 2017; Hauruseu and Koblížek, 2012). *Methylobacterium* can stimulate plant growth (Abanda-Nkpwatt *et al.*, 2006; Pattnaik *et al.*, 2017) through the synthesis and secretion of phytohormones, such as auxins (Alessa *et al.*, 2021; Ivanova *et al.*, 2001; Omer *et al.*, 2004; Pattnaik *et al.*, 2017) and cytokinins (Ivanova *et al.*, 2000; Koenig *et al.*, 2002; Lee *et al.*, 2006). It can also protect its host against pathogens by triggering an induced systemic resistance (Madhaiyan *et al.*, 2004, 2006).

Methylobacterium is a diverse genus, estimated to include up to 104 species (of which 59 are described) distributed in four clades (A, B, C, and D) (Leducq *et al.*, 2022b). Although *Methylobacterium* strives in a wide range of natural and anthropogenic habitat types such as root nodules, soils, sediments, freshwater, and medical equipment (Green, 2006), the phyllosphere is the main reservoir of diversity, especially regarding clades A, B and D (Leducq *et al.*, 2022b). In addition to spatiotemporal variations, *Methylobacterium* community composition has been shown to be significantly impacted by host species (Knief *et al.*, 2010; Tani *et al.*, 2015), including in temperate forests (Leducq *et al.*, 2022a). This host species-driven selection for bacterial taxa and functions have also been well documented for whole bacterial communities on tree leaves, in all biomes (Kembel *et al.*, 2014; Kim *et al.*, 2012; Laforest-Lapointe *et al.*, 2016a; Lambais *et al.*, 2006; Liu *et al.*, 2023; Redford *et al.*, 2010). While it is established that distinct *Methylobacterium* assemblages, which include generalists and host specialist taxa (Dourado *et al.*, 2012), inhabit the leaf surface of different host species, no study has investigated which bacterial genes are contributing to these symbiotic associations between bacterial taxa and plant species so far.

Pangenomes offer a valuable framework to study the ecology and evolution of highly diversified taxa. A pangenome consists of all genes that are contained in the genomes of all strains or species belonging to a taxon of interest, usually a species or a genus. It is composed of core genes present in all genomes, and accessory genes shared by only a subset of genomes or found in a single genome (Tettelin *et al.*, 2005, 2008). Pangenomes can be closed or open, based on whether or not new genes are systematically detected when a new genome is added in the taxon of interest (Medini *et al.*, 2005; Tettelin *et al.*, 2008). Generalist taxa inhabiting a wide variety of habitats tend

to have open pangenomes (Bobay and Ochman, 2018), which seems to be the case for *Methylobacterium* (Leducq *et al.*, 2022b). Pangenomes are shaped by events of gene gain, through horizontal gene transfer (HGT) (Ochman *et al.*, 2000; Treangen and Rocha, 2011), and of gene loss (Domingo-Sananes and McInerney, 2021). Pangenome composition evolves under both selection and drift, but the relative importance of these processes, and thus the question of whether pangenomic diversity is adaptive, is subject to debate (Andreani *et al.*, 2017; Bobay and Ochman, 2018; Domingo-Sananes and McInerney, 2021; Douglas and Shapiro, 2021; McInerney *et al.*, 2017; Shapiro, 2017). However, the high rates of HGT in bacteria (Nowell *et al.*, 2014; Vos *et al.*, 2015), especially among populations sharing a same niche (Smillie *et al.*, 2011), generally results in genomic content that reflects a microorganism ecological niche, rather than its phylogeny (Alneberg *et al.*, 2020). Furthermore, because of the prevalence of HGT, bacterial DNA can be considered as a shared resource within the environment (McInerney *et al.*, 2020), leading to the hypothesis that each ecological niche contains a local reservoir of adaptive genes for bacteria (Nowell *et al.*, 2014; Polz *et al.*, 2013). HGT has potentially played an important role in *Methylobacterium* evolution (Leducq *et al.*, 2022b). The pangenome of the genus have been studied to link phenotypes to genes (Alessa *et al.*, 2021), to reconstruct its phylogeny and to uncover its taxonomic and genetic diversity in the phyllosphere (Leducq *et al.*, 2022b). However, *Methylobacterium* pangenome's core and accessory gene content has not yet been functionally linked to the ecological niches of its natural populations.

Metapangenomics is an approach that integrates pangenomics and metagenomics (Delmont and Eren, 2018; Scholz *et al.*, 2016), the culture-independent study of genomes in natural environments using shotgun DNA sequencing (Quince *et al.*, 2017). By mapping metagenomic sequences to a taxon's pangenome, it is possible to uncover associations between genes and functions of natural populations and biotic or abiotic variables, such as habitat types (Utter *et al.*, 2020), geography (Delmont and Eren, 2018), or physico-chemical gradients (Boeuf *et al.*, 2021). Furthermore, this approach allows to investigate the nature of evolutionary processes (positive or purifying selection, neutral evolution) acting on individual genes and across whole genomes, which is evaluated by relating variations in polymorphism within populations to environmental factors (Dong *et al.*, 2023; Kiefl *et al.*, 2023; Li *et al.*, 2024a; Shenhav and Zeevi, 2020). Metapangenomics can thus reveal the genetic basis of niche adaptation, including the role of the accessory pangenome, while avoiding biases inherent to cultivation methods or *de novo* genome assembly from metagenomes (Meziti *et al.*, 2021; Youseif *et al.*, 2021).

Understanding host-associated *Methylobacterium* diversity is of fundamental interest. In experimental conditions, different strains exhibit contrasting influence on host growth (Sanjenbam *et al.*, 2022), and these effects can be host-dependent (Tani *et al.*, 2015). Among natural phyllosphere *Methylobacterium* communities, host-driven variation in community structure can plausibly be attributed to host-specific leaf characteristics. The leaf of each plant species differs in terms of morphology and phytochemistry (i.e., metabolites produced) (Gargallo-Garriga *et al.*, 2020; Mercier and Lindow, 2000; Nicotra *et al.*, 2011; Schweiger *et al.*, 2021). Consequently, each host species provides to phyllosphere epiphytic bacteria a unique set of environmental conditions at their leaf surface, requiring an ensemble of functional adaptations arising from genetic units or variation (Sheppard *et al.*, 2018). Indeed, leaf phytochemistry has been shown to influence bacterial community structure (Gaube *et al.*, 2023; Karamanoli *et al.*, 2005; Yadav *et al.*, 2005). For example, in a field study on herbaceous hosts, Gaube *et al.* (2023) demonstrated that *Methylobacterium* abundance was negatively correlated with sugars and monoterpenoids, but positively correlated with green leaf volatiles (Matsui, 2006). Furthermore, leaf traits have been shown to impact bacterial communities' taxonomic (Kembel *et al.*, 2014) and functional (Lajoie *et al.*, 2020) structure in the tropical trees phyllosphere.

Our study mainly aims to understand how host tree species influence the ecology and evolution of *Methylobacterium* in the phyllosphere. We addressed this by investigating different taxonomic and biological units at three levels of ecological organization in the phyllosphere, zooming in from all bacterial genera, to species and genes within *Methylobacterium* communities, and finally to genes within *Methylobacterium* populations. Our first objective was to quantify the influence of host species on the composition of bacterial genera communities, with an emphasis on *Methylobacterium*. Our second objective was to quantify the influence of host species on the richness and composition of species and genes within *Methylobacterium* communities. Our third objective was to identify genes potentially responsible for *Methylobacterium* adaptations to different host species, using gene abundance and variations in codon sequences. Considering that each plant species provides distinct foliar ecological niches to bacteria, we hypothesized that different host species select for different bacterial genera, as well as for different *Methylobacterium* species and genes, thus influencing community structures in the phyllosphere. More specifically, we predicted that some genes conferring adaptation to a particular host species are either quantitatively associated with this species and/or are evolving under positive selection. Our last objective was to evaluate the contribution of the accessory pangenome in *Methylobacterium* host species adaptations. Considering that ecological niches possibly support

a local reservoir of adaptative genes that populations can acquire through HGT, we hypothesized that the accessory pangenome is in part adaptive and thus will be well represented among host-specific enriched genes and among genes subjected to positive selection.

In order to achieve our objectives, we combined the study of the *Methylobacterium* pangenome with the investigation of environmental metagenomes. More precisely, we constructed the *Methylobacterium* pangenome onto which we mapped short metagenomic reads from natural bacterial communities. Those communities were sampled on the foliar surface of five host species characterizing the understory canopy of a mesic forest site of the northern temperate deciduous forest of eastern Canada. Our approach allowed us to know exactly which *Methylobacterium* species and genes were more or less prevalent on each individual tree and to relate patterns in abundances of species and genes, as well as patterns of protein-coding gene polymorphisms, to host species or types. We revealed strong divergences in the taxonomic and genetic community structures between trees and shrubs, as well as between conifers and broadleaf species. We identified hundreds of potentially adaptive host-associated genes and functions belonging to both core and accessory pangenome, therefore linking *Methylobacterium* genes to its phyllosphere ecology.

1.4 Material and methods

1.4.1 Sampling site and host species

We collected natural phyllosphere bacterial communities at the Station de Biologie des Laurentides (SBL), the research field station of Université de Montréal located in Saint-Hyppolyte, in the province of Quebec, Canada. The SBL is located in the eastern maple-yellow birch bioclimatic domain. The landscape around the station is mainly composed of deciduous forests, small lakes and various wetlands such as marshes, peatlands and swamps (Ministère des Ressources naturelles et des Forêts, 2023). The area has an averaged annual temperature of 4.3°C (mean temperature in July of 19.0°C) and receives 1192.9 mm of precipitation annually (105.9 mm in July) (Environnement et Ressources naturelles, 2024).

The sampling site was a mesic, 70 years-old even-aged forest (with some patches of old uneven stands) on a mineral soil, with a 20–23 meters high canopy (Ministère des Ressources naturelles et des Forêts, 2023). The dominant tree species were white birch (*Betula papyrifera* Marshall), accompanied by poplars (*Populus* spp.), and maples (*Acer* spp.). For our study, we chose to

sample the leaf bacterial communities of five taxonomically diverse host plants, characteristic of the forest understory canopy : balsam fir (*Abies balsamea* (L.) Mill.), eastern white cedar (*Thuja occidentalis* L.), sugar maple (*Acer saccharum* Marshall), American beech (*Fagus grandifolia* Ehrhart), and beaked hazelnut (*Corylus cornuta* Marshall). *A. saccharum* and *F. grandifolia* are climax tree species of mesic sites in this bioclimatic domain. *A. balsamea* and *T. occidentalis* are shade tolerant conifer trees often associated with maples and birches at that latitude; balsam fir was ubiquitous, while the white cedar was mainly found closer to the forest edges near wetlands. *C. cornuta* is a 1–3 meter tall shrub species typically associated with white birch stands.

1.4.2 Sample collection and processing, DNA extraction and metagenomic sequencing

At SBL, we sampled bacterial communities from leaves of 5 individual trees of each host species (25 samples in total) on July 26, 2023 – a peak period for *Methylobacterium* diversity in this study forest (Leducq *et al.*, 2022a). For each sample, we collected approximately 10–20 grams of fresh subcanopy leaves (from a height range of 1.5–5 metres) using a pruning pole. We sealed the harvested leaves inside sterile plastic bags (Fisherbrand) and stored the samples at 4°C until processing (within 24 hours). We also collected two negative control samples by opening sterile bags on site for 10 seconds before resealing them.

The sample processing order was randomized to avoid potential batch effect. All processing steps occurred in sterile conditions under a laminar flow hood. We poured 50 mL of phosphate buffer (0.1M K_2HPO_4 mixed with 0.1M KH_2PO_4, pH 7.3) in each bag, then vigorously shook the sealed bags for 5 minutes. Bags were kept still for 10 minutes, then we transferred the phosphate buffer (~45 mL) containing the bacterial cells in sterile 50 mL falcon tubes, that we centrifuged for 30 minutes (3900 rpm, 4°C). We removed the supernatant until 1–2 mL of phosphate buffer was left in each tube. The pellet containing the microbial communities was resuspended in the remaining buffer and stored at –80°C until DNA extraction.

We performed DNA extraction in randomized sample order using the DNeasy PowerSoil Pro Kit (QIAGEN), according to the manufacturer's protocol, with the exception that ice-thawed microbial communities were centrifuged, then the supernatant was slowly removed to keep only 200 µL of phosphate buffer in which the pellet was resuspended and transferred with 800 µL of Solution CD1 in the PowerBead Pro tube. At the end of the protocol, DNA was eluted in 50 µL of Solution C6 and stored at –80°C. Three extraction kit negative controls were processed following the same protocol. The DNA concentration of each sample was quantified using a Qubit 4 Fluorometer

(ThermoFisher) with the dsDNA HS Assay Kit. No DNA was detected in any of the five negative controls (two field controls and three extraction kit controls).

DNA libraries were prepared at the Center of Excellence in Research on Orphan Diseases – Fondation Courtois (CERMO-FC) Genomics Platform at the University of Quebec in Montreal, using the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England Biolabs) following the manufacturer's protocol. 50 ng of DNA were used in combination with a 30 min fragmentation step to provide inserts of lengths 100–250bp. The clean-up step was performed with AMPure XP (Beckman Coulter) with double-size selection. NEBNext Unique Dual Index were used as primer pairs for PCR (5 cycles). Library quality control was done using a Qubit dsDNA HS Assay Kit for DNA quantification and a TapeStation with a High Sensitivity D1000 Screen Tape (Agilent Technologies) for analyzing read length profiles. Sequencing was performed at the Centre de recherche CHU de Québec (Laval University) on an Illumina NovaSeq 6000 S4 flow cell using Reagent Kit v1.5, to obtain approximately 85 million paired-end reads of 2x150 bp. This large sequencing depth allowed to adequately study low abundance populations and genes in the samples. Demultiplexed paired-end metagenomic reads were quality filtered using 'iu-filter-quality-minoche' function (with default parameters) implemented in anvi'o (v8; Eren *et al.*, 2015), which is based on an approach by Minoche *et al.* (2011).

1.4.3 Assembling and characterizing the *Methylobacterium* pangenome, and competitive mapping of metagenomic reads

In order to build the *Methylobacterium* pangenome, we started by selecting one reference genome for each of the 104 species identified in an extensive phylogenomic study by Leducq *et al.* (2022b). Of those 104 species, 59 are named and described species (out of the previously described 63 species, some were assigned as subspecies; for details see Alessa *et al.* [2021]), and 45 are new candidate species (*M. sp. 001–045*) with no type strain (Leducq *et al.*, 2022b). *Methylorubrum* was considered part of *Methylobacterium* (Alessa *et al.*, 2021; Leducq *et al.*, 2022b). For species with more than one genome, we selected the genome with the largest N50. A table containing all selected genomes' strains, assemblage statistics and download sources can be found online at https://github.com/lauzonj/methylobacterium_tree_hosts/blob/main/genomes_methylobacterium.csv. Some selected species' genome belonged to strains that were previously isolated from this study sampling forest and from the Gault Nature Reserve at Mont-Saint-Hilaire (MSH), another deciduous forest in Quebec during the summer of 2018 (Table 1.1) (Leducq *et al.*, 2022a).

We performed all steps of the assembly and characterization of the *Methylobacterium* pangenome with anvi'o, using native functions and integrated softwares. We used the function 'anvi-gen-contigs-db' to generate an anvi'o database containing the 104 genomes, to be used for subsequent steps. Within each genome, Prodigal (v2.6.3; Hyatt *et al.*, 2010) was used to predict genes (i.e. detect open reading frames). Genes were functionally annotated using four databases – Clusters of Orthologous Groups (COG; National Center for Biotechnology Information; v.20; Galperin *et al.*, 2021) (function 'anvi-run-ncbi-cogs'), KEGG Kofam (Aramaki *et al.*, 2020) (function 'anvi-run-kegg-kofams'), and Pfam (Mistry *et al.*, 2021) (function 'anvi-run-pfams').

We used bowtie2 (Langmead and Salzberg, 2012) to competitively map metagenomic short reads to the 104 *Methylobacterium* genomes with the default parameters (--sensitive), then we used *samtools* (Danecek *et al.*, 2021) to sort and index alignments. Using the functions 'anvi-profile', 'anvi-merge' and 'anvi-summarize', we uncovered from the BAM files (storing the read recruitment results) statistics on the horizontal coverage (i.e., coverage breath) and the coverage depth for each gene of every *Methylobacterium* species in each sample, as well as the horizontal coverage and mean coverage depth for each *Methylobacterium* genome in each sample. Horizontal coverage of a genome was measured as the proportion of nucleotides having at least 1X coverage (note from the authors : in anvi'o language, 'horizontal coverage' is called 'detection', but we prefer the use of 'horizontal coverage' for simplicity and to avoid confusion). At the profiling step, we also excluded from analyses contigs < 1000 bp from further analyses, and we calculated single codon variants (SCVs).

We used the function 'anvi-pan-genome' (parameters --minbit 0.5, --mcl-inflation 10), which integrates DIAMOND (Buchfink *et al.*, 2015), MCL (Van Dongen, 2008; van Dongen and Abreu-Goodger, 2012) and MUSCLE (Edgar, 2004), to identify and align gene clusters and compute the pangenome (for detailed method, see Utter *et al.* [2020] and Delmont and Eren [2018]). Genes (i.e., open reading frames) were grouped in gene clusters based on the similarity of their amino acid sequences; these homologous genes can be orthologs or paralogs. We used the anvi'o function 'anvi-script-compute-bayesian-pan-core' to predict whether gene clusters were considered as core genes or accessory genes. This function integrates the 'mOTUpan' method (Buck *et al.*, 2022), a Bayesian approach to estimate the likelihood that genes belong to the core genome. This method was initially implemented in the mOTUlizer software.

Table 1.1 *Methylobacterium* strains (*n* = 18) previously isolated from the sampling site (Station de Biologie des Laurentides; SBL) or from another deciduous forest (Gault Nature Reserve at Mont-Saint-Hilaire, QC, Canada; MSH), and whose genomes were used to construct the pangenome.

***Methylobacterium* species**	**Strain**	**Host species**	**Forest of origin**
M. sp. 002	J-068	*F. grandifolia*	MSH
M. sp. 003	J-078	*F. grandifolia*	MSH
M. sp. 004	J-090	*F. grandifolia*	SBL
M. sp. 014	E-045	*A. saccharum*	MSH
M. sp. 018	E-025	*A. saccharum*	MSH
M. sp. 020	E-005	*F. grandifolia*	MSH
M. sp. 021	E-065	*F. grandifolia*	SBL
M. sp. 022	J-026	*F. grandifolia*	SBL
M. sp. 023	J-030	*F. grandifolia*	SBL
M. sp. 024	J-048	*F. grandifolia*	SBL
M. sp. 025	J-070	*F. grandifolia*	SBL
M. sp. 026	J-077	*F. grandifolia*	MSH
M. sp. 027	J-088	*A. saccharum*	SBL
M. sp. 029	E-066	*F. grandifolia*	SBL
M. sp. 030	J-001	*A. saccharum*	MSH
M. sp. 033	J-067	*F. grandifolia*	MSH
M. sp. 034	J-076	*F. grandifolia*	MSH
M. sp. 036	J-043	*F. grandifolia*	MSH

1.4.4 Taxonomic annotation of metagenomic reads of complete bacterial communities

In order to study how the host plant species influenced whole bacterial communities beyond *Methylobacterium*, we needed to identify and quantify all bacterial taxa found in the phyllosphere metagenomes. We used the software kraken2 (Wood *et al.*, 2019) with the standard database to perform a k-mer based taxonomic annotation of all metagenomic reads in each of the 25 samples. We classified reads at the genus level for bacteria. The abundances (counts) of paired-end reads for every genus in every sample were estimated using bracken (Lu *et al.*, 2017), and gathered in a bacterial genera community table. We incorporated the reads assigned to *Methylorubrum* into *Methylobacterium* read counts (Alessa *et al.*, 2021; Leducq *et al.*, 2022b).

1.4.5 Statistical analyses

All statistical analyses were performed on R (v.4.4.0; R Core Team, 2022). Data was manipulated with the help of the R packages tidyr (Wickham *et al.*, 2023a), dplyr (Wickham *et al.*, 2023b), and

picante (Kembel *et al.*, 2010). Figures were made with ggplot2 (Wickham, 2016). All linear model's assumptions were verified. When these assumptions (i.e., normality of residuals) could not be met for analyses of variance (ANOVAs) using raw or transformed data (i.e., logarithmic transformation for log-normal distributed variables), we used the nonparametric Kruskal-Wallis (KW) test as an alternative. η^2 was used as an effect size measure for Kruskal-Wallis tests. In the context of ANOVAs and Permutational multivariate analyses of variance (PERMANOVAs; Anderson, 2001), when post-hoc tests were needed to clarify which host species were driving the overall host effect we 1) compared samples from *C. cornuta* to all other samples (the shrub vs. trees comparison) and 2) compared samples from *A. balsamea* and *T. occidentalis* to samples from *A. saccharum* and *F. grandifolia* (the conifer vs. broadleaf trees comparison). We did not include *C. cornuta* samples in the broadleaf group in the second type of comparison since we wanted to separate the effect of *C. cornuta* from the effect of *A. saccharum* and *F. grandifolia*. These decisions of post-hoc comparison types were made after observing recurrent patterns in the data, and because the low number of samples from each host species (n = 5) did not allow enough statistical power for multiple host species pairwise comparisons. Post-hoc tests of ANOVAs were made with Student t-tests or Welch t-tests in the case of unequal variances. Post-hoc tests of PERMANOVAs were performed by subsequent PERMANOVAs comparing the composition between the groups of interest. A Bonferonni correction was applied to p-values of post-hoc tests. Statistical significance threshold for all statistical analyses was set at α = 0.05.

1.4.5.1 Metagenomic sequencing yield and Methylobacterium *read mapping results*

We tested if the number of quality-filtered read pairs obtained from shotgun sequencing varied among host species using an ANOVA. We employed Kruskal-Wallis tests to verify if the number of single reads that mapped to the *Methylobacterium* pangenome varied among host species, as well as to verify if the percentage of single mapped reads varied among host species. We tested if there was a correlation between the number of mapped reads (log-transformed data) and the amount of filtered read pairs with a Pearson r coefficient correlation test.

1.4.5.2 Phyllosphere complete bacterial community composition

Our first objective consisted in quantifying the host species effect on the whole bacterial community taxonomic composition at the level of genera, including the relative abundance of *Methylobacterium* in those communities. As a preliminary analysis, we used a linear regression model to investigate the correlation between the raw number of single reads assigned to *Methylobacterium* (including *Methylorubrum*) by bracken and the number of single reads mapped

to the *Methylobacterium* pangenome using bowtie2, to assess the precision of both methods in detecting metagenomic reads from a bacterial genus.

In order to calculate the relative abundance of *Methylobacterium* in the bacterial communities, we started by performing one random subsampling ("rarefying", *sensu* Schloss [2024]) set at the lowest number of total paired-end reads of a sample (n = 7,597,648) with the function 'rrarefy' in vegan R package (Oksanen *et al.*, 2022). Based on this subsampled data, we first compared *Methylobacterium* relative abundance among hosts using an ANOVA, followed by post-hoc Welch and Student t-tests. We then calculated the relative abundances of all genera composing bacterial communities, and illustrated these differences using a barplot. We computed a Spearman correlation matrix (and calculated Benjamini-Hochberg adjusted p values) based on the rarefied read counts, excluding genera that did not make up a total of 0.5% of bacterial reads across samples.

Taxonomic assignment by kraken2 can result in high numbers of very low abundance taxa being detected. For subsequent beta diversity analyses, we decided to suppress from the community table all genera detection values that did not account for at least 0.1% relative abundance (less than 1 read per 1000 reads) of the total community for a given sample. We then performed a rarefaction (*sensu* Schloss [2024]) set at the lowest number of total paired-end reads of a sample (n = 6,048,614) with the function 'avgdist' in vegan (100 iterations) to obtain a rarefied Bray-Curtis distance matrix. To investigate if bacterial community composition at the genus level varied among host species, we performed a PERMANOVA with the function 'adonis2' in vegan, followed by subsequent post-hoc PERMANOVAs and analyses of multivariate homogeneity of variances with function 'betadisper' in vegan (based on Anderson [2006] and Anderson *et al.* [2006]), to verify if PERMANOVAs results could be due to differences in compositional variance among host species or groups of species. We visualized the compositional differences among samples with a Principal Coordinates Analysis (PCoA) based on the Bray-Curtis distance between communities.

1.4.5.3 Assessing the presence of a Methylobacterium *species in a sample*

To perform *Methylobacterium* species-level analyses, we first needed to assess the presence of these species in each phyllosphere sample. Many studies used a threshold of 50% of horizontal coverage (i.e., half of a genome's nucleotides have a coverage depth ≥ 1X) to assess the detection of a genome (i.e., the presence of a population from the related species) in a given sample (Delmont and Eren, 2018; Giacomini *et al.*, 2023; Utter *et al.*, 2020). However, because horizontal

coverage data of our study had a log-normal distribution (Appendix A, Figure A.1a) there was no distinct value over which genomes appeared to be clearly detected. Instead of choosing an arbitrary threshold, we decided to use a threshold based on outlier values of horizontal coverage. We log-transformed the variable's data for all genomes in all samples to a normal distribution, and we calculated *z*-scores on log values. We used a threshold of $z \geq 1.96$ to assess outliers data points which we each considered as a genome detected in a sample (Appendix A, Figure A.1b). The equivalent threshold in horizontal coverage value was 34%. In other words, a genome which had 34% of its nucleotides covered at least once by metagenomic reads in a given sample was considered detected in that sample.

1.4.5.4 Methylobacterium *species communities' diversity, composition and phylogenetic relations*

Part of our second objective was to quantify the host species effect on *Methylobacterium* species community structure. We first built a *Methylobacterium* species community table keeping for every individual sample only species which had their genome detected in that sample using the method described above. For *Methylobacterium* species considered as present in a sample, we used the mean coverage depth of their genomes to assess their abundance. Before performing alpha and beta diversity analyses on *Methylobacterium* species communities among host species, we tested with a linear model if there was a correlation between species richness and the number of filtered read pairs in a sample, to verify if differences in sequencing depth could introduce a bias in our interpretation of the results. Since we found no correlation, we did not normalize the data (i.e., no rarefaction) prior to subsequent *Methylobacterium* species analyses.

As measures of alpha diversity, we calculated species richness, Shannon index and Pielou evenness index for every sample. We used ANOVAs followed by post-hoc Welch t-tests to compare the species richness and the Pielou index among communities of different hosts, and a Kruskal-Wallis test to compare the Shannon index. We calculated a linear regression to test the relationship between species richness and the Pielou index to investigate how species diversity was distributed within samples. We illustrated the abundance (based on coverage depth) of *Methylobacterium* species in communities of every sample using a barplot.

We used the function 'beta.div.comp' in R package adespatial (Dray *et al.*, 2024) to 1) compute a distance matrix based on the percentage difference dissimilarity coefficient (i.e., Bray-Curtis) and 2) partition the overall beta diversity of our samples in two components : balanced variation of

abundances (i.e., turnover of species and/or reshuffled abundances among similar species) and abundance gradients (i.e., nestedness and/or monotonic differences in abundances of shared species) (*sensu* Baselga [2013] and Legendre [2014]). We similarly partitioned the within host species beta diversity. Furthermore, we used a PERMANOVA on the Bray-Curtis distance matrix to test the difference in composition of *Methylobacterium* species communities among hosts, followed by post-hoc PERMANOVAs and multivariate homogeneity of variances tests. We computed a Principal Coordinates Analysis (PCoA) to visualize the compositional differences between samples.

In addition, we investigated how *Methylobacterium* taxonomic community structure related to phylogeny. To perform phylogenetic analyses on *Methylobacterium* species communities, we used a phylogenetic tree previously constructed with RAxML (Stamatakis, 2014) from the concatenated alignments of 384 core gene nucleotide sequences, from 213 *Methylobacteriaceae* genomes (Leducq *et al*., 2022b). This tree included all the strains whose genomes we used to construct the *Methylobacterium* pangenome. Firstly, we tested if *Methylobacterium* communities were composed of phylogenetically related species. We computed an interspecific phylogenetic distance matrix of *Methylobacterium* species with the function 'cophenetic' in the stats package. We then calculated the standardized effect size of mean pairwise distance (MPD) with the function 'ses.mpd' in the picante package, without taking into account species abundance and using shuffled tip labels of a pruned tree as a null model (1000 iterations on 999 runs). The obtained *z*-scores were used to compare the MPD of observed communities to null communities. Negative *z*-scores indicated phylogenetic clustering within communities, while positive *z*-scores indicated larger phylogenetic distance than expected by chance. We also compared the *z*-scores among host species with an ANOVA to test if host species was having an effect on the degree of phylogenetic relatedness of communities. Secondly, we tested if different host species harboured distinct phylogenetic groups of *Methylobacterium* species. We measured the intercommunity MPD (i.e., the mean phylogenetic distance between every pair of species from two different communities) with the function 'comdist' in picante. We then used this distance matrix to perform a PERMANOVA. Finally, we plotted a heatmap to visualize the relation between phylogeny and community composition.

1.4.5.5 Assessing gene clusters presence and abundance in Methylobacterium *communities*

In order to perform *Methylobacterium* gene-level analyses, we first needed to measure the abundance of each gene cluster – the basic units of analyses – in each sample. First, to assess

the presence of a given unique gene (i.e., ORF predicted by Prodigal) of a *Methylobacterium* species in a given sample, we used a horizontal coverage threshold of ≥ 50% and a coverage depth threshold of ≥ 0.5X. We thus kept any unique genes that had at least half of their nucleotides covered a minimum of 1X in a given sample, regardless of whether the gene belonged to a detected *Methylobacterium* species in the sample. We included genes from all *Methylobacterium* reference genomes to avoid missing genes contained in genomes of non-detected species that could potentially have been horizontally transferred to detected species. Afterward, we rounded the coverage depth value of unique genes to integers to get count values, and we summed up the count values of every unique gene belonging to the same gene cluster. These steps resulted in a gene cluster "community" table containing gene clusters count values (i.e., abundance) for every sample. These abundance values were used in all subsequent gene-level analyses. Of the 70,695 gene clusters identified in the *Methylobacterium* pangenome, 22,190 clusters were present across all samples, with a total abundance of 7,476,568 counts. *Methylobacteirum* gene cluster abundances varied greatly among samples (from 12,004 to 1,081,178).

1.4.5.6 Methylobacterium *gene communities' diversity and composition*

Another part of our second objective was to investigate the host-driven differences in *Methylobacterium* genetic community structure. To conduct beta diversity analyses on gene cluster composition of *Methylobacterium* communities, we performed a rarefaction set at 44,221 counts on the gene cluster community table using the function 'avgdist' (100 iterations) to obtain a rarefied Bray-Curtis distance matrix. This rarefaction resulted in the loss of one *A. saccharum* and one *F. grandifolia* sample (23 samples remained). We tested if gene clusters composition of *Methylobacterium* communities varied among host species by performing a PERMANOVA, followed by the usual post-hoc PERMANOVAs and multivariate homogeneity of variances tests. We computed a PCoA to visualize gene cluster compositional differences among host species and host types.

Before performing alpha diversity and differential abundance analyses on gene clusters of *Methylobacterium* communities, we generated one random subsampling of the original gene cluster community table, then filtered remaining low abundant gene clusters. We rarefied again at 44,221 counts, but this time with the function 'rrarefy'. After rarefying, there were 17,488 gene clusters left (representing 78.8% of all detected clusters), distributed in 23 samples (the same samples that were lost in the previous 'avgdist' rarefaction). We filtered remaining low abundant gene clusters by removing clusters whose counts were < 0.001% of total gene cluster counts

across samples (i.e., clusters that had less than 1 count for every 100,000 counts). After filtering, there were 10,091 gene clusters left (representing 57.7% of rarefied gene clusters, and 45.5% of all detected clusters), while 97.4% of rarefied gene clusters counts remained. Most removed gene clusters had an initial (i.e., pre-rarefying and filtering) very low total abundance (mean count = 18, median = 9) compared to kept gene clusters (pre-rarefying and filtering : mean count = 719.3; median = 463; post-rarefying and filtering; mean count = 98.2; median = 61). These steps resulted in a rarefied and filtered gene cluster community table. We used this table to compare gene cluster richness and Shannon index among host species using an ANOVA and a Kruskal-Wallis test respectively.

In order to address our third objective, to identify potentially host-adaptive *Methylobacterium* genes through variation in abundance, we performed differential abundance analyses on gene clusters using DESeq2 (Love *et al.*, 2014). Although their study was focusing on operational taxonomic units (OTUs), Weiss *et al.* (2017) demonstrated that DESeq2 was sensitive enough to work with small sample size (i.e., 5 samples per group), and that rarefying decreased the false discovery rate for datasets with large variation in the average library size among groups, which is the case in our study (Appendix A, Figure A.7). We thus used the rarefied and filtered gene cluster data with DESeq2 to identify and quantify the differentially abundant gene clusters (DAGCs) among species and host types, more precisely between the four tree species (*A. balsamea*, *T. occidentalis*, *A. saccharum*, *F. grandifolia*) and the shrub species (*C. cornuta*); between conifers (*A. balsamea*, *T. occidentalis*) and all broadleaf species (*A. saccharum*, *F. grandifolia*, *C. cornuta*); between conifers (*A. balsamea*, *T. occidentalis*) and broadleaf trees (*A. saccharum*, *F. grandifolia*); between conifers (*A. balsamea*, *T. occidentalis*) and the shrub species (*C. cornuta*); between broadleaf trees (*A. saccharum*, *F. grandifolia*) and shrub (*C. cornuta*); between *A. balsamea* and *T. occidentalis*; and between *A. saccharum* and *F. grandifolia*. When a gene cluster was identified as differentially abundant in more than one comparison, we considered it only in the broadest comparison. For example, a gene cluster that was more prevalent in all broadleaf species compared to conifers, as well as more prevalent in broadleaf trees compared to conifers, was only considered as a DAGC in the first type of comparison. Furthermore, we calculated the abundance of DAGCs in each COG categories in each type of comparison, excluding the category “Function unknown” as well as unassigned functions (i.e., no functional annotation). Gene clusters assigned to more than one COG category were counted once in each assigned category.

1.4.5.7 pN/pS ratio based on single codon variants (SCVs) to quantify the type of selection acting on genes in a host-specific manner

To identify host-adaptive genes, we also analyzed variations in codon sequences to detect signals of selection. We chose the three *Methylobacterium* species – *M. sp. 018*, *M. sp. 021*, *M. sp. 022* – that occurred in the highest number of samples (with occurrences in at least three samples per host species) to conduct our selection analyses. As previously mentioned, single codon variability was calculated at the metagenomic read mapping profiling step. We used the anvi'o function 'anvi-gen-variability-profile' to gather SCV data for each of the three *Methylobacterium* species across all samples. The output tables contained statistics on competing codons (i.e., coverage depth, reference, consensus) on a per position-per gene basis, in each sample in which a given *Methylobacterium* species was detected. The function 'anvi-get-pn-ps-ratio' (--min-coverage 10) was then used to calculate the *pN*/*pS* ratio – which is the rate of non-synonymous polymorphism (*pN*) divided by the rate of synonymous polymorphism (*pS*) (Schloissnig *et al.*, 2013) – for each gene of each species in each sample. The *pN*/*pS* ratio was calculated relatively to the reference codons, with an added pseudo-count of 1.0^{-6} to the *pS* rate. A *pN*/*pS* ratio > 1 indicates positive selection, a ratio < 1 indicates purifying selection, while a ratio = 1 indicates neutral evolution. We used the thresholds of $pN/pS > 1.1$ for positive selection, $1.1 \geq pN/pS \geq 0.9$ for near-neutral evolution (or no selection), and $pN/pS < 0.9$ for purifying selection. For each *Methylobacterium* species, only polymorphism data from samples in which the species was present were considered.

We first calculated the contribution of the different types of selection (positive, near-neutral, and purifying) for each of the three *Methylobacterium* species across all genes and samples. We then compared these frequencies among *Methylobacterium* species with a Pearson Chi-squared test, in order to verify if the nature of selection differed among *Methylobacterium* species. We performed post-hoc Chi-squared tests for species pairwise comparison with Bonferroni adjusted *p*-values. We also performed a Chi-squared test to compare selection type frequencies between core and accessory genes across all *Methylobacterium* species and all samples.

All subsequent analyses were performed separately for each of the three *Methylobacterium* species. We calculated the frequencies of each type of selection across all genes of a *Methylobacterium* species on a per host species basis. We then performed a Pearson Chi-squared test to analyze if the relative contribution of different selection types to single codon variation differed among host species. Based on these results, we performed a second Chi-squared test to analyze the difference between conifers and broadleaf species for *M. sp. 018*.

Furthermore, in order to reveal genes under positive selection by one or more host species, we first identified all genes that had at least one *pN*/*pS* ratio > 1.1 in any sample, then we selected the data related to those genes only for host species in which a given gene showed codon-level variability in at least three samples. Subsequently, to limit the number of statistical tests to relevant comparisons, we identified gene-host species combinations with a mean *pN*/*pS* > 1.1 and a lower-bound standard deviation around the mean > 1. We then performed Student t-tests to test if the mean *pN*/*pS* ratio of the gene in the host's samples was greater than 1. We applied the Benjamini-Hochberg procedure to adjust *p*-values. We finally assessed if positively selected genes were also DAGCs.

1.5 Results

1.5.1 *Methylobacterium* sequences were more prevalent in *Corylus cornuta*'s metagenomes

We started by analyzing the variation in metagenomic sequencing yield and in pangenome mapping statistics. Among all 25 samples, between 54,301,257 and 123,574,867 raw paired-end reads per sample (mean of 85,081,000) were obtained from metagenomic shotgun sequencing, and a mean of 98% of pairs passed the quality filtering step (Appendix A, Table A.1). The amount of filtered read pairs obtained showed a tendency to vary with host species (ANOVA, $p = 0.063$, adj. $R^2 = 0.217$), with *A. saccharum* samples generally yielding fewer pairs (Appendix A, Figure A.2). The amount of filtered single reads that were mapped to the *Methylobacterium* pangenome varied with host species (KW, $p = 0.015$, $\eta^2 = 0.420$), with *C. cornuta* samples mapping more *Methylobacterium* reads than the other four host species (Figure A.3). *Methylobacterium* single reads accounted for a different percentage of the total metagenomic reads (which include unknown amounts of other bacterial, archaeal and eukaryotes DNA, including host DNA) in samples among host species (KW, $p = 0.021$, $\eta^2 = 0.377$). *C. cornuta* samples contained a larger percentage of *Methylobacterium* reads (mean of 3.46%) than the other four host species (Figure 1.1). Overall, there was no clear correlation between the number of mapped reads to the *Methylobacterium pangenome* and the amount of filtered read pairs ($p = 0.112$, Peason $r = 0.326$), which indicated that our sequencing depth was generally large enough to capture most of *Methylobacterium* genomic DNA in our samples. Nevertheless, we detected a tendency within samples of *A. balsamea* and *T. occidentalis* to show an increase in the number of mapped reads with increasing filtered pairs (Appendix A, Table A.2).

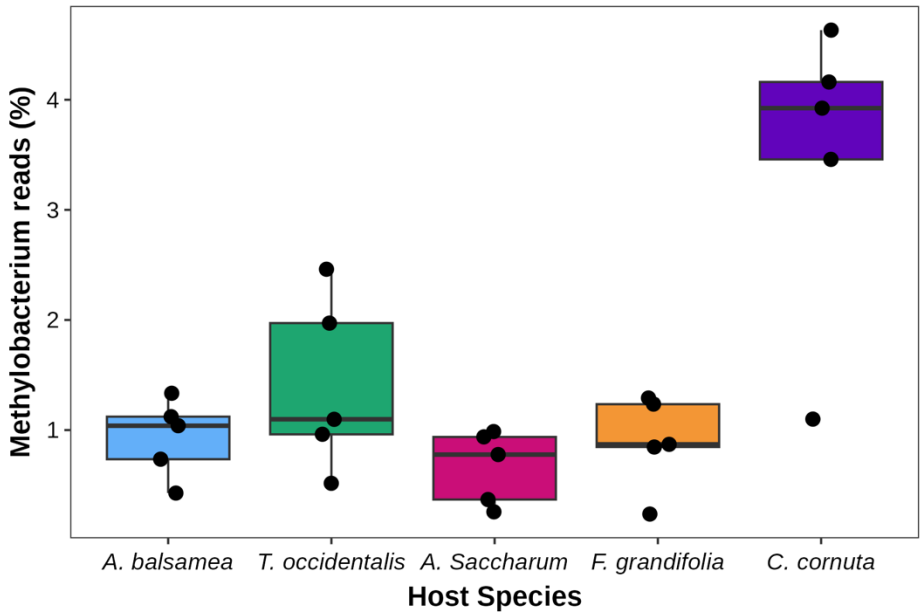


Figure 1.1 The metagenomes of *C. cornuta* were generally composed of a larger proportion of *Methylobacterium* sequences than the metagenomes of other host species (Kruskal-Wallis, $p = 0.021$, $\eta^2 = 0.377$). The boxplot illustrates the percentage of metagenomic single reads that mapped to *Methylobacterium*'s pangenome in samples of the five host species.

1.5.2 Total bacterial community composition varied among host species

We analyzed how host species influenced the genus composition of the entire leaf bacterial communities, with a particular emphasis on *Methylobacterium*'s relative abundance. The number of single reads that were assigned to *Methylobacterium* (including *Methylorubrum*) by bracken (i.e., k-mer based taxonomic assignment) was strongly correlated with the number of single reads that were mapped to *the Methylobacterium* pangenome (regression, $p < 0.001$, adj. $R^2 = 0.970$). The relative abundance of *Methylobacterium* in bacterial communities varied from 2.7 to 16.7% (mean of 6.77 ± 3.98% standard deviation) across all samples, and varied among hosts (ANOVA, $p < 0.001$, adj. $R^2 = 0.692$). *Methylobacterium* was 2.6 times more prevalent on *C. cornuta* (mean of 13.42 ± 3.97%) than on tree species (mean of 5.11 ± 1.47%) but did not vary significantly between conifers and broadleaf tree species (Figure 1.2; Appendix A, Table A.3). The most abundant bacterial genera across all 25 phyllosphere samples were *Mycobacterium* (8.42 ± 3.09%), *Xanthomonas* (8.41 ± 2.37%), *Streptomyces* (7.49 ± 1.48%), *Methylobacterium* (6.77 ± 3.98%), *Sphingomonas* (5.60 ± 2.33%) and *Pseudomonas* (3.05 ± 1.79%; Figure 1.3). The read counts of three genera covaried positively with *Methylobacterium* (*Spirosoma*, *Sphingomonas*, *Hymenobacter*), while four genera covaried overall negatively with *Methylobacterium* (*Bacillus*, *Mycobacterium*, *Granulicella*, *Xanthomonas*) (Figure 1.4).

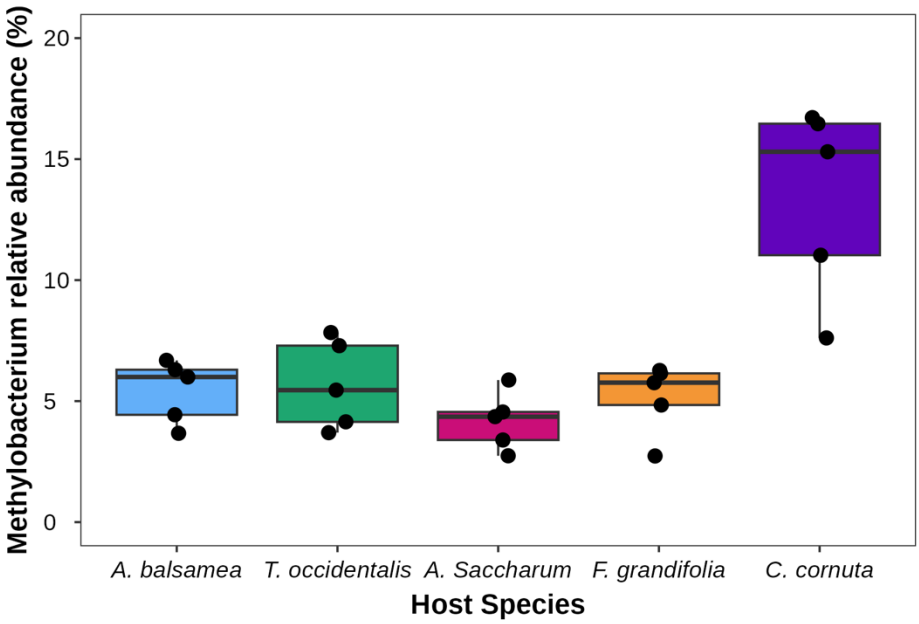


Figure 1.2 The relative abundance of *Methylobacterium* was higher in *C. cornuta* phyllosphere bacterial communities (Welch t-test, $p = 0.017$). The boxplot illustrates *Methylobacterium*'s relative abundance in samples of the five host species, measured as the percentage of bacterial reads assigned to *Methylobacterium* (or *Methylorubrum*) based on a k-mer classification.

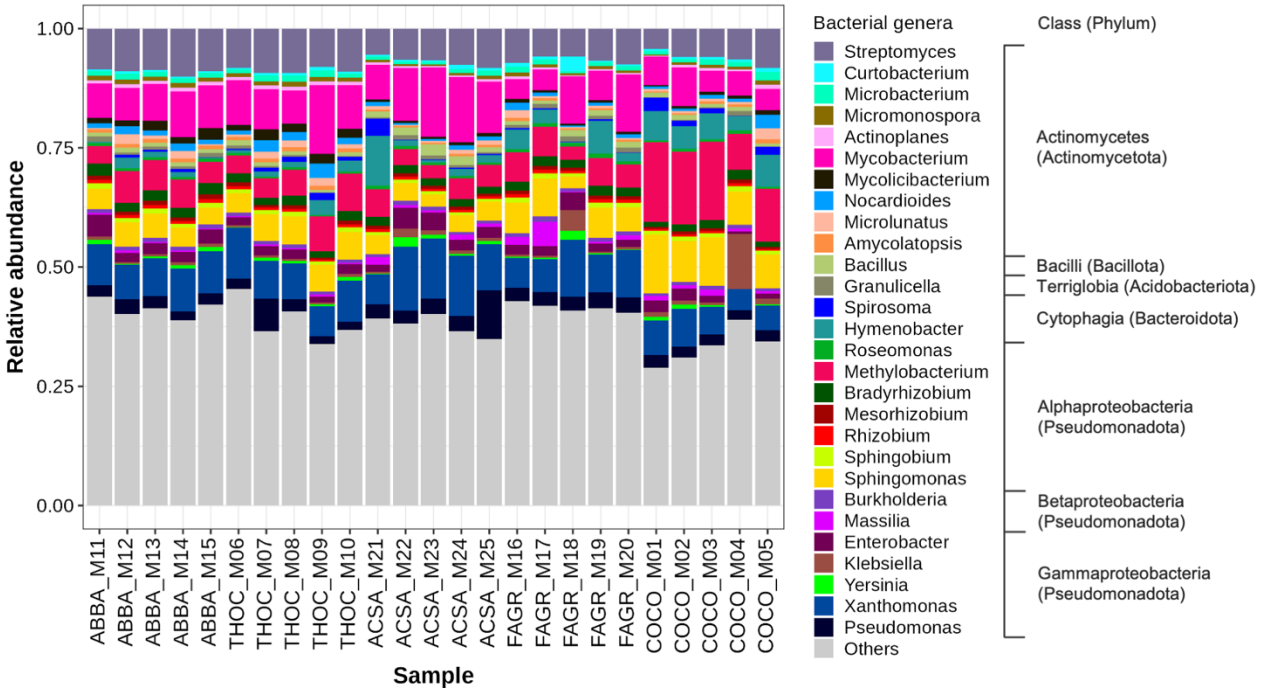


Figure 1.3 The proportion of bacterial genera composing the phyllosphere communities varied among the 25 samples and among the five host species. The barplot illustrates the 28 most abundant genera, ordered by class and phylum. "Others" category includes all genera whose reads did not account for at least 0.5% of the total bacterial reads across samples.

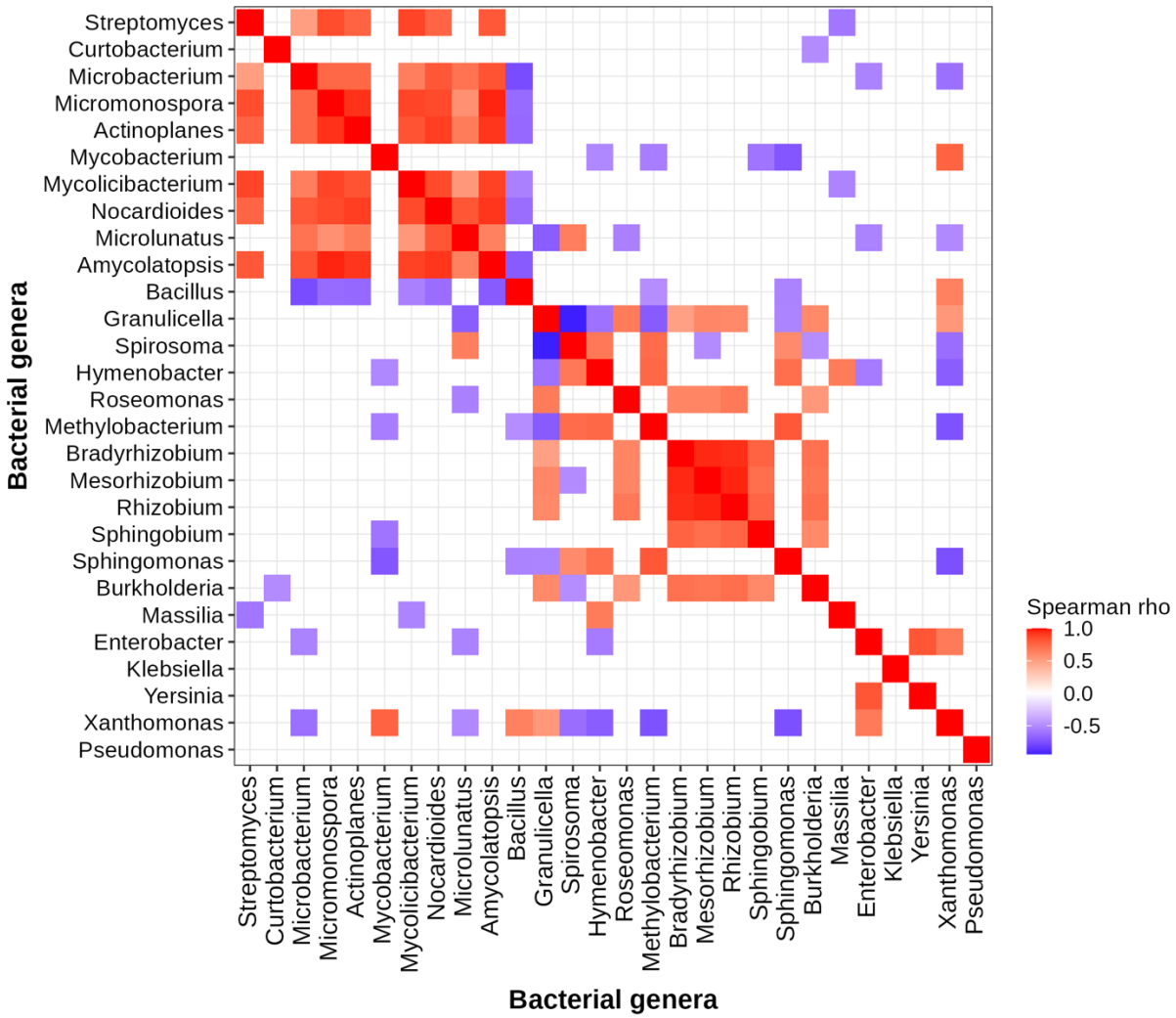


Figure 1.4 *Spirosoma*, *Sphingomonas*, and *Hymenobacter* genera covaried positively with *Methylobacterium* in the phyllosphere (notwithstanding the host species), while *Bacillus*, *Mycobacterium*, *Granulicella*, and *Xanthomonas* genera covaried negatively. Only the 28 most abundant genera in bacterial communities are shown (> 0.5% of bacterial reads across samples). Colours, representing Spearman rho, have been added only for significatively covarying pairs ($p < 0.05$).

By removing read counts from genera that did not account for at least 0.1% of the total bacterial reads of a given community, we kept 82.07% of all reads across samples shared among 198 genera (9.59% of genera originally "detected" by kraken2). We found that host species explained 47.05% of genera compositional variation among host species' bacterial communities (PERMANOVA, $p < 0.001$, $R^2 = 0.470$), while the within host species variances were similar (betadisper, $p = 0.407$). Furthermore, genus composition also differed between *C. cornuta* communities and tree species communities, as well as between conifers and broadleaf trees, even

though the conifers harboured more homogenous communities than broadleaf trees (Appendix A, Table A.4). The first axis of the PCoA explained 37.91% of the compositional variation, and highlighted the distinction between *C. cornuta* and other host species communities (Figure 1.5). The second axis explained 21.95% of the variation, and highlighted the distinction in composition between conifers and broadleaf host species.

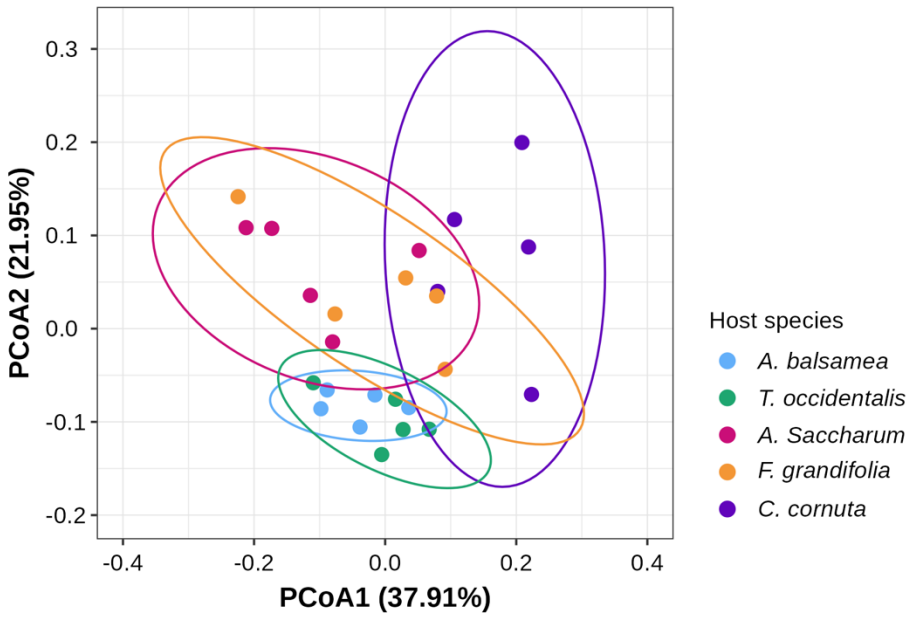


Figure 1.5 Genus-level community composition differed between *C. cornuta* and tree species, as well as between conifer and broadleaf tree species (PERMANOVAs, p = 0.001), as illustrated by the two main axes of this Principal Coordinates Analysis. The relatively smaller distance among conifer communities compared to the broadleaf communities reflect their higher homogeneity in genus composition (betadisper, p = 0.011). Bray-Curtis dissimilarity coefficient was used to calculate distances between pairs of samples. Ellipses indicate 95% confidence intervals.

1.5.3 *Methylobacterium*'s pangenome is mainly composed of accessory genes

Together, the 104 genomes were composed of a total of 620,585,276 nucleotides spanning 14,881 contigs. Prodigal identified 597,777 genes on these genomes. After removing 3,126 contigs < 1000 bp, we lost 0.25% of total nucleotides, and 3724 genes were excluded from the pangenome. The pangenome contained a total of 594,053 genes grouped in 70,695 gene clusters based on the similarity of their amino acid sequences. Of these genes, COG functions were assigned to 435,527 genes, KO functions to 306,272 genes, and Pfam functions to 448,027 genes. We identified 2214 core gene clusters (3.13% of the pangenome), of which 1844 gene clusters were present in at least 100 genomes. We identified 68,481 accessory gene clusters (96.87% of

the pangenome), among which 38,190 (54.02% of the total pangenome) were singletons (i.e., gene clusters found in only one genome) (Appendix A, Figure A.4). Of the 2214 core gene clusters, 105 were not assigned to any function (4.74% of core pangenome), while 36,889 accessory gene clusters (53.87% of accessory pangenome) were not functionally assigned. The pangenome we constructed for our analyses was open (Appendix A, Figure A.5).

1.5.4 The diversity and composition of *Methylobacterium* species communities diverged among host species

From whole phyllosphere bacterial communities, we zoomed in our analyses to focus only on *Methylobacterium* species assemblages, with the objective of quantifying the influence of host plant species on the taxonomic diversity and composition of these assemblages. *Methylobacterium* species richness was not correlated with the number of filtered read pairs per sample (regression, p = 0.497, adj. R^2 = –0.023). Across all individual host trees sampled, we detected a total of 23 *Methylobacterium* species in 24 samples. Each species was detected in 1 to 22 individual samples. One *A. saccharum* sample didn't recruit any *Methylobacterium* genomes. *Methylobacterium sp. 018*, *M. sp. 021*, and *M. sp. 024* were the most abundant species at our study site in terms of total genome coverage across all hosts (Figure 1.6).

Methylobacterium's community species richness differed among host species (ANOVA, p = 0.007, adj. R^2 = 0.403) and was on average higher in *C. cornuta* communities (mean of 13.6 ± 4.3 species) than on those of trees (6.8 ± 3.2 species) (Figure 1.7) (Appendix A, Table A.5). The Shannon index of *Methylobacterium* communities did not vary with host species (KW, p = 0.336, η^2 = 0.029), neither did the Pielou index (ANOVA, p = 0.176, adj. R^2 = 0.125), although we observed a tendency for *C. cornuta* samples to harbour slightly less even communities (Figure 1.8). Pielou index decreased with species richness (regression, slope = -0.013, p = 0.004, adj. R^2 = 0.297), indicating that communities harbouring more species are dominated by a few very abundant ones (Figure 1.6).

In tree samples, 14 different *Methylobacterium* species were detected, while the shrub species *C. cornuta* harboured 23 different species. Conifer species harboured 12 *Methylobacterium* species (*A. balsamea* = 11 species, *T. occidentalis* = 12 species), and broadleaf trees harboured 11 *Methylobacterium* species (*A. saccharum* = 10 species, *F. grandifolia* = 11 species). Eight *Methylobacterium* species were present on at least one sample of all host species: *M. sp. 018*, *M. sp. 021*, *M. sp. 022*, *M. sp. 023*, *M. sp. 024*, *M. sp. 026*, *M. sp. 029*, and *M. sp. 030*. Nine

species were only present on *C. cornuta* samples : *M. mesophilicum*, *M. pseudosasicola*, *M. sp. 004*, *M. sp. 011*, *M. sp. 034*, *M. haplocladii*, *M. sp. 003*, *M. thuringiense*, and *M. trifolii* – the last four species being present in one *C. cornuta* sample only. *M. sp. 025* was only absent from *F. grandifolia* samples, while *M. sp. 027* was absent from both conifer species. Except for *M. cerastii*, all the *Methylobacterium* species that were the most prevalent in our samples are new candidate species whose strains have previously been isolated from either our study site or from another deciduous forest of southern Quebec (Leducq *et al.*, 2022a). Those strains had been isolated from either *A. saccharum* or *F. grandifolia* (Table 1.1), but we did not observe any pattern of association between the original strain's host species and the abundance of the corresponding *Methylobacterium* species on the original host (Figure 1.6).

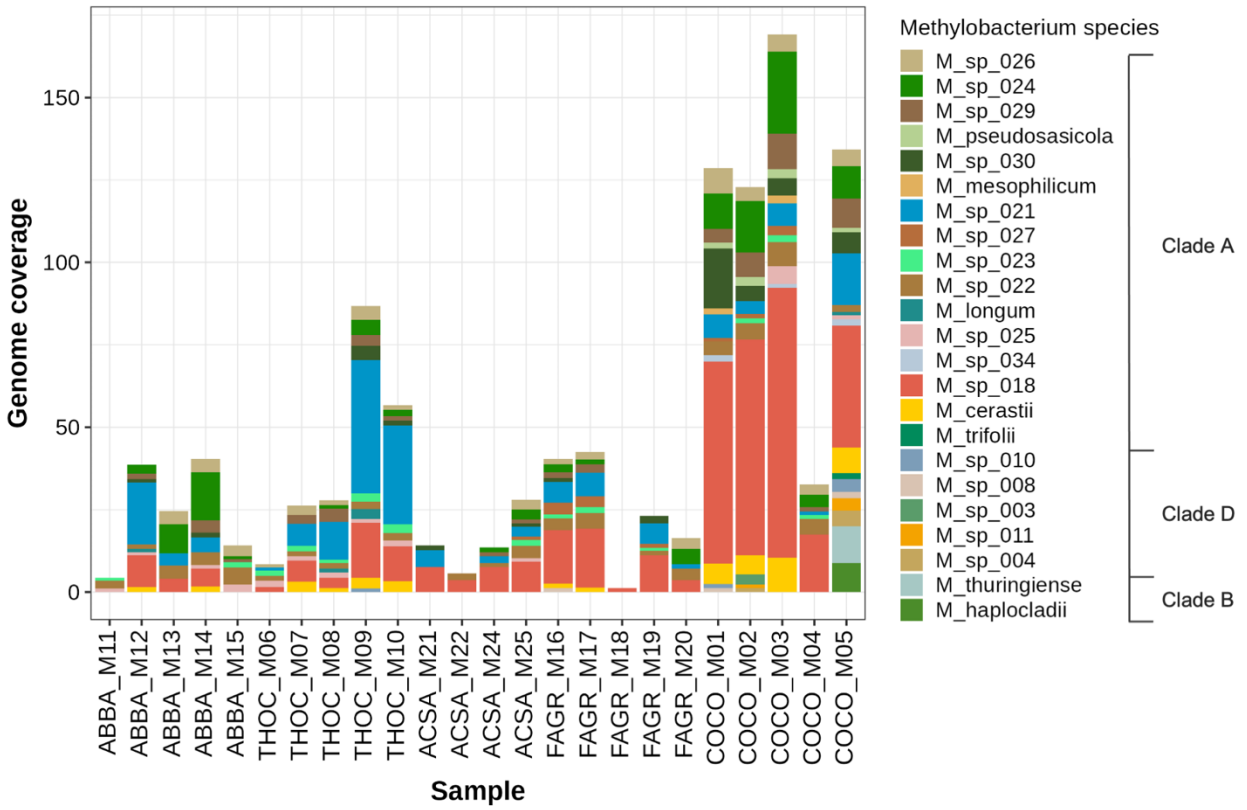


Figure 1.6 *Methylobacterium sp. 018*, *M. sp. 021*, and *M. sp. 024* were the most abundant species in terms of total genome coverage across all samples and hosts. A total of 23 *Methylobacterium* species – colour-coded and listed according to their corresponding clade – were detected in 24 samples. ABBA, *Abies balsamea*; ACSA, *Acer saccharum*; COCO, *Corylus cornuta*; FAGR, *Fagus grandifolia*; THOC, *Thuja occidentalis.*

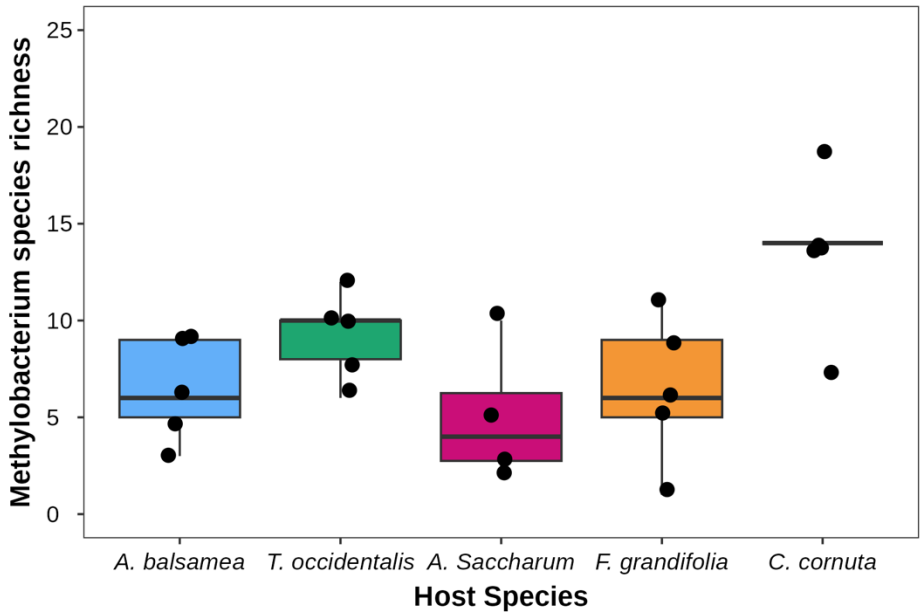


Figure 1.7 *Methylobacterium*'s community species richness was higher in *C. cornuta* samples (p = 0.040). The shrub species harboured in average over 13 *Methylobacterium* species, while tree species generally harboured around 7 species.

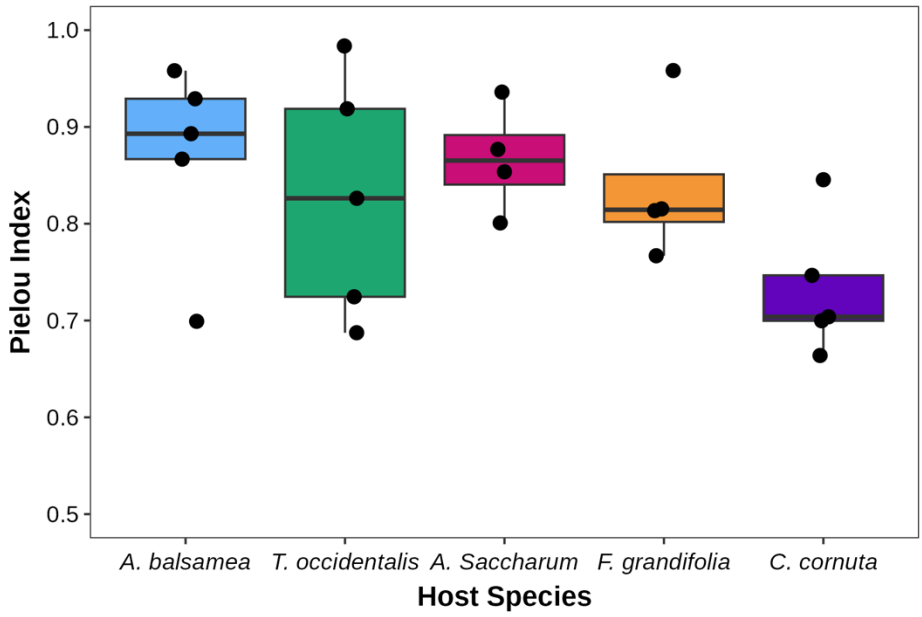


Figure 1.8 *Methylobacterium*'s community evenness, as measured by the Pielou index, did not significantly differ among the five host species (ANOVA, p = 0.176), although *C. cornuta* communities, which are richer in species, tended to be slightly less even.

Beta diversity partitioning analysis revealed that balanced variation in abundances explained 34.43% of the measured differences in composition among all samples, while abundance gradients explained 65.57% of the differences (Table 1.2). Within host species beta diversity was also largely due to abundance gradients, except for *A. balsamea* samples in which balanced variation accounted for 55.77% of the total beta diversity. Host species explained 35.5% of the compositional differences in *Methylobacterium* species communities (PERMANOVA, $p = 0.001$, $R^2 = 0.355$), and group variances did not differ significantly (betadisper, $p = 0.762$). *C. cornuta* samples harboured the most distinct *Methylobacterium* species communities (Figure 1.9) which were significatively different from the communities of the tree hosts. Conifer and broadleaf trees also differed in species composition (Appendix A, Table A.6).

Table 1.2 The beta diversity partitioning between balanced variation in abundances and abundance gradients assessed the nature of compositional differences among all samples as well as among samples of the same host species.

	Total *β* div.	**bal. var.**	**abund. grad.**	**bal. var./ Total *β* div.**	**abund. grad./ Total *β* div.**
All samples	0.3183	0.1096	0.2087	0.3443	0.6557
A. balsamea	0.3290	0.1835	0.1455	0.5577	0.4423
T. occidentalis	0.2574	0.0419	0.2155	0.1627	0.8373
A. saccharum	0.2418	0.0811	0.1607	0.3354	0.6646
F. grandifolia	0.3127	0.0679	0.2448	0.2172	0.7828
C. cornuta	0.2180	0.0803	0.1377	0.3682	0.6318

Total *β* div., total beta diversity; bal. var., balanced variation in abundances; abund. grad., abundance gradients.

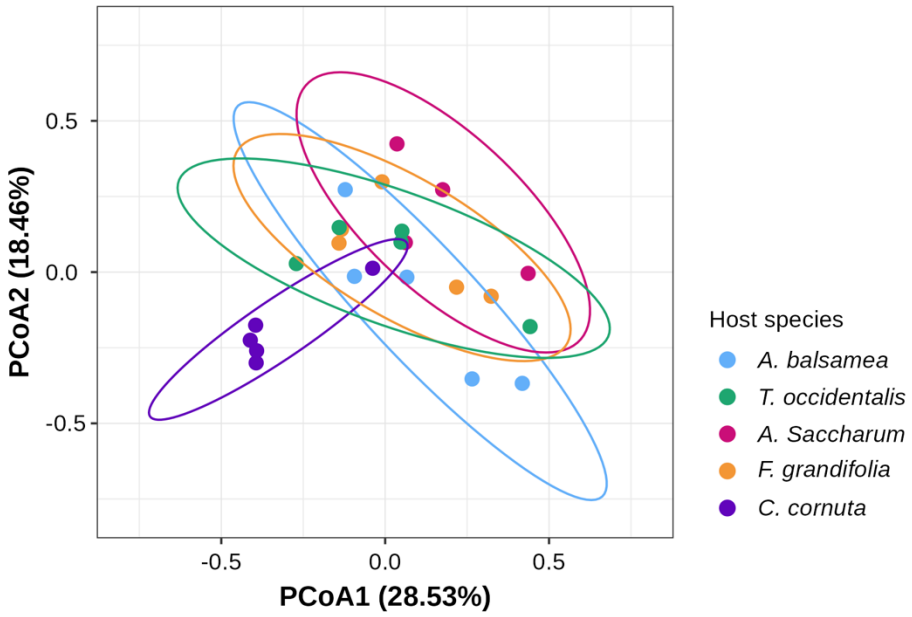


Figure 1.9 *Methylobacterium* taxonomic community composition differed markedly between *C. cornuta* and tree species (PERMANOVA, $p < 0.001$), as illustrated by the first axis of this Principal Coordinates Analysis. The composition also varied between conifer and broadleaf trees communities (PERMANOVA, $p = 0.048$), but the difference is less clearly illustrated by the first two axes of the ordination. Bray-Curtis dissimilarity coefficient was used to calculate distances between pairs of samples. Ellipses indicate 95% confidence intervals.

Finally, we looked for phylogenetic patterns in *Methylobacterium* taxonomic community structure. Most *Methylobacterium* communities were composed of phylogenetically closely related species as indicated by the standardized effect size of MPD (mean of z-scores = –2.61 ± 1.41) (Appendix A, Table A.7). However, host species did not influence phylogenetic relatedness (ANOVA, $p = 0.729$, adj. $R^2 = –0.098$) (Appendix A, Figure A.6), nor did they harbour different phylogenetic groups of *Methylobacterium* species (PERMANOVA, $p = 0.215$, $R^2 = 0.184$). Figure 1.10 illustrates the genome coverage of *Methylobacterium* species in different samples in relation to their phylogeny, highlighting the abundance of clade A (blue branches) in communities, compared to clades D (red) and B (orange).

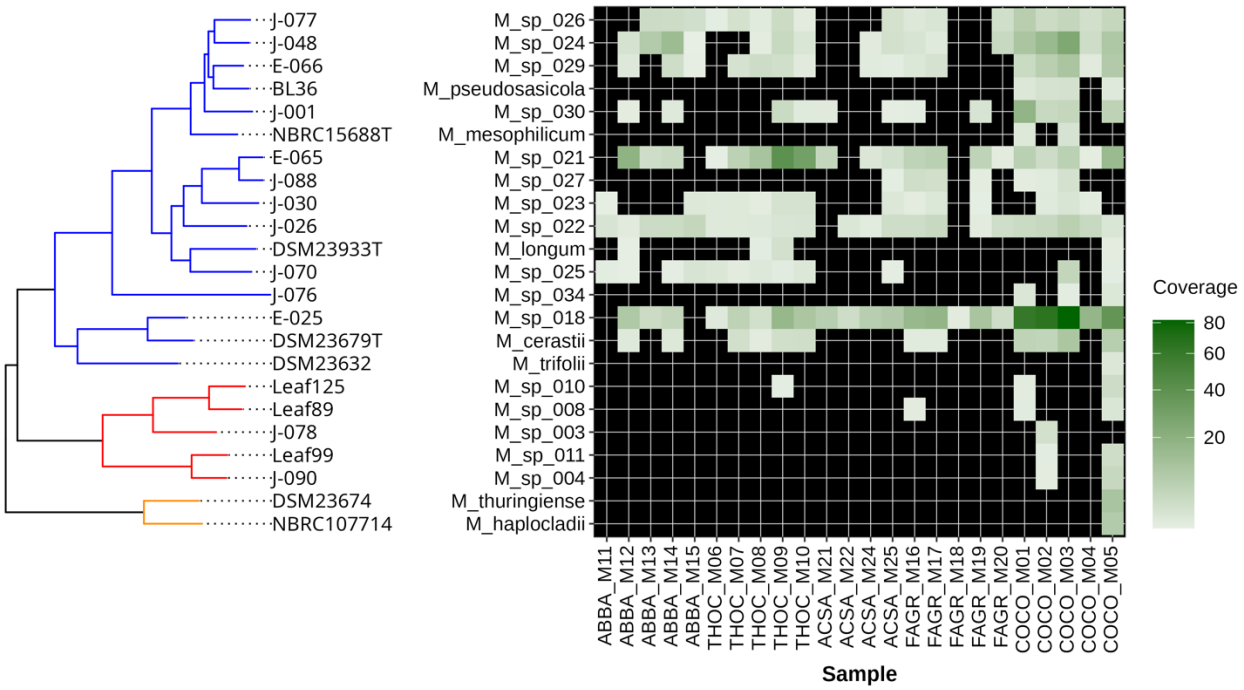


Figure 1.10 Genome coverage values of the 23 detected *Methylobacterium* species, presented alongside their phylogenetic relationships, highlight the phylogenetic clustering of *Methylobacterium* communities and the dominance of clade A. Host species did not influence the degree of clustering (ANOVA, p = 0.729) nor the phylogenetic composition of *Methylobacterium* communities (PERMANOVA, p = 0.215). Tree tip labels correspond to strain names whose genomes were used to represent the corresponding species. Orange-coloured tree branches represent clade B; red branches, clade D; and blue branches, clade A.

1.5.5 Host species explained 30% of the genetic composition of *Methylobacterium* communities

From species-level analyses, we zoomed in further to investigate the host species effect on the genetic structure of *Methylobacterium* communities. Gene cluster alpha diversity metrics did not vary among host species; richness was similar (ANOVA, p = 0.566, adj. R^2 = –0.046), as well as the Shannon index (KW, p = 0.346, η^2 = 0.026). As for beta diversity, host species explained close to 30% of the variation in gene cluster composition of *Methylobacterium* communities (PERMANOVA, p < 0.001, R^2 = 0.296). The compositional variance was homogenous among host species (betadisper, p = 0.469). Gene cluster composition varied between *C. cornuta* and the four tree species, but the compositional variance was higher within tree samples. Gene cluster composition also varied between conifers and broadleaf tree species, but with equal group compositional variance (Appendix A, Table A.8). The differences in gene cluster composition of *Methylobacterium* communities were illustrated by a PCoA analysis in which the first axis

explained 21.56% of the variation and differentiated gene communities of *C. cornuta* and trees, and in which the second axis explained 10.68% of the variation and differentiated conifer and broadleaf hosts (Figure 1.11).

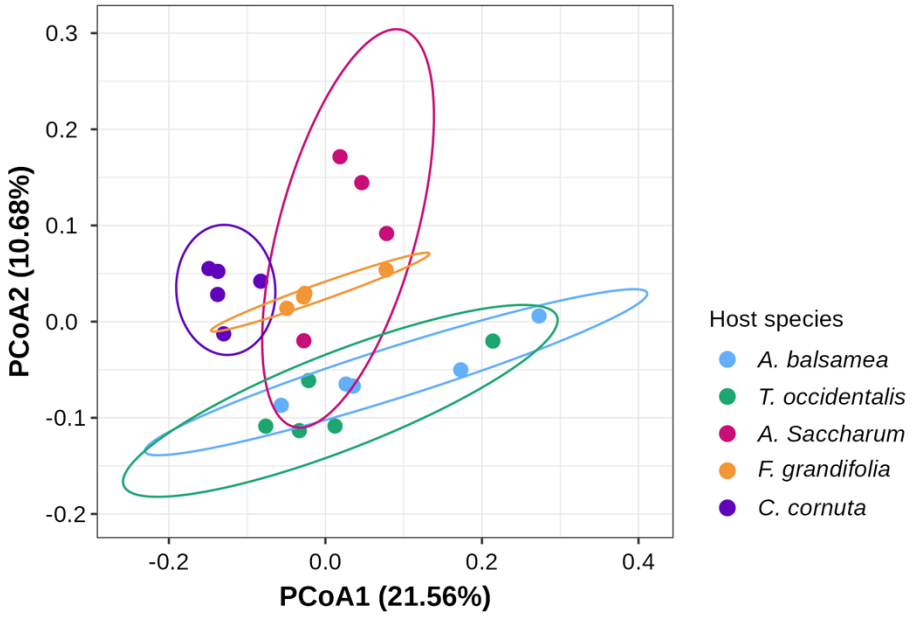


Figure 1.11 Gene cluster composition of *Methylobacterium* communities differed between *C. cornuta* and tree species, as well as between conifer and broadleaf species (PERMANOVAs, p = 0.001), as respectively illustrated by the first and second axes of this Principal Coordinates Analysis. Shrub-inhabiting communities had a more homogenous gene composition than tree-inhabiting communities (betadisper, p = 0.003), as reflected by the *C. cornuta* samples' cluster on the ordination. A Bray-Curtis dissimilarity matrix was computed to calculate distances between pairs of samples by performing rarefaction on gene clusters counts (n = 44,221 counts; 100 iterations). Ellipses indicate 95% confidence intervals.

1.5.6 Hundreds of potentially adaptive *Methylobacterium* gene clusters were associated with host species and host types

In order to identify *Methylobacterium* genes potentially conferring adaptation to different host species, we analyzed the variation in gene cluster abundance. A total of 1159 gene clusters (5.22% of originally present gene clusters) were differentially abundant among host species and host types, of which 1050 (90.6%) were accessory gene clusters, and 757 (65.31%) were functionally annotated. The log(2) fold change between the focus hosts and the reference hosts varied between 0.5265 (1.44X more abundant) and 5.3860 (41.82X more abundant). A figure illustrating all 1159 DAGCs including their normalized abundance within samples, accessory/core status and log(2) fold change for the associated comparison can be found online at

https://github.com/lauzonj/methylobacterium_tree_hosts/blob/main/DAGCs_heatmap.pdf.
A table containing all DAGCs' functions (COG, KOfam, Pfam) and differential abundance analyses' statistics can be found at
https://github.com/lauzonj/methylobacterium_tree_hosts/blob/main/DAGCs_stats_functions.csv.

We found 216 DAGCs (18.6% of all DAGCs) between tree hosts and the shrub species, of which 176 DAGCs were associated with trees, and 40 DAGCs were associated with *C. cornuta* (Figure 1.12). We found 724 DAGCs (62.5% of DAGCs) between broadleaf hosts and conifer hosts, of which 454 DAGCs were associated with broadleaves, and 270 DAGCs were associated with conifers. Only 15 DAGCs (1.3%) were uniquely found between broadleaf tree hosts (without considering the shrub *C. cornuta*) and conifer hosts (11 DAGCs associated with broadleaf trees, and 4 DAGCs associated with conifers), highlighting the fact that most broadleaf-associated DAGCs were not tree-form specific. Supporting this observation and revealing that *C. cornuta* is closer in gene cluster composition to broadleaf trees than to conifer species, only 20 DAGCs (1.7%) were uniquely found between broadleaf trees and the broadleaf shrub (16 DAGCs associated with broadleaf trees, and 4 DAGCs with *C. cornuta*), while 183 DAGCs (15.8%) were uniquely found between conifer species and the shrub (127 DAGCs associated with conifers, and 56 DAGCs with *C. cornuta*). No DAGC was found between the two broadleaf tree species (*A. saccharum* vs. *F. grandifolia*), nor between the two conifer species (*A. balsamea* vs. *T. occidentalis*).

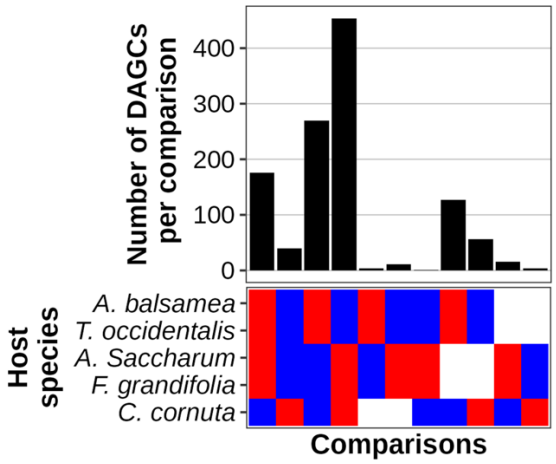


Figure 1.12 The number of differentially abundant gene clusters (DAGCs) differed among host species and host types within each comparison type. The numbers of DAGCs include gene clusters without functional annotations. Comparison types are illustrated as red and blue tiles, where red represents the focal host species or group, and blue represents the reference level against which the focal host(s) is(are) compared.

Among the functionally annotated DAGCs, the most abundant COG categories represented across all DAGCs were ‘Amino acid transport and metabolism’, ‘Signal transduction mechanisms’, ‘General function prediction only’, ‘Transcription’, ‘Inorganic ion transport and metabolism’, and ‘Carbohydrate transport and metabolism’ (Figure 1.13, right panel). Those 6 categories made more than 50% of functionally annotated DAGCs. There was 47 DAGCs uniquely assigned to the ‘Function unknown’ category (not shown in figure). Figure 1.13 (centre panel) illustrates the number of DAGCs related to COG categories in each comparison type. ‘Energy production and conversion’, ‘Translation, ribosomal structure and biogenesis’, and ‘Amino acid transport and metabolism’ were the 3 most prevalent COG categories represented by tree-enriched gene clusters, while ‘Mobilome’ and ‘Inorganic ion transport and metabolism’ were the two most prevalent COG categories represented by shrub-enriched gene clusters. Conifer-enriched gene clusters were more related to ‘Amino acid transport and metabolism’, ‘Lipid transport and metabolism’, and ‘General function predictions only’, whereas broadleaf-enriched gene clusters were more related to ‘Cell wall/membrane/envelope biogenesis’, ‘Signal transduction mechanisms’, ‘Transcription’, and ‘Defence mechanisms’.

Figure 1.14 and Figure 1.15 illustrate all functionally annotated DAGCs with a *p*-value < 0.01 and an absolute log(2) fold change value ≥ 2, respectively more prevalent within host tree species samples or the shrub species *C. cornuta* (Appendix A, Figure A.8), and within broadleaf or conifer host species samples (Appendix A, Figure A.9). We found 23 functionally annotated gene clusters, including 5 core clusters, that matched the above-mentioned criteria (*p*, fold change) and that were more abundant within tree host samples than within the shrub samples, and found 8 gene clusters more abundant in *C. cornuta.* We found 45 functionally annotated gene clusters matching the above criteria that were more abundant within broadleaf samples, and 19 within conifer samples.

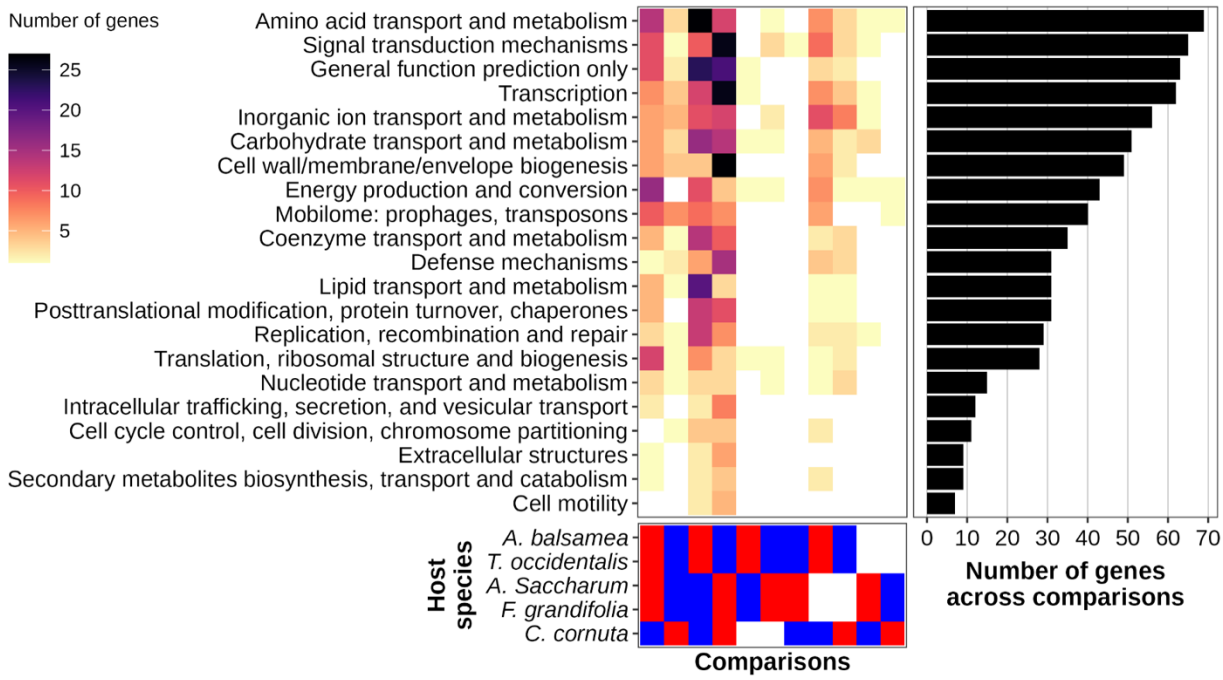


Figure 1.13 The abundance of COG functional categories of DAGCs varied among host species and host types within each comparison type (centre heatmap). DAGCs assigned to the category 'Function unknown' were excluded from the figure, as well as unannotated DAGCs. The right panel illustrates the total number of DAGCs in each COG category across all comparison types. Comparison types (bottom) are illustrated as red and blue tiles, where red represents the focal host species or group, and blue represents the reference level against which the focal host(s) is(are) compared. The number of genes (DAGCs) related to each COG category across comparison types is illustrated (right panel).

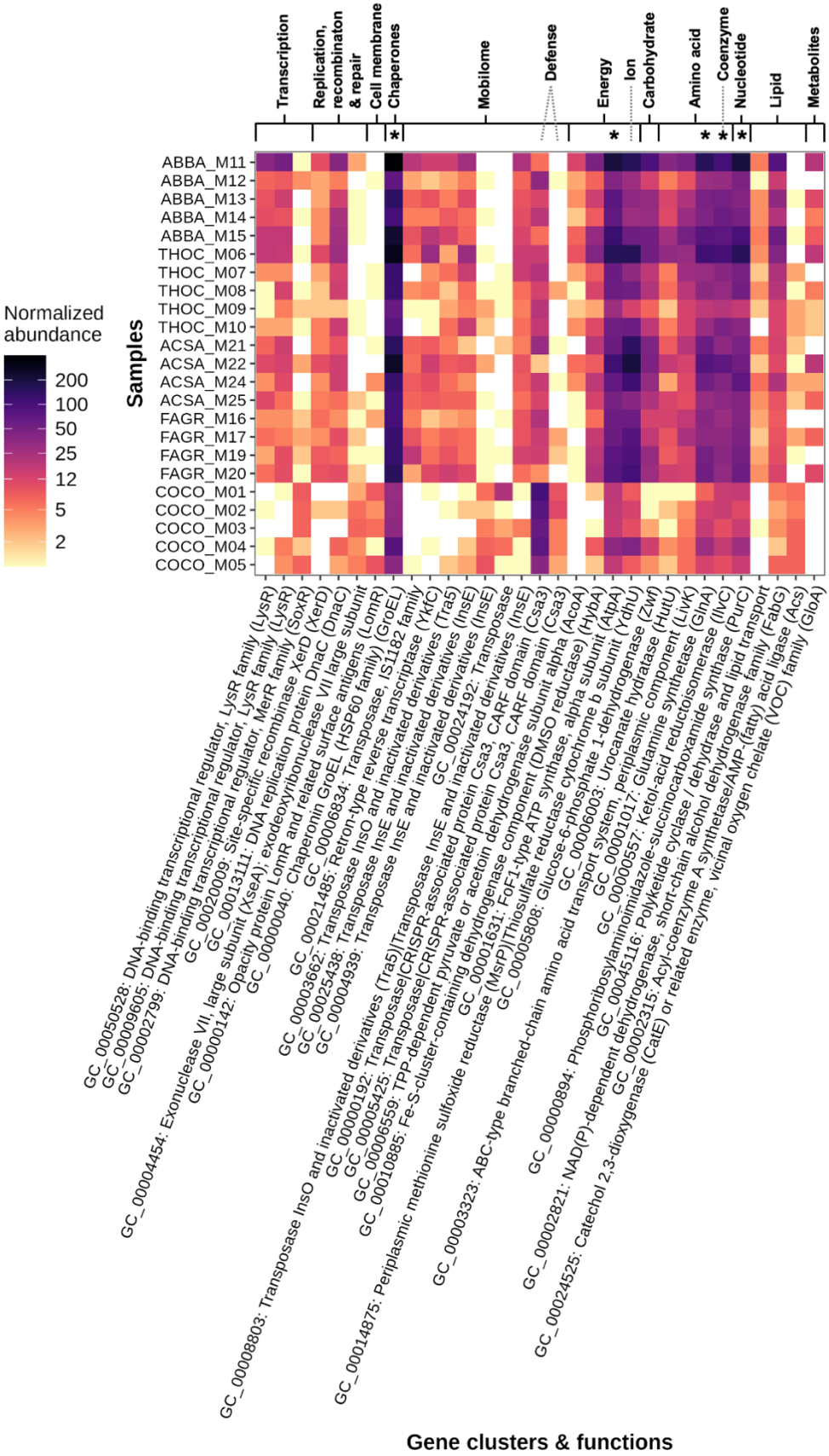

Transcription
Replication, recombinaton & repair
Cell membrane
Chaperones
Mobilome
Defense
Energy
Ion
Carbohydrate
Amino acid
Coenzyme
Nucleotide
Lipid
Metabolites
Normalized abundance
200
100
50
25
12
5
2
Samples
ABBA_M11
ABBA_M12
ABBA_M13
ABBA_M14
ABBA_M15
THOC_M06
THOC_M07
THOC_M08
THOC_M09
THOC_M10
ACSA_M21
ACSA_M22
ACSA_M24
ACSA_M25
FAGR_M16
FAGR_M17
FAGR_M19
FAGR_M20
COCO_M01
COCO_M02
COCO_M03
COCO_M04
COCO_M05
GC_00050528: DNA-binding transcriptional regulator, LysR family (LysR)
GC_00009605: DNA-binding transcriptional regulator, LysR family (LysR)
GC_00002799: DNA-binding transcriptional regulator, MerR family (SoxR)
GC_00020009: Site-specific recombinase XerD (XerD)
GC_00013111: DNA replication protein DnaC (DnaC)
GC_00004454: Exonuclease VII, large subunit (XseA); exodeoxyribonuclease VII large subunit
GC_00000142: Opacity protein LomR and related surface antigens (LomR)
GC_00000040: Chaperonin GroEL (HSP60 family) (GroEL)
GC_00003662: Transposase InsO and inactivated derivatives (Tra5)|Transposase InsE and inactivated derivatives (InsE)
GC_00021485: Retron-type reverse transcriptase, IS1162 family
GC_00025438: Transposase InsO and inactivated derivatives (Tra5)
GC_00004939: Transposase InsE and inactivated derivatives (InsE)
GC_00008803: Transposase InsO and inactivated derivatives (Tra5)|Transposase InsE and inactivated derivatives (InsE)
GC_00000192: Transposase/CRISPR-associated protein Csa3, CARF domain (Csa3)
GC_00005425: Transposase/CRISPR-associated protein Csa3, CARF domain (Csa3)
GC_00006559: TPP-dependent pyruvate or acetoin dehydrogenase component (DMSO reductase subunit alpha (AcoA)
GC_00014875: Periplasmic methionine sulfoxide reductase (MsrP)|Thiosulfate reductase (HybA)
GC_00010885: Fe-S-cluster-containing dehydrogenase component (DMSO reductase subunit (AtpA)
GC_00024192: Transposase
GC_00001631: FoF1-type ATP synthase, alpha subunit (YdhU)
GC_00005808: Glucose-6-phosphate cytochrome b subunit (Zwf)
GC_00006003: Urocanate hydratase (HutU)
GC_00003323: ABC-type branched-chain amino acid transport system, periplasmic component (LivK)
GC_00000894: Phosphoribosylaminoimidazole-succinocarboxamide synthetase (GlnA)
GC_00001017: Glutamine synthetase (GlnA)
GC_00045116: Polyketide cyclase / dehydrase and lipid transport
GC_00000557: Ketol-acid reductoisomerase (IlvC)
GC_00002821: NAD(P)-dependent dehydrogenase, short-chain alcohol dehydrogenase family (PurC)
GC_00002315: Acyl-coenzyme A synthetase/AMP-(fatty) acid ligase (FabG)
GC_00024525: Catechol 2,3-dioxygenase (CatE) or related enzyme, vicinal oxygen chelate (VOC) family (Acs)
GC_00000940: Aldehyde dehydrogenase (NAD+) (GloA)
Gene clusters & functions

Figure 1.14 The normalized abundances of 31 DAGCs in the 24 phyllosphere samples highlight annotated genes that were the most significantly enriched in communities inhabiting tree species or the shrub species *C. cornuta*. DAGCs are grouped by functional categories (based on COG). Gray dotted lines indicate alternative category. Core gene clusters are marked with an asterisk. The DAGCs shown are those that were functionally annotated and whose *p*–value < 0.01 and absolute log(2) fold change value ≥ 2. ABBA, *Abies balsamea*; ACSA, *Acer saccharum*; COCO, *Corylus cornuta*; FAGR, *Fagus grandifolia*; THOC, *Thuja occidentalis*.

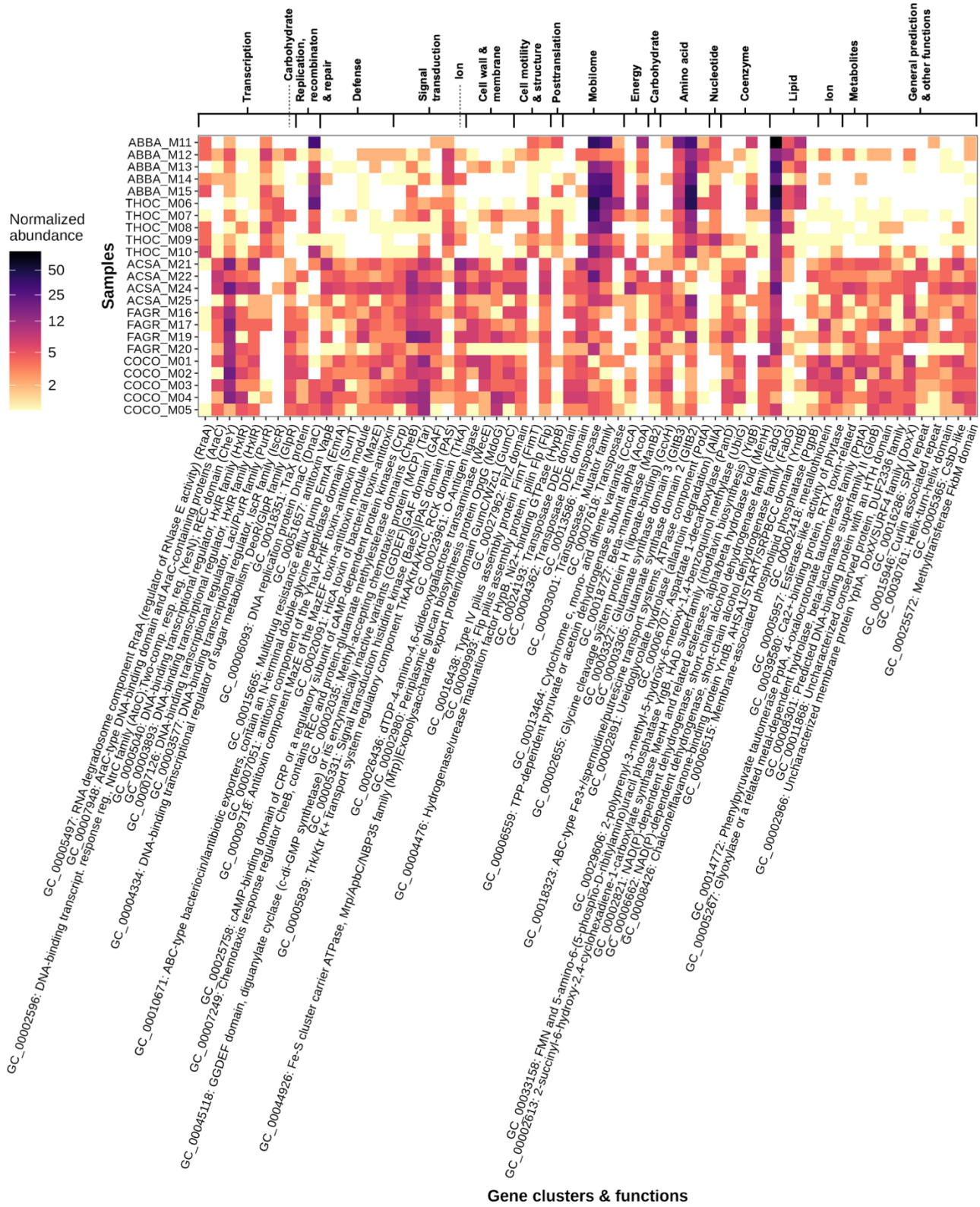

Transcription
Carbohydrate
Replication, recombination & repair
Defense
Signal transduction
Ion
Cell wall & membrane
Cell motility & structure
Posttranslation
Mobilome
Energy
Carbohydrate
Amino acid
Nucleotide
Coenzyme
Lipid
Ion
Metabolites
General prediction & other functions
Normalized abundance
50
25
12
5
2
Samples
ABBA_M11
ABBA_M12
ABBA_M13
ABBA_M14
ABBA_M15
THOC_M06
THOC_M07
THOC_M08
THOC_M09
THOC_M10
ACSA_M21
ACSA_M22
ACSA_M24
ACSA_M25
FAGR_M16
FAGR_M17
FAGR_M19
FAGR_M20
COCO_M01
COCO_M02
COCO_M03
COCO_M04
COCO_M05
GC_00005497: RNA degradosome component RraA (regulator of RNase E activity) (RraA)
GC_00002596: DNA-binding transcript
GC_00007948: AtaC-type DNA-binding domain and AraC-containing proteins (AraC)
GC_00007949: response reg., NtrC family (AtoC):Two-comp. resp. reg. (YesN): REC domain (CheY)
GC_00005040: DNA-binding transcriptional regulator, HxlR family (HxlR)
GC_00003893: DNA-binding transcriptional regulator, LacI/PurR family (HxlR)
GC_00007126: DNA-binding transcriptional regulator, IscR family (PurR)
GC_00003577: DNA-binding transcriptional regulator of sugar metabolism, DeoR/GlpR family (IscR)
GC_00004334: DNA-binding transcriptional regulator, TrmB family (GlpR)
GC_00006093: DNA replication protein DnaC (TrmB)
GC_00018351: TraX protein
GC_00051657: antitoxin VapB
GC_00010671: ABC-type bacteriocin/lantibiotic exporters, contain an N-terminal double-glycine peptidase domain (EmrA)
GC_00015665: Multidrug resistance efflux pump EmrA (EmrA)
GC_00007051: antitoxin component of the YhaV-PrlF toxin-antitoxin module (SunT)
GC_00009716: Antitoxin component MazE of the MazEF toxin-antitoxin module (MazE)
GC_00045118: GGDEF domain, diguanylate cyclase (c-di-GMP synthetase) or its enzymatically inactive variants (GGDEF)
GC_00007249: Chemotaxis response regulator CheB, contains REC and protein-glutamate methylesterase domains (CheB)
GC_00025758: cAMP-binding domain of CRP or a regulatory subunit of cAMP-dependent protein kinases (Crp)
GC_00002091: HicA toxin of bacterial toxin-antitoxin module (Tar)
GC_00002035: Methyl-accepting chemotaxis protein (MCP) (Tar)
GC_00005331: Signal transduction histidine kinase (BaeS)/GAF domain (GAF)
GC_00044926: Fe-S cluster carrier ATPase, Mrp/ApbC/NBP35 family (Mrp)/Exopolysaccharide export protein/domain GumC/WZZ (GumC)
GC_00005839: TrkA/Ktr K+ transport system regulatory component TrkA/KtrA/KtrC, RCK domain (TrkA)
GC_00026436: dTDP-4-amino-4,6-dideoxygalactose transaminase (WecE)
GC_00002980: Periplasmic glucan biosynthesis protein OpgG (MdoG)
GC_00009993: Flp pilus assembly protein, pilin Flp (Flp)
GC_00044476: Hydrogenase/urease maturation factor HypB, Ni2+-binding GTPase (HypB)
GC_00024193: Transposase DDE domain
GC_00004362: Transposase DDE domain
GC_00003001: Transposase, Mutator family
GC_00013586: Transposase, Mutator family
GC_00013464: Cytochrome c, mono- and diheme variants (CccA)
GC_00002655: Glycine cleavage system protein H (lipoate-binding) (GcvH)
GC_00018727: Beta-mannanase (ManB2)
GC_00003327: Glutamate synthase, ATPase component 2 (GltB2)
GC_00003305: Glutamate synthase, ATPase component 1 (GltB3)
GC_00002891: Ureidoglycolate hydrolase (allantoin degradation) (UgH)
GC_00018323: ABC-type Fe3+/spermidine/putrescine transport systems, ATPase component (UbiG)
GC_00029606: 2-polyprenyl-3-methyl-5-hydroxy-6-metoxy-1,4-benzoquinol methylase (PanD)
GC_00006707: Aspartate 1-decarboxylase (PanD)
GC_00033158: FMN and 5-amino-6-(5-phospho-D-ribitylamino)uracil phosphatase YigB, HAD superfamily (AhlA)
GC_00002613: 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase MenH and Related esterases, alpha/beta hydrolase fold (MenH)
GC_00002821: NAD(P)-dependent dehydrogenase, short-chain alcohol dehydrogenase family (FabG)
GC_00006662: NAD(P)-dependent dehydrogenase, short-chain alcohol dehydrogenase family (FabG)
GC_00008426: Chalcone/flavanone-binding protein YndB, AHSA1/START/SRPBCC domain (YndB)
GC_00006515: Membrane-associated phospholipid phosphatase (PgpB)
GC_00014772: Phenylpyruvate tautomerase PptA, 4-oxalocrotonate tautomerase (PptA)
GC_00039580: Ca2+ Esterase-like activity or phytase
GC_00005957: Esterase-like activity or phytase
GC_00002418: metallothionein
GC_00005267: Glyoxylase or a related metal-dependent hydrolase beta-lactamase superfamily II (GloB)
GC_00008301: Predicted DNA-binding protein, RTX toxin-related
GC_00011868: Uncharacterized conserved protein YphA, DoxX/SURF4 family (DoxX)
GC_00002966: Uncharacterized membrane protein YphA, DoxX/SURF4 family, DUF2336 family
GC_00015946: Cutin associated repeat
GC_00030761: Helix-turn-helix domain
GC_00005365: CsbD-like
GC_00025572: Methyltransferase FkbM domain
Gene clusters & functions

Figure 1.15 The normalized abundances of 64 DAGCs in the 24 phyllosphere samples highlight annotated genes that were the most significantly enriched in communities inhabiting conifer or broadleaf species. DAGCs are grouped by functional categories (based on COG). Gray dotted lines indicate alternative category. The DAGCs shown are those that were functionally annotated and whose *p*–value < 0.01 and absolute log(2) fold change value ≥ 2. ABBA, *Abies balsamea*; ACSA, *Acer saccharum*; COCO, *Corylus cornuta*; FAGR, *Fagus grandifolia*; THOC, *Thuja occidentalis*.

1.5.7 Host-associated accessory genes under positive selection were mostly found in the *Methylobacterium sp. 018* population inhabiting *Corylus cornuta* leaves

Finally, we analyzed how different types of selection (positive, near-neutral and negative) varied among *Methylobacterium* populations, host species and accessory/core status of genes, based on the ratio of non-synonymous rate to synonymous rate of polymorphism (*pN*/*pS* ratio) measured within codon sequences. Purifying (i.e., negative) selection was the most prevalent type of selection acting on the three *Methylobacterium* species across all phyllosphere samples. Nevertheless, the proportion of positive selection, purifying selection, and near-neural evolution were different among *Methylobacterium* species ($p < 0.001$). *M. sp. 021* had a higher proportion of both positively selected protein-coding genes and near-neutral variation across samples than *M. sp. 022* and *M. sp. 018*. *M. sp. 022* had a higher proportion of positive selection than *M. sp. 018* (Figure 1.16) (Appendix A, Table A.9).

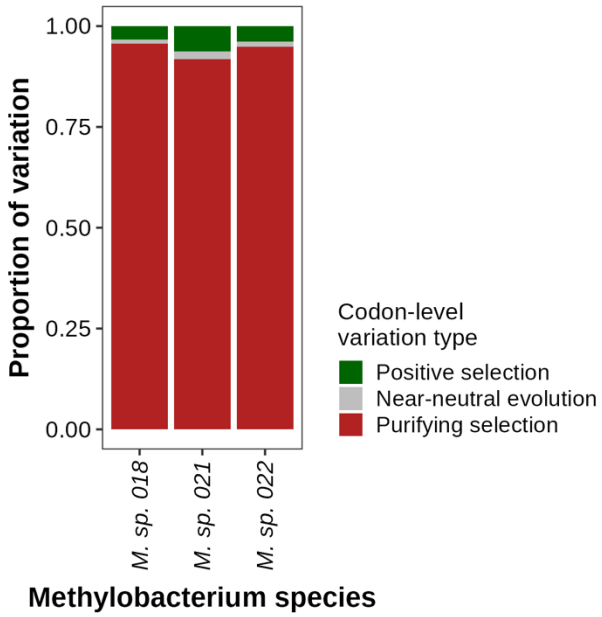


Figure 1.16 *Methylobacterium sp. 021* population contained a higher proportion of positively selected genes than *M. sp. 018* and *M. sp. 022* populations, across all host species ($p < 0.001$). Colours illustrate the proportions of the three selection types, assessed by the *pN*/*pS* ratio based on single codon variants within protein-coding genes.

The type of selection associated with codon-level variation differed among host species within *M. sp. 018* ($p < 0.001$), but not within *M. sp. 021* ($p = 0.067$) nor *M. sp. 022* ($p = 0.128$). In *M. sp. 018*, genes under positive selection were more prevalent in broadleaves than conifers ($p < 0.001$) (Figure 1.17). Overall, we detected more positive selection-driven variation in

accessory genes than in core genes ($p < 0.001$) as Figure 1.18 illustrates for *M. sp. 018* (a), *M. sp. 021* (b), and *M. sp. 022* (c).

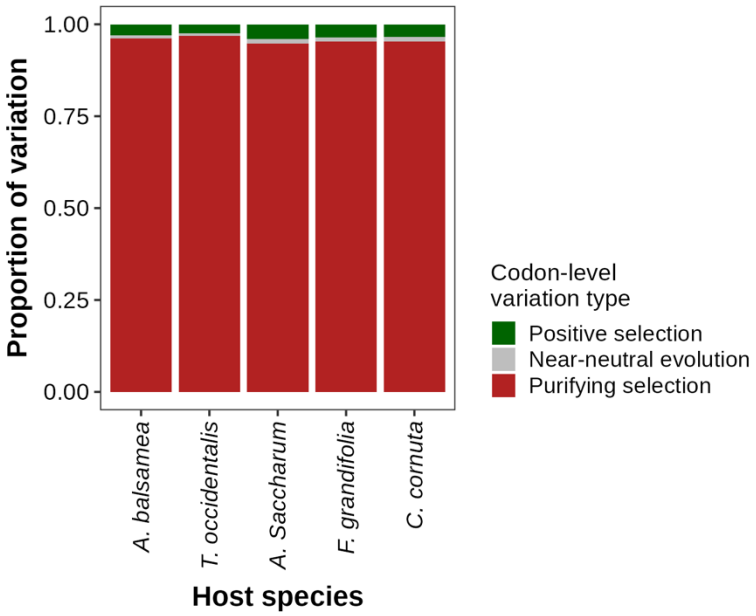
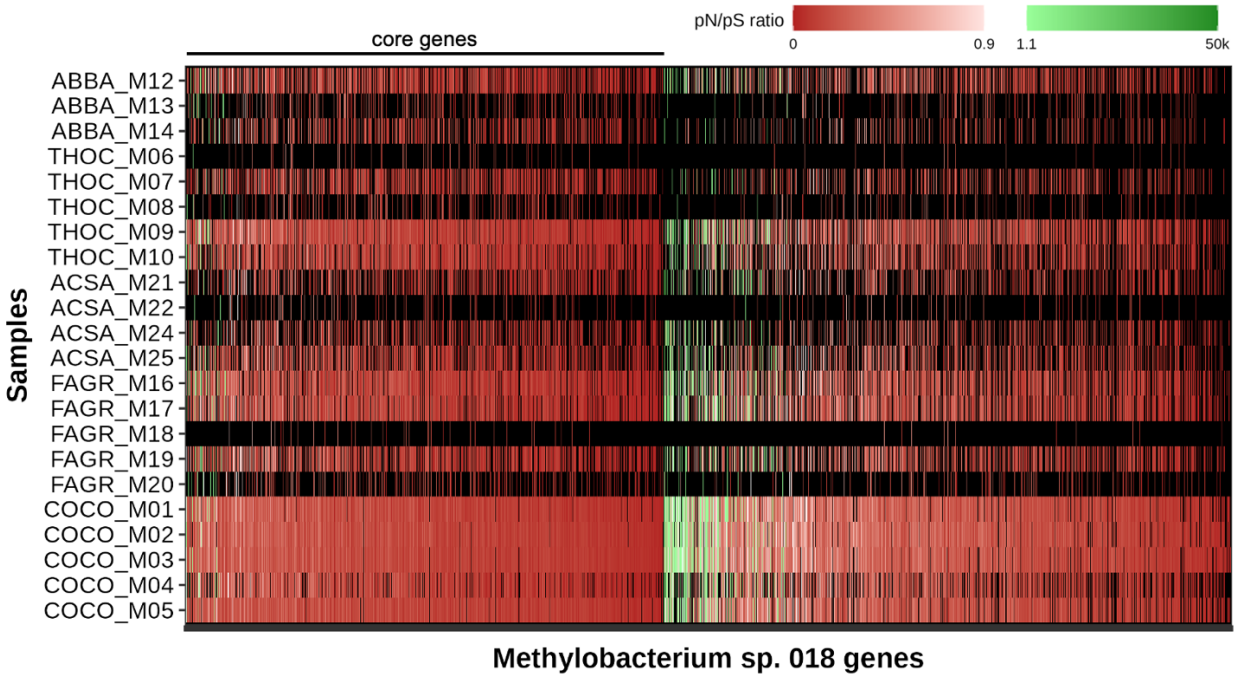
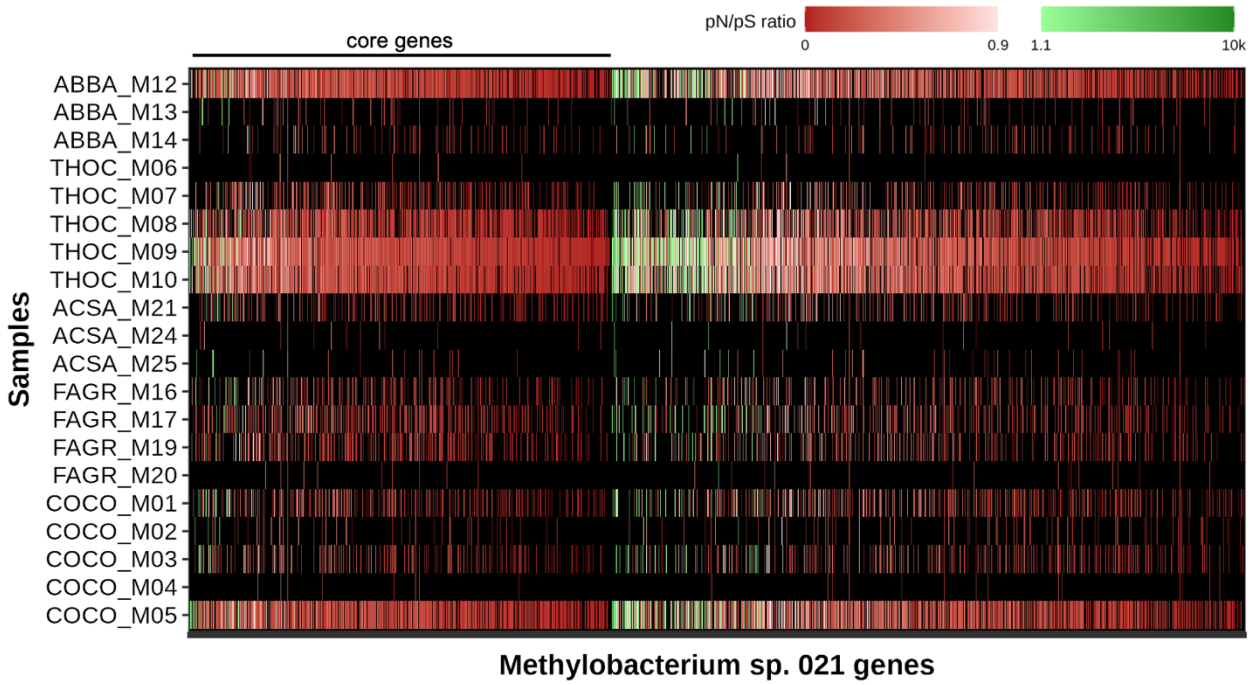


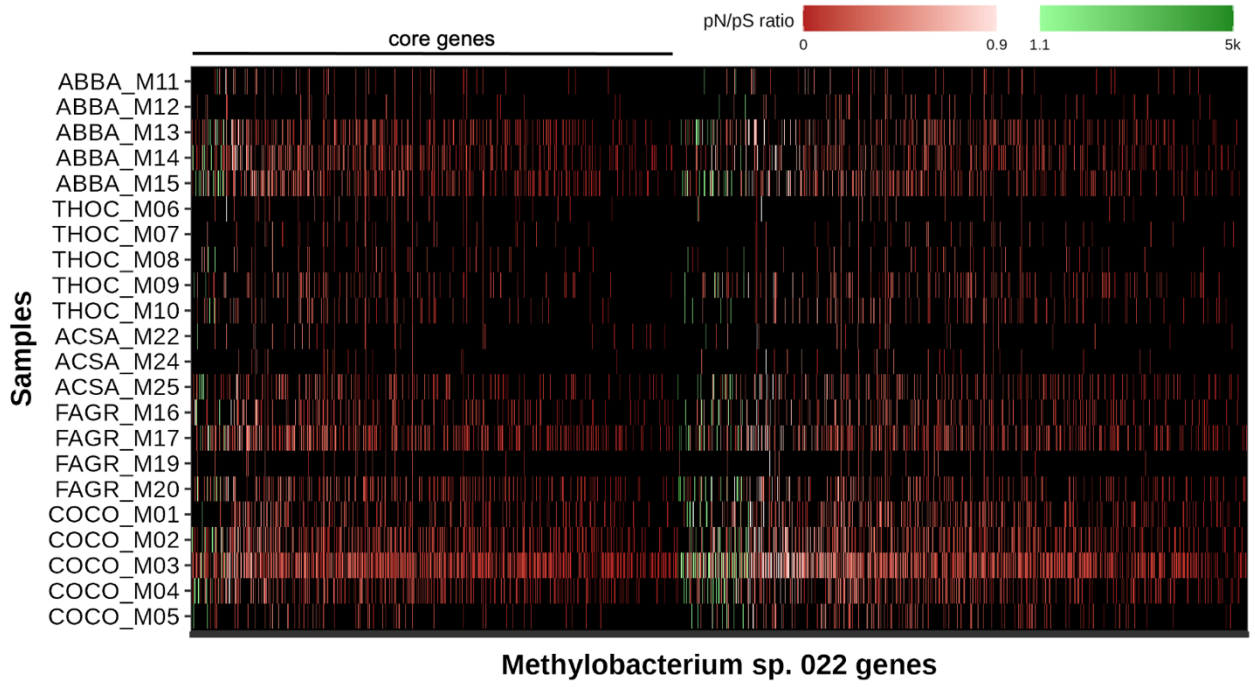
Figure 1.17 Within *Methylobacterium sp. 018* population, positive selection was more prevalent in broadleaf host species than in conifers ($p < 0.001$). Colours illustrate the proportions of the three selection types, assessed by the *pN*/*pS* ratio based on single codon variants within protein-coding genes.



(a)



(b)



(c)

Figure 1.18 Positive selection was relatively more prevalent in accessory genes than in core genes within populations of *Methylobacterium sp. 018*, *M. sp. 021*, and *M sp. 022*, across all host species ($p < 0.001$). The heatmap indicates the *pN*/*pS* ratio of all genes within genomes of *M. sp. 018* (a), *M. sp. 021* (b), and *M. sp. 022* (c), in which SCVs were detected in samples in which each species was present. Each column is a gene; genes are organized by core/accessory status, then by decreasing mean *pN*/*pS* ratio across samples. Green indicates positive selection (*pN*/*pS* > 1.1), red indicates purifying selection (*pN*/*pS* < 0.9), white indicates near-neutral evolution (1.1 ≥ *pN*/*pS*

≥ 0.9), and black indicates an absence of data (whether due to insufficient coverage depth or to the lack of codon-level variation).

Within the *M. sp. 018* recruited metagenome, we revealed 63 unique protein-coding genes (including 3 core genes) under positive selection (*pN*/*pS* > 1.1) detected on three host species, *T. occidentalis* (1 gene), *F. grandifolia* (8 genes) and *C. cornuta* (60 genes), for a total of 69 'gene-host' combinations (Table 1.3). We found that 25 of these genes were functionally annotated, and 25 genes were also DAGCs, mostly associated with broadleaves (in comparison to conifers). Within the *M. sp. 021* recruited metagenome, we revealed 13 unique accessory genes under positive selection, all detected on *T. occidentalis*. Five of those genes were functionally annotated, but none were DAGCs. Finally, within the *M. sp. 022* recruited genome, we revealed only one accessory gene under positive selection on *C. cornuta*, functionally unannotated and not differentially abundant. All positively selected genes' *pN*/*pS* statistics and t-test results, along with associated host species driving selection, core/accessory status, DAGCs type (if applicable) and functions (COG, KOfam, Pfam) can be found at https://github.com/lauzonj/methylobacterium_tree_hosts/blob/main/all_pos_sel_genes.csv.

Table 1.3 Thirty functionally annotated genes belonging to *Methylobacterium sp. 018* and *M. sp. 021* were identified as being under host-specific positive selection on either *C. cornuta*, *T. occidentalis*, and *F. grandifolia*. All genes presented have a mean *pN*/*pS* ratio > 1 (adj. $p < 0.05$) in samples of their associated host species. Functions presented are inferred from COG database; if no COG annotation was available for a given gene, then the function presented is from KOfam (KEGG), then Pfam.

gene id	mean *pN*/*pS*	*Methylobacterium* species > host species > gene function
		– *Methylobacterium sp. 018* –
		Host : *Thuja occidentalis*
66377	>100.00	VapC family ribonuclease, toxin component of the VapBC toxin-antitoxin module, contains PIN domain (VapC)
		Host : *Fagus grandifolia*
552276	2.98	Transposase | CRISPR-associated protein Csa3, CARF domain (Csa3)
66964*	2.71	Ceramidase
165750	2.70	ADP-ribosylglycohydrolase (DraG)
66581	1.53	Glutathione S-transferase (GstA)
		Host : *Corylus cornuta*
66716	>100.00	Plastocyanin (PetE)
165872	>100.00	Ribonuclease toxin, BrnT, of type II toxin-antitoxin system
66377	20.56	VapC family ribonuclease, toxin component of the VapBC toxin-antitoxin module, contains PIN domain (VapC)
378643	14.08	mRNA-degrading endonuclease RelE, toxin component of the RelBE toxin-antitoxin system (RelE)
166007	6.07	Predicted acetyltransferase, GNAT family
285352	2.75	NAD(P)-dependent dehydrogenase, short-chain alcohol dehydrogenase family (FabG)
552276	2.58	Transposase|CRISPR-associated protein Csa3, CARF domain (Csa3)
66844	2.46	Chemotaxis response regulator CheB, contains REC and protein-glutamate methylesterase domains (CheB)
66714	2.45	MHYT domain, NO-binding membrane sensor (MHYT) | Cyclic di-GMP metabolism protein, combines GGDEF and EAL domains with a 6TM membrane domain
66836	2.34	AraC-type DNA-binding domain and AraC-containing proteins (AraC)
551239	2.06	Serine kinase of the HPr protein, regulates carbohydrate metabolism (HprK)
285704	1.97	Putative restriction endonuclease
66827	1.96	DNA-binding transcriptional regulator YiaG, XRE-type HTH domain (YiaG)
66964*	1.89	Ceramidase
552146*	1.62	Sulfate permease or related transporter, MFS superfamily (SUL1)
458130	1.56	CheY-like REC (receiver) domain, includes chemotaxis protein CheY and sporulation regulator Spo0F (CheY) | PAS domain (PAS) | DNA-binding transcriptional response regulator, NtrC family, contains REC, AAA-type ATPase, and a Fis-type DNA-binding domains (AtoC) | Signal transduction histidine kinase regulating C4-dicarboxylate transport system
66535	1.50	Antitoxin component of toxin-antitoxin stability system, DNA-binding transcriptional repressor (Phd)
285765	1.39	toxin (YhaV)
166043	1.23	Flp pilus assembly protein, pilin Flp (Flp)
285443	1.23	Transposase InsO and inactivated derivatives (Tra5)
		– *Methylobacterium sp. 021* –
		Host : *Thuja occidentalis*
554371	>100.00	Transposase
461004	>100.00	Copper chaperone CopZ (CopZ)
554373	>100.00	DNA-directed RNA polymerase specialized sigma subunit, sigma24 family (RpoE)
288810	3.03	Glutathione S-transferase (GstA)
70696	1.95	Selenocysteine lyase/Cysteine desulfurase (CsdA)

* core gene

1.6 Discussion

1.6.1 The leaf surface of temperate woody species harbours diverse *Methylobacterium* species

Methylobacterium's ubiquity in the phyllosphere was reflected by its prevalence across all foliar epiphytic bacterial communities of the host species that were typical of the understory canopy on our forest plot, including conifers, broadleaves, trees and a shrub species. Based on a k-mer short-read classification, *Methylobacterium* genus accounted for a substantial proportion of leaf bacterial communities in the forest: across all our samples, around one bacterial cell out of fifteen (6.8%) belonged to *Methylobacterium* (Figure 1.3). This relatively high proportion of the phyllosphere communities for this single genus is in accordance with studies of herbaceous host species (2-20%) (Delmotte *et al.*, 2009; Knief *et al.*, 2012a). However, studies of the temperate tree phyllosphere in southern Quebec have generally found a lower relative abundance of *Methylobacterium* than measured in our analyses, varying from 1.26 to 2.3% across forest plots, tree species and growth season (Laforest-Lapointe *et al.*, 2016a, 2017b; Leducq *et al.*, 2022a).

Our metapangenomic approach allowed us to detect 23 distinct *Methylobacterium* species on five different host species (Figure 1.6), representing over a fifth of the genus's known taxonomic diversity. This high species richness associated with the temperate tree phyllosphere is in line with the study by Leducq *et al.* (2022a), who identified 15 *Methylobacterium* ASVs based on the 16S rRNA marker gene – across two forests, seven host species, and five months – as well as 200 ASVs based on the *rpoB* gene, a highly polymorphic single-copy core gene offering a strain-level resolution (Vos *et al.*, 2012) within the *Alphaproteobacteria* class. Accurate measurement of true bacterial species diversity is hard in nature (Malaterre, 2017; Øvreås and Curtis, 2011) as results depend on the reference framework used, the identification and measurement approach (Breitwieser *et al.*, 2019; Pérez-Cobas *et al.*, 2020; Tessler *et al.*, 2017), as well as how prokaryote species are delineated (Ereshefsky, 2010). True *Methylobacterium* species richness in our sampled communities might be higher than our results show since 1) some species might be present in very low relative abundance and therefore not recruit enough reads by their representative genome to pass our horizontal coverage threshold; 2) novel species might be present for which we did not have any reference genome; 3) highly genomically divergent populations might be ongoing speciation but were mapped to the same reference genome; and 4) reads from a truly present but undetected species might have been mapped to different genomes of phylogenetically close species, hence "diluting" the reads among many poorly covered genomes. Furthermore, bacterial assemblages evolve during the growth season (Leducq *et al.*,

2022a; Redford and Fierer, 2009) as leaf phytochemistry (Eller *et al.*, 2012; Tukey, 1970) and climatic conditions change. While the month of July corresponds to a peak in *Methylobacterium* diversity (Leducq *et al.*, 2022a), our study is only a temporal snapshot; additional species might have been detected earlier and later in the growth season. Furthermore, additional species might be present in the highest layers of the canopy, where the quantity of light imposes other cellular and metabolic constraints (Carvalho and Castillo, 2018) as well as leaf characteristics (Dörken and Lepetit, 2018).

Unsurprisingly, most detected species belonged to clade A (*n* = 16; Figure 1.10), again corroborating with the findings of Leducq *et al.* (2022a, 2022b) who demonstrated that most phyllosphere-associated *Methylobacterium* species were members of this clade. No species of clade C were detected, confirming that this clade is not associated with the temperate forest phyllosphere. Most communities were dominated by either *M. sp. 018* or *M. sp. 021*, two species from clade A who might potentially be keystone taxa. Furthermore, 13 out of the 23 detected species are species inferred from a phylogenomic study (Leducq *et al.*, 2022b) for which the corresponding reference genomes belonged to strains that had been isolated from leaves of *A. saccharum* or *F. grandifolia* trees in our study plot in 2018 (Leducq *et al.*, 2022a). All ten other detected species' reference strains have also been isolated from plant leaves. Importantly, we did not observe any relationship between host species of origin and host-specific abundance. Likewise, Leducq *et al.* (2022a) had not found an effect of host species on isolates diversity.

By analyzing the whole bacterial community structure at the genus-level, we found that *Methylobacterium* abundances covaried positively with three other genera: *Sphingomonas*, *Hymenobacter*, and *Spirosoma* (Figure 1.4). *Sphingomonas* (phylum *Pseudomonadota*, class *Alphaproteobacteria*) accounted for 5.6% (± 2.6%) of phyllosphere bacterial communities across all our samples. *Sphingomonas* is a major and ubiquitous genus of the phyllosphere (Delmotte *et al.*, 2009; Sohrabi *et al.*, 2023), including in the temperate forest (Laforest-Lapointe *et al.*, 2016a). It has notably been shown to protect against the leaf pathogen *Pseudomonas syringae* (Innerebner *et al.*, 2011). *Hymenobacter* (*Bacteroidetes*, *Cytophagia*) accounted for 2.8% (± 2.6%) of phyllosphere bacterial communities across all our samples. Even though *Hymenobacter*'s ecological roles in the phyllosphere are not abundantly documented (Bashir *et al.*, 2022), it can be one of the most abundant genera in the phyllosphere, often in association with *Spingomonas* and sometimes *Methylobacterium* (Aydogan *et al.*, 2020; Janakiev *et al.*, 2020; Leveau and Tech, 2011). An experimental study by Aydogan *et al.* (2020) showed a decreased in abundances of

Hymenobacter, *Sphingomonas*, and particular *Methylobacterium* taxa under a 2°C warming treatment (associated with an increase in *Pseudomonas*); while other studies have suggested an implication of this particular genera combination in mitigating atmospheric pollution (Franzetti *et al.*, 2020; Kończak *et al.*, 2024; Smets *et al.*, 2016). The potential role of this three genera combination in plant health and growth in temperate forests would be worth investigating further. We also found that *Methylobacterium* abundances covaried negatively with *Mycobacterium*, *Xanthomonas*, *Granulicella*, and *Bacillus* (Figure 1.4). *Mycobacterium* (*Actinomycetota*, *Actinomycetes*) can be highly abundant on plant leaves (Vorholt, 2012) where some species could play a key role in plant health (de Matos *et al.*, 2024). *Xanthomonas* (*Proteobacteria*, *Gammaproteobacteria*) was the second most prevalent genus in our phyllosphere samples (8.4%). This well-studied genus is known for its diversity of pathogenic species that affect hundreds of plant species often in a host-specific manner including important crops such as rice and banana (Jacques *et al.*, 2016; Ryan *et al.*, 2011; Timilsina *et al.*, 2020), as well as trees and shrubs (Goychuk *et al.*, 2023; La Porta *et al.*, 2023), including *Corylus* species (Kałużna *et al.*, 2021; Webber *et al.*, 2020). It has been demonstrated that inoculating some *Methylobacterium* strains on rice and tomato leaves could greatly reduce leaf blight symptoms (Oeum *et al.*, 2024; Yim *et al.*, 2014), but more work is needed to better understand the relationship between those genera, especially in natural forest ecosystems.

1.6.2 Host tree and shrub species are drivers of bacterial community structure at the level of genera, species, and genes

Host species influenced the taxonomic and genetic components of leaf bacterial communities in the temperate forest, in line with numerous studies of the tree phyllosphere (Laforest-Lapointe *et al.*, 2016a; Lajoie and Kembel, 2021; Leducq *et al.*, 2022a; Redford *et al.*, 2010; Wang *et al.*, 2023a). Host species explained almost half of the variation in community composition at the genus level (47%), and around a third of the variation at the *Methylobacterium* species level (36%) and gene cluster level (30%) (Figures Figure **1.5**, Figure **1.9** and Figure **1.11**). Such a strong effect of host species within assemblages of species from a same genus suggests that, while genera compositional variation can provide a general portrait of community dynamics, a higher taxonomic resolution is needed to adequately understand ecological mechanisms. Interestingly, we found that no plant species, except *Corylus cornuta*, harboured host-exclusive *Methylobacterium* species. These results indicate that most *Methylobacterium* species are generalists that are well adapted to a wide variety of host species – in line with the conclusion by Knief *et al.* (2010) –, and that only a few ubiquitous dominant taxa (in our study, *M. sp. 018* and *M. sp. 021*) shape

community structure by variation in abundance. Furthermore, the fact that host-driven compositional variations in *Methylobacterium* species and gene clusters were of similar importance suggests that there is not a marked disconnection between the taxa and their respective gene content in the phyllosphere, unlike in the gene ecology model (Shapiro *et al.*, 2016). In other words, even if some genetic variation probably arises due to gene gain, duplication and loss (which we did not measure), species appear to be coherent units with distinct sets of genes. Nevertheless, different host species clearly filtered for different genes: over 10% of gene clusters – most of them belonging to the accessory pangenome – were found to be differentially abundant among host types and species. In addition, the representation of functional categories related to those gene clusters also differed among hosts (Figure 1.13). One of the only aspects of community structure on which host species did not show any influence was the phylogenetic patterns of assembly (Figure 1.10). Globally, *Methylobacterium* species communities were phylogenetically structured; they were composed of evolutionary closely related species, as it is often the case in phyllosphere bacteria communities (Kembel *et al.*, 2014; Lajoie and Kembel, 2021; Leducq *et al.*, 2022a).

Two recurrent patterns of variation were pervasive in all our analyses. The strongest and most consistent pattern of divergence in community structure was observed between the trees and the shrub species, at every level of biological resolution. Notably, *Methylobacterium* was the most abundant bacterial genera living on *C. cornuta* leaves (relative abundance of 13.4%), where it occupied a significantly larger proportion of bacterial communities than on trees (relative abundance of 5.1%) (Figure 1.2). *C. cornuta* also harboured more *Methylobacterium* species than other hosts species (Figure 1.7). *C. cornuta* was unique in harbouring host-specific species; those nine species, including two species from clade B, represented 40% of the total *Methylobacterium* species richness across all samples (Figure 1.10). Clades A and D are the two clades most often associated with the phyllosphere, but clade B contains many species whose representative strain was isolated from the soil (Leducq *et al.*, 2022b). The reference genome we used for one of the two detected clade B species, *M. haplocladii*, have been sequenced from a strain isolated on bryophytes leaves (Tani and Sahin, 2013) – plants living in close contact with the soil. Therefore, one of the factors that could possibly explain the distinctiveness of *C. cornuta* leaf communities is the proximity of the shrub leaves to the forest soil, compared to tree leaves. Observational as well as experimental studies have effectively shown that soil microorganisms have the ability to colonize plant leaves and can potentially account for a large fraction of the phyllosphere communities (Massoni *et al.*, 2021; Zarraonaindia *et al.*, 2015). Alternatively, colonizing

populations could migrate from tree leaves to shrub leaves through rainfall (Bittar *et al.*, 2018). A vertical gradient in phyllosphere community structure has previously been shown (Herrmann *et al.*, 2021; Stone and Jackson, 2019), but the exact drivers are still not fully understood. Although we collected *C. cornuta* leaves at a similar height for each sample, future studies of shrub microbiomes should consider leaf distance from the soil as an important variable, and if possible sample the microorganisms present in the soil and in rainfall.

C. cornuta was also the only host species on which we observed a clear dominance of a single *Methylobacterium* species, *M. sp. 018*, thus leading to an uneven community structure compared to tree species (Figure 1.8). Interestingly, this pattern of higher richness and lower evenness in the lower strata of the canopy has also been observed by Stone and Jackson (2019). Within *M. sp. 018* population, the number of host-associated positively selected genes was much higher on *C. cornuta* than on other host species (Table 1.3). This host-driven ongoing positive selection could also potentially reflect a higher diversity of novel biotic interactions due to a sustained immigration, leading to an increase in the strength of selection, in line with the Red Queen Hypothesis (Brockhurst *et al.*, 2014), and favouring beneficial genetic variation. Indeed, many positively selected genes on *C. cornuta* were toxins and antitoxins and membrane-associated proteins.

In spite of all these differences between *C. cornuta* and the tree hosts, we found that, in terms of gene cluster variation, the angiosperm leaves conditions exerted a stronger filter than did the shrub leaf, resulting in 4.5X more genes associated with all broadleaves (including *C. cornuta*) compared to genes associated with *C. cornuta* only (Figure 1.12) – thus spanning a higher diversity of functional categories (Figure 1.13). This is also reflected by a lower dissimilarity in *Methylobacterium* genetic composition between *C. cornuta* and broadleaf trees than between *C. cornuta* and conifers (Figure 1.11). This strong broadleaf ecological filtering effect, which acts stronger than plant form on gene composition, is in line with the findings that bacterial community functional structure is strongly influenced by host phylogeny, especially at high taxonomic levels (Lajoie and Kembel, 2021).

Pronounced differences in community structure between broadleaves and conifers effectively accounted for another important observed pattern of divergence in the temperate phyllosphere, in accordance with other studies (Laforest-Lapointe *et al.*, 2016a; Lajoie and Kembel, 2021). Putting aside the unique *C. cornuta* host species, we revealed significant differences in community

composition between broadleaf and conifer tree host species at all levels of analysis. Broadleaf trees and conifers harboured distinct assemblages of bacterial genera (Figure 1.5), although both showed similar relative abundance of *Methylobacterium* (Figure 1.2). Those bacterial assemblages were more homogenous on conifer hosts. This might be a direct effect of more similar leaf phytochemistry on *A. balsamea* and *T. occidentalis* than on *A. saccharum* and *F. grandifolia*. Alternatively, the homogeneity in bacterial communities could reflect leaf phenology. Conifer leaves persist longer than deciduous leaves, and at the level of an individual tree the conifer phyllosphere habitat persist as long as the tree is alive. Since the habitat is more stable, bacterial communities could have more time to evolve into a stable state in which stochastic assembly processes lose importance over deterministic processes (Dini-Andreote *et al.*, 2015) which could lead to homogenous selection. In contrast, deciduous leaf communities could be more strongly shaped by massive yearly stochastic colonization events associated with leaf emergence. In support of this leaf phenology hypothesis, it has been demonstrated that *Methylobacterium* taxonomic communities progressively became more homogenous within a forest as the growth season unfolded (Leducq *et al.*, 2022a).

At the levels of *Methylobacterium* species, broadleaf and conifer trees also harboured different communities (Figure 1.9). We found one broadleaf specific (including *C. cornuta*) species, *M. sp. 027,* who was found on all broadleaf hosts, but not on conifers. In terms of presence, *M. sp. 025* showed a strong association with conifers, although its relative abundance was low within each individual community; it was present in 90% of conifers samples, but only in 20% of broadleaf samples. *M. sp. 018* and *M. sp. 021* were the most prevalent species respectively on broadleaves and *T. occidentalis*. Contrasting with the pattern observed at the level of bacterial genera, both conifer and broadleaf trees were equally heterogenous in terms of species compositional variation. In addition, *A. balsamea* was the only host species within which species turnover exerted a higher impact than abundance gradients (i.e., nestedness) on total compositional variation (Table 1.2) – a sign of taxonomic heterogeneity which could be attributable to the fact that no particular species dominate on *A. balsamea* compared to other host species. The decreased influence of potential keystone taxa (*M. sp. 018* and *M. sp. 021*) could result in the remodeling of interspecies dynamics, and thus community structure (Banerjee *et al.*, 2018). Our contrasting results at the level of bacterial genera and *Methylobacterium* species highlights the need to better understand how phyllosphere temporal dynamics differ between conifer and broadleaf species, and what are the impacts of these distinctions at different taxonomic resolutions, from phylum to species. Furthermore, conifers and broadleaf trees harboured a similar

Methylobacterium species richness (Figure 1.7), contrasting with the findings of Laforest-Lapointe *et al.* (2016a) who found that conifer leaf communities had a higher alpha diversity. Differences in gene abundances were uncommon among tree host pairs of the same clade, but as mentioned earlier they were plentiful between all broadleaves (including *C. cornuta*) and conifers. We found 1.7X more broadleaf-associated gene clusters than conifer-associated clusters (Figure 1.12). Furthermore, within *M. sp. 018* population, positive selection on genes was substantially more frequent on broadleaves than on conifers (Figure 1.17), again possibly due to the fact that the conifer phyllosphere is a more stable environment than the leaves of deciduous species. The stability of conifer leaf surfaces could call for genetic conservatism (i.e., purifying selection). The ephemeral nature of broadleaf species, leading to constantly reshuffled biotic interactions both with co-occurring microorganisms and hosts, could call for genetic innovation by cells of colonizing populations, again in line with the Red Queen hypothesis (Brockhurst *et al.*, 2014) and with the observation that in pathogens, positive selection is often more prevalent in genes involved in host and microbiota interactions (e.g., cell surface proteins) (Petersen *et al.*, 2007).

1.6.3 *Methylobacterium* adaptations to host types are reflected in their genes

Since no gene clusters were found to be differentially abundant between *A. saccharum* and *F. grandifolia*, nor between *A. balsamea* and *T. occidentalis*, we could not identify host-specific potentially adaptive genes for individual tree species. This absence of distinguishable gene clusters between those host species is most possibly attributable to a lack of statistical power due to the low number of samples from each species (n = 5). We thus focused our analyses mostly on gene associations with host types, principally between trees and *C. cornuta*, and between all broadleaf species and conifers. Instead of analyzing the completeness of metabolic pathways, we decided to interpret in their ecological context the differentially abundant genes (i.e., genes enriched in particular host types) and their functional categories, as well as genes evolving under positive selection in a host-associated manner.

Mechanisms leading to genes enrichment in a given ecological niche compared to others include novel gene gain through HGT, gene duplication or gene loss (Andersson and Hughes, 2009; Ochman *et al.*, 2000). Those genes can be more abundant because they confer an adaptive advantage to a population; however, those genes can also be neutral or even slightly deleterious and can be hitchhiking along other selected portions of the genome in a population (Ares-Arroyo *et al.*, 2024) without offering niche-associated benefits. Therefore, we were careful in our interpretation of differentially abundant genes, since we are aware that some of our result-driven

emergent hypotheses would need to be experimentally tested. In addition, partly because phyllosphere microorganisms are poorly represented in functional databases, a third of all differentially abundant gene clusters were not annotated, thus representing hypothetical proteins for which we cannot predict the function. Investigation of those bacterial hypothetical proteins – either by comparing nucleotide sequences with well-annotated bacterial genomes or by analyzing the predicted protein structure – could potentially reveal important novel niche adaptive genes, functions and pathways (Ijaq *et al.*, 2022).

We observed that *Methylobacterium* communities on tree hosts were broadly enriched in functions related to amino acids, translation and energy (Figure 1.13), hinting at the importance of energy-related protein production and turnover. Importantly, we found enriched genes involved in photosynthesis, oxidative phosphorylation, and protection against heat and oxidative stresses. Tree leaves are higher in the canopy than shrub leaves, and thus in a closed forest receive more sunlight, including ultraviolet radiation which can engender oxidative stress through the production of reactive oxygen species (Santos *et al.*, 2012, 2013). Photohetereotrophic bacterial populations capable of harvesting that light more efficiently while being better able to protect themselves from heat and UV damages would be favoured. We thus hypothesize that sunlight exposition is one of the main factors responsible for the divergence of tree and shrub-inhabiting bacterial communities in a forested ecosystem, in line with other studies demonstrating that sunlight, and particularly UV radiation, exert a strong selective pressure in the assembly of phyllosphere communities (Carvalho and Castillo, 2018; Jacobs and Sundin, 2001; Truchado *et al.*, 2017).

In contrast, *Methylobacterium* communities living on *C. cornuta* were mostly enriched in genes related to the mobilome and to inorganic ions, and to a lesser extent to transcription and cell membrane (Figure 1.13). The mobilome of a bacteria can be defined as the entirety of its mobile genetic elements – such as plasmids, transposons, and prophages, acquired through HGT – and play a crucial role in biotic interactions (virus-bacteria, bacteria-bacteria and bacteria-host), antimicrobial resistance, virulence, as well as in niche-driven community assembly and evolution, including in the phyllosphere habitat (Carr *et al.*, 2021; Holtappels *et al.*, 2024). The divergence from trees in terms of enriched functional categories highlights the possibly elevated importance of environmental interactions on the shrub species compared to trees, potentially reflecting a higher immigration rate from soil or through rainfall.

Conifer *Methylobacterium* communities were mostly enriched in genes related to amino acid, carbohydrate and lipid metabolisms (Figure 1.13). We notably found more abundant genes involved in methylamine oxidation, as well as genes involved in long chain fatty acid and glycerophospholipid metabolisms – these latter could play a role in population survival to large yearly temperature variation by modulating the cell membrane composition and flexibility. Broadleaf *Methylobacterium* communities were notably enriched in gene clusters involved in cell wall and membrane biogenesis, defence mechanisms, transcription, and signal transduction (Figure 1.13), suggesting higher levels of novel biotic interactions due to more frequent stochastic colonization events. These divergences in enriched functional categories between conifers and broadleaves strengthened our previously mentioned leaf phenology hypothesis.

Although we found a few gene-host associations under positive selection, purifying selection was by far the most dominant type of selection acting on genes in the three *Methylobacterium* population studied living on every host species (Figure 1.16), in line with numerous studies of bacterial populations along variation in environmental conditions within their natural habitats (Dong *et al.*, 2023; Kiefl *et al.*, 2023; Li *et al.*, 2024a; Schloissnig *et al.*, 2013; Shenhav and Zeevi, 2020). Furthermore, the fact that we found a much higher abundance of positively selected genes across all samples than the number of detected gene-host associations under positive selection suggests that there is most certainly other factors than host species driving natural selection in the temperate forest. Finally, many host-specific key *Methylobacterium* genes could be under extremely strong purifying selection. Our selection results are thus only a limited, albeit informative, portrait of how *Methylobacterium* genes evolved in host-associated populations.

1.6.3.1 Tree-enriched gene clusters

The core gene cluster encoding the Bacteriochlorophyllide reductase subunit BchX was one of the energy-related tree-enriched clusters (although our analyses indicated a statistical association with conifer trees). This oxidoreductase is involved in the reactions leading to the synthesis of bacteriochlorophyll *a* (Chew and Bryant, 2007) which is used by aerobic anoxygenic photoheterotrophs (Yurkov and Hughes, 2017) such as *Methylobacterium*. In addition, three gene clusters involved in proton transportation during ATP synthesis via oxidative phosphorylation were also enriched in tree communities: the single copy core gene clusters encoding the alpha and beta subunits of the FoF1-type ATP synthase (AtpA and AtpD), and the heme/copper-type cytochrome/quinol oxidase subunit 1 (CyoB). These enrichments could be related to an elevated photosynthetic activity or to the higher demand in energy to face oxidative stress (Kumar *et al.*,

2016). Interestingly, the differential abundance of single copy core genes could only plausibly result from a gene duplication event. Furthermore, many gene clusters encoding stress response proteins were also enriched, such as RecA, a recombinase involved in DNA reparation following UV-B exposition (Gunasekera and Sundin, 2006), as well as GroEL (HSP60) and GroES (HSP10), respectively a chaperonine and a co-chaperonine involved in protein folding and assembly which play an important role in bacterial tolerance to high temperature (Fayet *et al.*, 1988; Maleki *et al.*, 2016; Segal and Ron, 1998), as well as to other kinds of abiotic stresses such as ultraviolet radiation (Krueger and Walker, 1984). Finally, an accessory gene cluster encoding for the enzyme Glucose-6-phosphate 1-dehydrogenase (Zwf) was enriched in tree communities. Zwf is involved in the pentose phosphate pathway, an alternative pathway to glycolysis for glucose oxidation. Studies have shown that redirecting the carbohydrate flux from glycolysis to the pentose phosphate pathway was a way to respond to oxidative stress by increasing the reduction of $NADP^+$ to NADPH (Christodoulou *et al.*, 2018; Ralser *et al.*, 2007; Rui *et al.*, 2010). Interestingly, the genomes which recruited the most reads for this accessory gene cluster in tree samples are not the ones representing species that were detected in our study plot, thus strongly suggesting HGT.

1.6.3.2 *Shrub-enriched gene clusters*

Mobilome genes were abundant in *Methylobacterium* communities inhabiting the leaf surface of every host species, but they made up the largest portion of *Corylus cornuta* host-associated gene clusters. Although in our study we could not evaluate the direct functional impact of those genes, they could potentially play an important role in mediating elevated microbe-microbe interactions on *C. cornuta*, in line with the possibility that shrub communities are subjected to a sustained immigration. Outside the mobilome, shrub-enriched genes of diverse functional categories were identified as potentially playing important ecological roles regarding biotic interactions. The accessory gene cluster encoding the Acyl-coenzyme A synthetase/AMP-(fatty) acid ligase (Acs) (or long-chain acyl-CoA synthetase) was enriched in *C. cornuta* communities. Long-chain fatty acids are involved in many biological processes, among which cell membrane biosynthesis, signalling in cell processes regulation, virulence, quorum sensing and antibiosis (Kumar *et al.*, 2020; Mitchell and Ellermann, 2022). Long-chain fatty acids have also been shown to be involved in auxin transport and development of *Arabidopsis thaliana* (Roudier *et al.*, 2010). Long-chain fatty acids could thus potentially play a major role in plant growth and phyllosphere host-bacteria mutualistic interactions. In addition, in the *C. cornuta*-inhabiting *M. sp. 018* population, we detected positive selection on a gene encoding a NAD(P)-dependent dehydrogenase FabG (short-chain alcohol dehydrogenase family). FabG is another enzyme involved in fatty acid biosynthesis;

among other functions, it can participate in the synthesis of long-chain acyl-CoA along the enzyme Acs. Among enriched genes related to cell membrane and transport regulation, we found a cluster coding for the Opacity protein LomR – a homolog to OmpX (Sun *et al.*, 2021), a membrane immunogenic protein. These beta-barrel proteins can serve as porins and transporters, and can play a role in signal transduction, pathogenesis, as well as in membrane biogenesis (Fairman *et al.*, 2011; Kolodziejek *et al.*, 2007; Wimley, 2003). Finally, a gene encoding for Flp/PilA, a pilus assembly protein, was found to be positively selected in *M. sp. 018* population inhabiting *C. cornuta* (the associated gene cluster is also enriched in broadleaves). Flp is involved in biofilm formation and surface adhesion which favour host colonization (Kachlany *et al.*, 2000; Wang and Chen, 2005), as well as in motility, phage transduction and interactions with eukaryotes cells, and pathogenicity (Craig *et al.*, 2004).

1.6.3.3 Conifer-enriched gene clusters

Lipid transport and metabolism was one of the most well represented functional categories among genes associated with conifer-inhabiting *Methylobacterium* communities. An accessory gene cluster coding for a phosphatidylethanolamine (phosphatidyl-N-methylethanolamine N-methyltransferase) was enriched in *Methylobacterium* conifer communities. This enzyme is involved in the glycerophospholipid metabolism, where it catabolizes phosphatidylethanolamine to phosphatidylcholine. Chwastek *et al.* (2020) demonstrated the adaptive potential of headgroup changes in acyl chains, notably to temperature variation. Two of the acyl chains that showed potential for adaptation through variation are phosphatidylethanolamine and phosphatidylcholine – the latter have notably been suggested to protect bacterial cells of *Pseudomonas aeriginosa* to survive freezing (Wilderman *et al.*, 2002) and to play an important role in the interactions between bacteria and their host plants (Aktas *et al.*, 2010; Minder *et al.*, 2001). A potentially higher production of phosphatidylcholine could favour populations living on conifer leaves by providing an enhanced protection against freezing temperatures in winter. Interestingly, the enriched gene cluster is not present in the genome representing *M sp. 018*; its absence might partially explain the lower relative abundance of this species on conifers. Furthermore, accessory gene clusters coding for two subunits (B and C) of the *N*-methylglutamate synthase were enriched in conifers' communities. The *N*-methylglutamate pathway is the principal way for *Methylobacterium* to consume methylamine (Alessa *et al.*, 2021; Minami *et al.*, 2016), a one-carbon compound. The alternative enzyme to oxidize methylamine is the methylamine dehydrogenase; strains from only five *Methylobacterium* species encode its subunits. In our study, none of the genes encoding the methylamine dehydrogenase recruited any read, whereas *N*-methylglutamate synthase genes

mapped abundantly to *M. sp. 021* – the most abundant species on *T. occidentalis* – as well as to three other poorly covered species of clade A, suggesting a past acquisition of the genes by *M. sp. 021* through HGT. Our results demonstrate that in the temperate phyllosphere, methylamine is exclusively oxidized through the *N*-methylglutamate pathway by species who possess the needed accessory genes. Interestingly, Shiraishi *et al.* (2015) demonstrated that as the leaves of *A. thaliana* aged the methylamine secretion increased. This phenomenon could potentially explain why *N*-methylglutamate synthase genes are enriched in communities inhabiting conifers which have older leaves than deciduous trees.

1.6.3.4 Broadleaf-enriched gene clusters

On broadleaves, many defence-related genes coding for toxin/antitoxin components were enriched, notably the antitoxin vapB of the vapBC operon, which alleviate the growth inhibition effect of vapC (Robson *et al.*, 2009); the antitoxin PrlF or the YhaV-PrlF toxin-antitoxin module that counteracts the RNAse activity of YhaV (Schmidt *et al.*, 2007); the MazE antitoxin of the MazEF toxin-antitoxin module, which counteracts MazF cell death pathway that can be used to prevent spread of phage infection in population (Engelberg-Kulka *et al.*, 2005); and the HicA toxin of the HicAB toxin-antitoxin module, which prevents cell growth through mRNA cleavage (Jørgensen *et al.*, 2009). Furthermore, the gene encoding for the multidrug resistance efflux pump EmrA, involved in xenobiotic (antibiotics, metals, detergents, etc.) detoxification by transmembrane exportation (Hinchliffe *et al.*, 2014), was enriched in broadleaf communities. Toxin/antitoxin components were also relatively abundant in positively selected genes, notably a *M. sp. 018* gene encoding the VapC toxin which showed signs of positive selection on *C. cornuta* but also on the conifer *T. occidentalis*. In addition, we identified a *M. sp. 018*-*C. cornuta* positive selection on the mRNA-degrading endonuclease RelE, the toxin component of the RelBE toxin-antitoxin system (Pedersen *et al.*, 2003); on BrnT, the ribonuclease toxin of BrnTA toxin-antitoxin module which is regulated in response to environmental stress (Heaton *et al.*, 2012); on YhaV toxin; and on a gene coding for the antitoxin phd which prevents cell death by the toxin doc (Smith and Magnuson, 2004). These results suggest that in broadleaves and in shrub *Methylobacterium* populations, an enhanced control of cell growth and activity in response to the environment is crucial. In addition, on broadleaf communities, numerous transcription-related gene clusters were enriched, notably two coding for a DNA-binding transcriptional regulator from the HxlR family, which in presence of formaldehyde positively induce the expression of the hxlAB operon involved in formaldehyde oxidation (Yurimoto *et al.*, 2005). This transcriptional regulator thus plays a key role in *Methylobacterium* methanol consumption. Finally, three accessory gene cluster encoding

CheB, a chemotaxis response regulator with a methylesterase activity (Stewart *et al.*, 1990), were enriched on broadleaves, suggesting that enhanced motility capacity might possibly be more important on deciduous leaves because of their larger surface area, a hypothesis shared by Lajoie and Kembel (2021).

1.6.4 The accessory pangenome of *Methylobacterium* is partly adaptive

Although the focus of our study was principally on revealing associations between *Methylobacterium* genes and host species – rather than investigating the evolutionary mechanisms shaping the structure of *the Methylobacterium* pangenome – we found substantial indications supporting our hypothesis stating that the accessory pangenome is, at least partially, adaptive.

Indeed, while identifying host-enriched gene clusters, we found clear evidence for the role of accessory genes in conferring host adaptations. Firstly, many differentially abundant accessory gene clusters recruited reads by genomes whose species were poorly covered, suggesting adaptive HGT between *Methylobacterium* populations. Secondly, some host-enriched accessory gene clusters recruited reads by every genome whose species were detected in the host's communities, suggesting that populations are maintaining these accessory genes in their genomes. Thirdly, many proteins coded by enriched accessory gene clusters were also coded by numerous other accessory and core clusters, suggesting complex transfers, duplication and loss events leading to adaptations. Finally, and most importantly, we were able to bridge the link between host-specific *Methylobacterium* ecology and the functions of proteins encoded by many of its enriched accessory gene clusters – for example, oxidative stress response in tree-inhabiting communities, glycerophospholipid metabolism on conifers, defence mechanism on broadleaves, etc. In addition, in the three most prevalent *Methylobacterium* populations of the forest phyllosphere, *M. sp. 018*, *M. sp. 021*, and *M. sp. 022*, we found a higher proportion of positively selected genes in the accessory pangenome (Figure 1.18). The accessory genes undergoing positive selection in a host-associated manner were mostly coding for proteins involved in functions related to environmental interactions, notably many components of toxin/antitoxin modules, thus supporting their adaptive nature. Furthermore, numerous transposases (all accessory genes) were differentially abundant among host types or undergoing positive selection, although we were not able to assess if their abundance or evolution is adaptive for the bacterial host cell of for the mobile genetic elements themselves (Brockhurst *et al.*, 2019).

One of our objectives was to evaluate the contribution of *Methylobacterium* accessory pangenome to host adaptations, and not to resolve the long-standing debate on the nature of pangenome evolution. However, our conclusion that the accessory pangenome is in part adaptive is in accordance with the theoretical reasoning of McInerney *et al.* (2017) which stated that large populations have large and open pangenomes because they occupy many different niches whose selective pressure, albeit sometimes small, can maintain the accessory gene diversity. Our results are in line with this rationale. The *Methylobacterium* pangenome we assembled was open (Figure A.5). We demonstrated that *Methylobacterium* is a phyllosphere generalist genus containing some populations of very large sizes showing slight variation in host types preference (as well as some ubiquitous small size populations), and that some accessory genes are more abundant or evolve under positive selection in host-associated communities or populations. In addition, the adaptive nature of a substantial part of *Methylobacterium* accessory genes has also been suggested in a laboratory evolution experiment (Lee and Marx, 2012).

1.7 Conclusion

Our study revealed a strong influence of host species on bacterial community structure, as well as on *Methylobacterium* species assemblages and genetic variation, including within the accessory pangenome. Our data suggest that host types (i.e., tree, shrub, broadleaves, conifers), to a greater extent than host species *per se*, drove the main patterns of divergence in the temperate forest phyllosphere by exerting contrasting selective pressures. In light of our results, we hypothesize that in closed canopy forests sunlight play a key role in differentiating bacterial community structure between trees and shrub species, as reflected in the tree-enriched genes related to photosynthesis, oxidative phosphorylation, and stress response. On the other hand, shrub-inhabiting *Methylobacterium* communities could potentially be more strongly influenced by a higher diversity of biotic interactions resulting from a sustained migration of microorganisms due to soil proximity or rainfall. This could be reflected by the higher number of *Methylobacterium* species, by their enriched genes involved in biotic interactions, as well as by the relatively high amount of positively selected genes related to defence. In addition, we hypothesize that phyllosphere bacterial communities are also extensively shaped by leaf phenology. Since most conifers do not shed their leaves every fall, they offer a persistent habitat, while deciduous species leaves regularly offer a novel habitat with renewed colonizing opportunities. According to our leaf phenology hypothesis, conifer-inhabiting *Methylobacterium* communities are more strongly influenced by climatic fluctuations and leaf age, as reflected in the homogeneity of their residing bacterial genera, as well as in their enriched genes involved in fatty acid and glycerophospholipid

metabolisms; whereas broadleaf-inhabiting communities are more strongly influenced by a higher diversity of biotic interactions resulting from massive yearly stochastic colonization, as reflected in their higher proportion of positively selected genes in the largest population, as well as in their enriched genes involved in defence, transcription regulation, and environmental interactions. While many studies have focused on finding associations between leaf and plant traits and the phyllosphere community structure to explain broad patterns of taxonomic, phylogenetic and functional divergence (Kembel *et al.*, 2014; Lajoie *et al.*, 2020; Lajoie and Kembel, 2021), we advocate that ecological processes-driven hypotheses – such as our proposed leaf phenology hypothesis – should be used as experimental starting points beyond hosts or traits *per se* to further increase our understanding of the mechanical underpinnings of phyllosphere niche adaptations. The temperate forest is a wide-ranging terrestrial biome that has unique ecosystem dynamics due to its strong seasonality. Our study highlighted the key role played by its floristic biodiversity in shaping microbial ecology and evolution, and proposed clear hypotheses to guide future research on the phyllosphere microbiome.

1.8 Acknowledgements

We would like to thank Zihui Wang and Élanore Favron for their help in collecting and processing samples; Gabriel Lanthier for granting us access to the Station de Biologie des Laurentides; Geneviève Bourret, research professional at the Center of Excellence in Research on Orphan Diseases – Fondation Courtois (CERMO-FC) for preparing the libraries; the CHU de Québec-Université Laval Research Center for DNA sequencing; and David Ross for his support in running some metagenomic analyses.

CHAPITRE 2

RICHNESS AND COMPOSITION OF PHYLLOSPHERE *METHYLOBACTERIUM* COMMUNITIES CAUSE VARIATION IN *ARABIDOPSIS THALIANA* GROWTH

Article submitted to *Oikos*

Jocelyn Lauzon[1], Jérémie Pelletier[1], Élanore Favron[1], Zihui Wang[2] & Steven W. Kembel[1]

[1] *Département des sciences biologiques, Université du Québec à Montréal*
[2] *Department of Botany, University of British Columbia*

2.1 Résumé

La phyllosphère – la partie aérienne des plantes – forme un vaste habitat microbien. Les feuilles abritent des communautés bactériennes diversifiées qui jouent des rôles clés dans le fonctionnement des écosystèmes. Conséquemment, la surface foliaire est un système d'étude démontrant un potentiel élevé pour investiguer la relation entre la biodiversité et le fonctionnement des écosystèmes. Des études ont révélé une corrélation positive entre la diversité bactérienne des feuilles et la productivité des écosystèmes, mais le lien de causalité n'a toujours pas été démontré. Pour comprendre comment la diversité et la composition des communautés bactériennes de la phyllosphère pourraient causer une variation dans la croissance de leurs plantes hôtes, nous avons assemblé des communautés synthétiques en utilisant des souches de *Methylobacterium* – un genre bactérien omniprésent dans la phyllosphère, qui y stimule la croissance de ses hôtes. Nous avons inoculé ces communautés synthétiques sur des plants de *Arabidopsis thaliana* cultivés en conditions gnotobiotiques. Nos hypothèses étaient que (1) l'augmentation de la diversité de *Methylobacterium* causerait une augmentation de la croissance de l'hôte; (2) les souches auraient des impacts différents sur la croissance de l'hôte; et (3) la relation entre la diversité bactérienne et la productivité des plantes changerait en fonction des souches présentes. Nos résultats ont soutenu nos hypothèses, mais ont révélé des patrons inattendus dans la manière dont la biomasse foliaire de *A. thaliana* variait avec la richesse et l'identité des souches de *Methylobacterium*. L'augmentation de la richesse bactérienne entraînait une plus grande biomasse foliaire, mais seulement après une réduction initiale de la biomasse, suggérant un effet d'allègement de la compétition apporté par les interactions entre de multiples espèces. Deux souches de *Methylobacterium* montraient des effets bénéfiques sur la croissance de *A. thaliana*, mais une souche s'est avérée néfaste pour la plante. La composition des communautés façonnait la relation entre diversité et productivité, soulignant l'importance des interactions mutualistes et antagonistes entre souches bactériennes. De plus, nous avons démontré que la complémentarité des niches était probablement le principal mécanisme écologique influençant la relation diversité-productivité dans notre système d'étude. En révélant la relation de cause à effet entre la diversité et la composition des communautés de *Methylobacterium* et la croissance de leurs plantes hôtes, notre expérience a mis en lumière l'importance des bactéries de la phyllosphère dans le fonctionnement des écosystèmes terrestres.

Mots clés : *Arabidopsis thaliana*, biodiversité-fonctionnement des écosystèmes, écologie des communautés, *Methylobacterium*, complémentarité des niches, phyllosphère, croissance des plantes, communautés synthétiques.

2.2 Abstract

The phyllosphere – the aerial parts of plants – forms a vast microbial habitat that harbours diverse bacterial communities playing key roles in ecosystem function. The foliar surface is thus a promising study system to investigate biodiversity-ecosystem function relationships. Researchers have found a positive correlation between leaf bacterial diversity and ecosystem productivity, but the causality of this relationship has yet to be demonstrated. To understand how the diversity and composition of phyllosphere bacterial communities could cause variation in the growth of their host plants, we assembled synthetic communities composed of different diversities and compositions of *Methylobacterium* strains – a plant growth-promoting bacterial genus ubiquitous in the phyllosphere – that we inoculated on *Arabidopsis thaliana* grown in gnotobiotic conditions. We hypothesized that (1) increasing *Methylobacterium* diversity should cause an increase in host growth; (2) strains should differ in their impact on host growth; and (3) the relationship between bacterial diversity and plant productivity should be strain-dependent. Our results supported our three hypotheses but revealed unpredicted patterns in how *A. thaliana* leaf biomass varied according to inoculated *Methylobacterium* strain richness and identity. Increasing bacterial richness induced a higher host leaf biomass, but only after an initial reduction in biomass, suggesting competition alleviation by multispecies interactions. Two *Methylobacterium* strains showed beneficial effects on *A. thaliana* growth, and one strain was detrimental for the plant. Community composition shaped the relationship between diversity and productivity, highlighting the importance of community mutualistic and antagonistic interactions. Furthermore, niche complementarity was likely the main ecological mechanism driving the diversity-productivity relationship in our study system. By demonstrating the causal effects of *Methylobacterium* community diversity and composition on host plant growth, our experiment shed light on the importance of phyllosphere bacteria in terrestrial ecosystem functioning.

Keywords: *Arabidopsis thaliana*, biodiversity-ecosystem functioning, community ecology, *Methylobacterium*, niche complementarity, phyllosphere, plant growth, synthetic communities.

2.3 Introduction

The role of biodiversity as one of the cornerstones of ecosystem dynamics and functioning is well known, whether through its effects on productivity, stability, or resilience (Cardinale *et al.*, 2012; Hooper *et al.*, 2005; Tilman *et al.,* 2014). A notable example of the link between biodiversity and ecosystem function is the positive effect of diversity on ecosystem productivity, as demonstrated by the well-studied positive relationship between plant community diversity and biomass production (Tilman *et al.,* 2001). This biodiversity effect can be explained by two mechanisms, namely complementarity and selection, depending on whether the ecosystem response to diversity is dictated by the complementarity of species niches versus the selection of highly productive species (Loreau and Hector, 2001). While the relationships between biodiversity and ecosystem function have been well studied within communities of organisms from the same trophic level or within food webs (Van Der Plas, 2019), including bacteria (Bell *et al.,* 2005; Jiang, 2007), we understand less about how the biodiversity of microorganisms can influence the functioning of their hosts.

Plant leaves harbour diverse microbial communities, most often dominated by bacteria (Vorholt, 2012). These bacteria exert an influence on plant and terrestrial ecosystems (Bashir *et al*., 2022). While bacteria are important plant pathogens (Baker *et al*., 2010; Xin *et al*., 2018), bacterial taxa can also promote host growth (Pattnaik *et al*., 2017), assist in nutrient acquisition (Batool *et al*., 2016), stimulate host protection against pathogens (De Mandal and Jeon, 2023) and abiotic stresses (Wang *et al*., 2023b), and play a critical role in carbon and nitrogen biogeochemical cycles (Moreira *et al*., 2021; Yurimoto and Sakai, 2023), atmosphere purification (Sandhu *et al*., 2007), and climate regulation (Bringel and Couée, 2015). The high diversity and ecological importance of the plant microbiome thus make the phyllosphere a promising ecosystem to study ecological theory including biodiversity-ecosystem functioning associations (Meyer and Leveau, 2012). A previous study (Laforest-Lapointe *et al*., 2017a) showed that phyllosphere bacterial diversity was positively correlated with plant community productivity, acting as a key explanatory variable for models linking plant diversity with ecosystem productivity, and thus explaining a significant part of the plant biodiversity-productivity relationship. However, this study could not determine the causality of the relationship between microbial biodiversity and host plant productivity, since it was based on an experiment where plant diversity but not microbial diversity was manipulated. The causal relation between phyllosphere bacterial diversity and plant productivity has thus not yet been clearly demonstrated.

The objectives of our study were to test for causal relationships between leaf bacterial diversity and host plant productivity; to reveal the role of bacterial taxa and their community interactions in driving this causation; and to shed light on the ecological mechanisms at play in the diversity-productivity relationship. To address these objectives, we investigated how the diversity and composition of phyllosphere bacterial communities influence the growth of *Arabidopsis thaliana* host plants by assembling and inoculating synthetic communities of diverse strains of *Methylobacterium* in a gnotobiotic experimental setup. Our choice to work with synthetic communities was motivated by the need to move beyond a correlational perspective to understand causal links between the phyllosphere microbiome and host and ecosystem function (Vorholt *et al.*, 2017). *Methylobacterium* is a ubiquitous bacterial genus of plant leaves, known for its methylotrophy and plant growth-promoting characteristic (Abanda-Nkpwatt *et al*., 2006). *Methylobacterium* is an important component of the natural *A. thaliana* microbiome (Knief *et al.*, 2010) and is often employed in synthetic community experiments (Carlström *et al.*, 2019; Innerebner *et al*., 2011).

We tested three hypotheses, using strain richness at the moment of inoculation as a diversity measure, and rosette leaf dry biomass at the end of the vegetative growth period as a measure of plant growth and productivity. Considering the previously observed positive correlation between leaf bacterial diversity and plant community productivity (Laforest-Lapointe *et al*., 2017a), as well as the demonstrated positive effect of inoculating *Methylobacterium* strains on plant growth (Madhaiyan *et al.*, 2004), our first hypothesis was that a greater *Methylobacterium* taxonomic diversity on plant leaves should be associated with increased host plant productivity. We predicted a positive logarithmic correlation between the number of strains contained in inoculated synthetic communities and *A. thaliana* leaf biomass, in line with the redundancy hypothesis (Naeem *et al.*, 1995). Our second hypothesis was that *Methylobacterium* taxa would differentially influence plant growth. *Methylobacterium* strains exhibit genetic, phenotypic, and ecological differences – including dissimilarities in their interactions with other microorganisms (Carlström *et al*., 2019), host preferences (Knief *et al*., 2010; Leducq *et al*., 2022a), substrate utilization, and metabolic pathways (Alessa *et al*., 2021). These differences could result in variation in the growth-promoting effect of different individual strains, as well as in differences in bacterial- and host-dependent community interactions driving ecosystem functions. We thus predicted that the presence of some specific *Methylobacterium* strains would lead to a higher-than-average leaf biomass of *A. thaliana*. Thirdly, considering the potential importance of microbial taxa and their community interactions in determining plant growth (Bell *et al*., 2005), our third hypothesis was that some *Methylobacterium*

strains present in communities would influence the relation between bacterial diversity and plant growth. In other words, the strength of the effect of bacterial diversity on plant growth should be dependent on the strains composing the leaf community. We thus predicted that strains having a larger-than-average positive effect on *A. thaliana* leaf biomass (while testing our second hypothesis) would also, when present in communities, increase the positive effects of strain richness on leaf biomass. Finally, without having an *a priori* hypothesis, we tried to disentangle the relative importance of complementarity and selection in driving the biodiversity effect of *Methylobacterium* strains on *A. thaliana* vegetative growth.

2.4 Material and methods

2.4.1 *Arabidopsis thaliana* growth

Seeds of *Arabidopsis thaliana* Col-0 were surface sterilized and stratified prior to seeding. For sterilization, seeds were soaked 2 min in 70% ethanol, rinsed with sterile water, then soaked 8 min in a 2.5% sodium hypochlorite solution containing 0.2% Triton, and finally rinsed 12 times with sterile water (protocol adapted from Innerebner *et al*. [2011] and Lindsey *et al*. [2017]). Stratification occurred at 4°C for 4 days. Plants were grown inside sterile Magenta boxes with vented lids (1 seed/box), on a sterile minimal salts medium (1 cm thick) supplemented with 1% sucrose, 0.7% plant agar, and 0.2 µm-filtered Gamborg's vitamins solution (pH = 5.6), at 20°C under a 16 hour photoperiod regime (PAR 400-700nm = 120 µmol photons $m^{-2}s^{-1}$).

2.4.2 *Methylobacterium* strains and synthetic community design

Twelve *Methylobacterium* strains, representing diverse species and clades of the genus (Leducq *et al*., 2022b), were chosen to form the different synthetic communities (Appendix B, Table B.1). These strains had been previously isolated from tree leaves as part of a temperate phyllosphere study (Leducq *et al*., 2022a). Our experimental design involved synthetic communities of six richness values (number of strains) for the diversity treatment. We assembled twelve monocultures (richness = 1), twelve different communities composed of a pair of strains (richness = 2), and six different communities for each of the other richness values (4, 6, 8, and 10) for a total of 48 distinct community compositions (Appendix B, Table B.2). All twelve strains were present the same number of times among the different communities of a given richness value.

2.4.3 Bacterial cultures and community assembly

Strains were grown for three weeks at 22°C on an R2A Agar medium supplemented with succinate

3.5 mM as a carbon source. Just before community synthesis and inoculation, bacterial colonies were washed with 2 mL 10mM $MgCl_2$ and gently detached from the medium's surface with a sterile glass loop. For every strain, cells from petri dish were collected in a single sterile 50 mL tube and the suspension was homogenized by pipetting up and down. An extra 20 mL of 10 mM $MgCl_2$ was added, and the suspension was gently vortexed. As a surrogate for bacterial cell concentration, the optical density at 600 nm (OD_{600}) was measured in triplicate on a BioTek Eon Microplate Spectrophotometer plate reader (Richmond Scientific). All twelve liquid strains were then diluted to OD_{600} 0.2 prior to building the communities. Strains were assembled in synthetic communities at equal ratio (1:1 ratio between every strain of a community). Those newly formed communities were finally diluted in 10 mM $MgCl_2$ to an OD_{600} of 0.02 before inoculation (Innerebner *et al*., 2011).

2.4.4 Bacterial inoculation on *A. thaliana*

Synthetic communities were inoculated on *A. thaliana* leaves after 14 days of plant growth post-stratification (Innerebner *et al*., 2011). At the moment of inoculation, the rosettes contained a mean of 5 true leaves longer than 1 mm and had reached around 15% of their final size (growth stage #1.05; Boyes *et al*., 2001). We inoculated synthetic communities on plants by distributing a total of 40 µL of bacterial suspension evenly over the entirety of leaf surfaces. Every plant was inoculated with the same quantity of bacterial cells, approximately 4×10^5 cells. Each of the 48 unique community compositions was inoculated on three plants, and we dropped sterile 10mM $MgCl_2$ on six control plants – for a total of 150 plant samples.

2.4.5 Harvest and dry leaf biomass measurement

Plants were harvested after 38 days of growth, to ensure the rosettes had reached their final size. Plants were cut into three parts: roots, rosette (leaves) and stem. Roots and stems were discarded. We dried rosettes in a Heratherm IMH750-S incubator (Thermo Scientific) at 70°C for 48 hours. The dry biomass was weighted immediately after drying.

2.4.6 Statistical analyses

All statistical analyses were performed using R (v4.2.2; R Core Team, 2021). Data was manipulated with the help of packages tidyr (v1.3.0; Wickham *et al*., 2023a) and dplyr (v1.1.2; Wickham *et al*., 2023b). Figures were made using the base R graphics package and the ggplot2 package (v3.4.3; Wickham, 2016). Model selection, as well as model exploration and averaging, were performed with the MuMIn package (v1.47.5; Bartoń, 2023). Samples whose box contained

fungal contaminants on the agar at the moment of harvesting (n = 5) were discarded, as well as 2 mislabelled samples. We did not consider control plants (n = 6) as samples having a "strain richness" of zero in our models, but we plotted their data in figures as a visual reference for the interpretation of the results. We used the remaining 137 samples for all statistical analyses. All models used dry leaf biomass as a continuous numerical response variable. To address our hypotheses, candidate models were evaluated based on the corrected Akaike Information Criterion (AICc) and assumptions of the selected models were visually assessed by plotting standardized residuals by fitted values (to test for the homoscedasticity of residuals) and by plotting sample quantiles by theoretical quantiles (Q-Q plot to test for the normality of the residuals' distribution). All models' assumptions were met (Appendix B, Figure B.1 to B.3).

We quantified the relationship between leaf biomass and inoculated strain richness by constructing two linear models: a linear regression (additive biodiversity effect hypothesis) and a second-degree polynomial regression (redundant biodiversity effect). Both models were compared to a null model to assess their fit to the observed data. Then, to determine whether specific strains had a higher-than-average positive influence on leaf biomass, without consideration of diversity effects, we built a saturated linear model that included twelve binary variables coding the presence or absence of each strain in samples, without taking into account interactions between strains. We used the function 'dredge' in MuMIn to generate every possible model based on the saturated model (Appendix B, Table B.3a). We reported coefficients and standard errors of included parameters in all models with a ΔAICc $\leq$ 2, to attest the effect of individual strains' presence on leaf biomass and to reveal which strains had a higher-than-average impact. We selected the three strains having a higher-than-average effect on leaf biomass to test if these interacted with diversity to influence the polynomial regression fit linking strain richness and dry leaf biomass. We explored the full model space with the function *dredge*, starting from a saturated linear model that included inoculated strain richness (with second-degree polynomial terms), the presence or absence of each of the three strains, and all possible interactions between those four variables (Appendix B, Table B.3b). To overcome model uncertainty, we performed multi-model inference with the function 'model.avg' in MuMIn to generate fully averaged parameter estimates based on all models with a ΔAICc $\leq$ 2. We used the function 'emmip' in package emmeans (v1.8.2; Lenth, 2022) to extract fitted values and 95% confidence intervals for each combination of the three strains' presence or absence, over the entire range of richness values.

Finally, to analyze whether complementarity or selection was driving the leaf biomass response to inoculated strain richness, we built a conceptual framework based on three alternative models (Appendix B, Figure B.4). We first calculated expected biomass values assuming that, for every mixed bacterial community, all strains' effects on biomass in mixtures would correspond to their mean effects in monoculture. We fitted a linear model (termed the "no diversity effect model") to these expected values (Appendix B, Figure B.5a). Secondly, we calculated expected biomass values assuming that the effect of mixed communities on biomass would correspond to the effect of the occurring strain that had the highest mean effect in monoculture – a pure positive selection effect. The "positive selection model" fitted on those expected values is a second-degree polynomial regression (Appendix B, Figure B.5b). Similarly, we built a "negative selection model" based on the occurring strain having the lowest associated biomass in monoculture – a pure negative selection effect (Appendix B, Figure B.5c). Predictions with higher values than the positive selection model would imply positive complementarity, while predictions lower than the negative selection model would imply negative complementarity. By plotting our predictions based on observed data (Appendix B, Figure B.5d) with these alternative models, we qualitatively assessed how the relative contribution of complementarity and selection varied with increasing inoculated strain richness.

2.5 Results

2.5.1 *Methylobacterium* inoculated strain richness caused variation in *A. thaliana* leaf biomass

The relation between inoculated strain richness and leaf biomass was best modelled by a second-degree polynomial function (AICc = 1055.4, weight = 0.965) (Table 2.1). Strain richness explained close to 8% of the variation in leaf biomass (adj. R^2 = 0.078). We observed an initial decrease in dry leaf biomass at low strain richness (from an average of 28.0 mg for monocultures to 23.4 mg at four strains), but a gradual increase at higher richness (from 23.4 mg at four strains to 36.2 mg at ten strains) (Figure 2.1).

Table 2.1 The second-degree polynomial regression (model ID 1) best explained the relation between *Methylobacterium* inoculated strain richness and *A. thaliana* vegetative growth, compared to a linear regression (model ID 2) and a null model.

Model ID	Parameters	Coefficients ± SE (mg)	*df*	*LL*	AICc	ΔAICc	weight	adj. R^2
1	-	-	4	-523.56	1055.4	0.00	0.965	0.077
	intercept	31.17 ± 2.61						
	richness	-3.56 ± 1.38						
	richness^2	0.41 ± 0.13						
2	-	-	3	-528.31	1062.8	7.38	0.024	0.019
	intercept	24.76 ± 1.65						
	richness	0.60 ± 0.32						
null	-	-	2	-530.10	1064.3	8.88	0.011	0
	intercept	27.27 ± 0.99						

n = 137. SE, standard error; *df*, degrees of freedom; *LL*, log-likelihood.

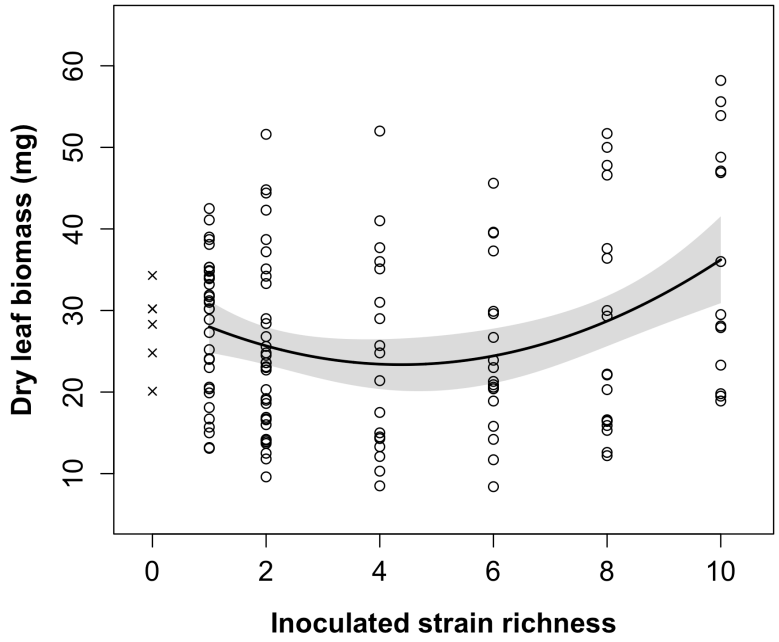


Figure 2.1 *A. thaliana* leaf biomass initially decreased with increasing inoculated taxonomic richness, but the relation became positive once *Methylobacterium* communities contained more than four strains. Solid line indicates fitted values of the model for the range of richness values used in the experiment. Shaded area indicates 95% confidence interval. *n* = 137. "x" data points represent control plants; those samples were not used in the model but were included as a visual assessment of the effect of *Methylobacterium* inoculation on host growth.

2.5.2 *Methylobacterium* inoculated strains influenced *A. thaliana* leaf biomass

We identified three strains with a higher-than-average effect on *A. thaliana* leaf biomass: *Methylobacterium sp.* J-067, *Methylobacterium sp.* E-045, and *Methylobacterium sp.* J-092, henceforth referred to as “high-impact strains” (the other nine strains will be referred to as “low-impact strains”). All three high-impact strains were present in every model with a ΔAICc ≤ 2 (Table 2.2). The first model (ID 1) in Table 2.2 showed that, when the other two high-impact strains were absent (but notwithstanding the presence or absence of the low-impact strains), the presence of J-067 increased biomass by a mean of 5.8 mg (standard error: ± 2.0 mg); J-092 increased biomass by 8.1 (± 2.1) mg, while E-045 decreased biomass by 5.3 (± 2.1) mg. Based on the adjusted R^2 of the six best models (ΔAICc $\leqslant$ 2), the prevalence of the three high-impact strains explained around 15.7% of the variation in leaf biomass. Full results of model diagnostics for all models are presented online (https://github.com/lauzonj/methylobacterium-arabidopsis-growth/blob/main/data/H2_all_models.csv). Figure 2.2 illustrates that communities including J-067 and/or J-092 were often associated with an increased leaf biomass, while the ones containing E-045 tended to show relatively low leaf biomass. In monocultures, only the inoculation of strain J-092 caused an obvious increase in leaf biomass. Furthermore, the effect on plant growth of most monocultures and mixed communities was highly variable, including those containing high-impact strains.

Table 2.2 The initial presence of *Methylobacterium sp.* J-067 or J-092 in communities enhanced *A. thaliana* growth, while E-045 had a detrimental effect.

ID	Intercept	E-046	J-078	J-088	J-059	J-043	J-067	J-048	J-076	E-045	J-092	E-005	J-068	*df*	*LL*	AICc	Δ	wgt	ajd. R^2
1	24.3 ±1.3						5.8 ±2.0			-5.3 ±2.1	8.1 ±2.1			5	-518.75	1048.0	0.00	0.035	0.153
2	24.8 ±1.4					-2.2 ±2.0	6.0 ±2.0			-5.2 ±2.1	8.7 ±2.1			6	-518.11	1048.9	0.92	0.022	0.161
3	24.5 ±1.4						6.9 ±2.3	-2.4 ±2.4		-5.8 ±2.2	9.0 ±2.2			6	-518.27	1049.2	1.22	0.019	0.159
4	24.7 ±1.4						6.0 ±2.0		-1.5 ±2.0	-5.2 ±2.1	8.1 ±2.1			6	-518.46	1049.6	1.61	0.016	0.156
5	24.5 ±1.4			-1.6 ±2.2			6.0 ±2.0			-5.2 ±2.1	8.7 ±2.2			6	-518.47	1049.6	1.62	0.015	0.156
6	24.0 ±1.4		1.3 ±2.0				5.6 ±2.0			-5.5 ±2.2	7.9 ±2.1			6	-518.56	1049.8	1.80	0.014	0.155

Parameters (intercept and strains) and their coefficients ± SE (mg) are shown for all models with a ΔAICc ≤ 2 (see Appendix B, Table B.3a for saturated model). Coefficients indicate, for a given strain in a given model, its positive or negative impact on host dry leaf biomass (mg) when the other strains included in the model are absent from communities, notwithstanding the presence or absent of strains not included in the model. Intercept indicates the mean biomass of samples when all strains included in the corresponding model are absent. *n* = 137. ID, model number; SE, standard error; *df*, degrees of freedom; *LL*, log-likelihood; Δ, ΔAICc; wgt, model weight.

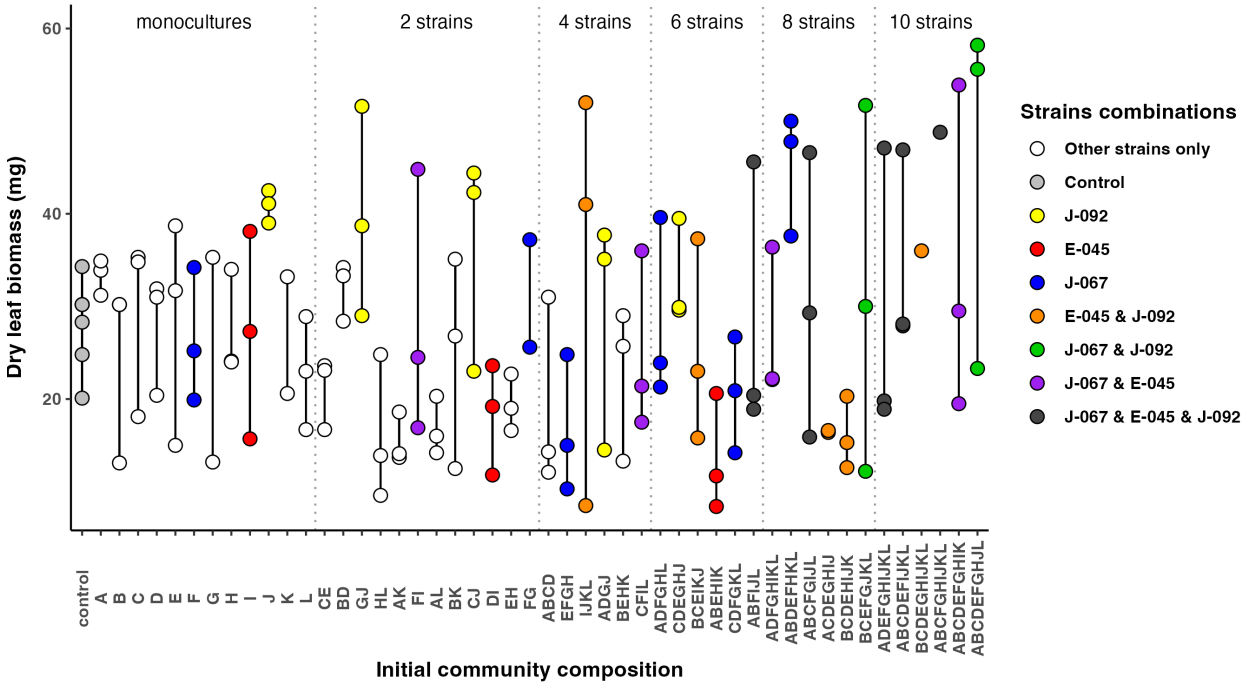


Figure 2.2 Dry leaf biomass of *A. thaliana* at the time of harvest in function of all the different initial compositions of synthetic communities. J-092 was the only strain in monoculture that caused an increase in leaf biomass. Most monocultures and mixed communities showed high variability in their effect on plant growth. Dots show observed data (n = 143) and are coloured according to the different high-impact strains combinations (see legend on the right). Grey dotted lines separate community compositions according to strain richness. Letters on the x-axis refers to the strains composing the communities: A: E-046, B: J-078, C: J-088, D: J-059, E: J-043, F: J-067, G: J-048, H: J-076, I: E-045, J: J-092, K: E-005, L: J-068 (Appendix B, Table B.1).

2.5.3 Interactions among strains and between strains and richness determined the biomass response to richness

To investigate the effect of specific strains on the relation between richness and leaf biomass, we explored all models that could be derived from a saturated model which included the three high-impact strains, the polynomial effect of richness and all possible interactions between those variables. The six best models (ΔAICc ≤ 2) all had a lower AICc (1032.4 to 1034.2) than our two previous hypotheses' models and explained between 31.3% and 35.8% of leaf biomass variation based on their adjusted R^2 (Table 2.3a). Full results of model diagnostics for all models are presented online (https://github.com/lauzonj/methylobacterium-arabidopsis-growth/blob/main/data/H3_all_models.csv).

Figure 2.3 shows predicted leaf biomass values in function of inoculated strain richness, under different composition scenarios, based on the averaged model (Table 2.3b). The general response pattern of biomass with increasing richness (an initial decrease in biomass followed by an increase) was found throughout all high-impact strains combinations, as well as in communities inoculated with only low-impact strains. At high diversity values, communities containing J-067 (without E-045 nor J-092), and those containing J-067 and J-092 (without E-045), showed the steepest positive richness-biomass relationships. E-045 was neutral in monoculture but, when inoculated into mixed assemblages without J-067 nor J-092, provoked a steep decrease in the richness-biomass function. E-045 also diminished the strength of the relationship (the curve flattened) when communities harboured J-067 or J-092. The weakest relations between diversity and productivity occurred when only low-impact strains were inoculated, as well as when communities contained all three high-impact strains.

Table 2.3 *Methylobacterium sp.* J-067, E-045, and J-092 interacted among themselves, as well as with inoculated strain richness, to modulate *A. thaliana* growth response to bacterial diversity. (a) Models with a ΔAICc ≤ 2 (see Table B.3b for saturated model). (b) Averaged model's full coefficients indicating the impact of diversity, strains, and their interactions on host biomass, to predict dry leaf biomass at different richness values and strain combinations.

a) Model exploration

	Parameters (strains)																			
ID	**J-067**	**E-045**	**J-092**	**D²**	**J-067:E-045**	**J-067:J-092**	**J-067:D²**	**E-045:J-092**	**E-045:D²**	**J-092:D²**	**J-067:E-045:J-092**	**J-067:E-045:D²**	**J-067:J-092:D²**	**E-045:J-092:D²**	**J-067:E-045:J-092:D²**	***df***	***LL***	**AICc**	**ΔAICc**	**weight**	**ajd. R^2**
1	+	+	+	+	+			+	+	+						13	-501.72	1032.4	0.00	0.110	0.339
2	+	+	+	+					+	+						11	-504.35	1032.8	0.41	0.089	0.313
3	+	+	+	+	+				+	+						12	-503.27	1033.1	0.65	0.079	0.324
4	+	+	+	+	+			+	+	+				+		15	-499.74	1033.5	1.05	0.065	0.358
5	+	+	+	+				+	+	+						12	-503.69	1033.9	1.49	0.052	0.320
6	+	+	+	+	+	+		+	+	+						14	-501.36	1034.2	1.76	0.045	0.343

"+" shows which parameters were selected in the different models; ":" denotes interactions; n = 137. D^2, polynomial strain richness; *df*, degrees of freedom; *LL*, log-likelihood.

b) Averaged model parameters and coefficients

Parameters	**coefficients ± adj. SE (mg)**
Intercept	32.92 ± 3.35
J-067	1.73 ± 3.87
E-045	1.10 ± 7.31
J-092	18.55 ± 7.20
D	-7.40 ± 2.26
D^2	1.00 ± 0.28
J-067:E-045	5.90 ± 5.86
J-067:J-092	-0.48 ± 2.33
E-045:J-092	11.24 ± 19.35
E-045:D	-0.47 ± 3.60
E-045:D^2	-0.26 ± 0.37
J-092:D	-1.23 ± 3.60
J-092:D^2	-0.20 ± 0.36
E-045:J-092:D	-2.26 ± 6.48
E-045:J-092:D^2	0.20 ± 0.55

D^2 and D, respectively 2nd and 1st-degree terms in polynomial function of strain richness; ":" denotes interactions; adj. SE, adjusted standard error.

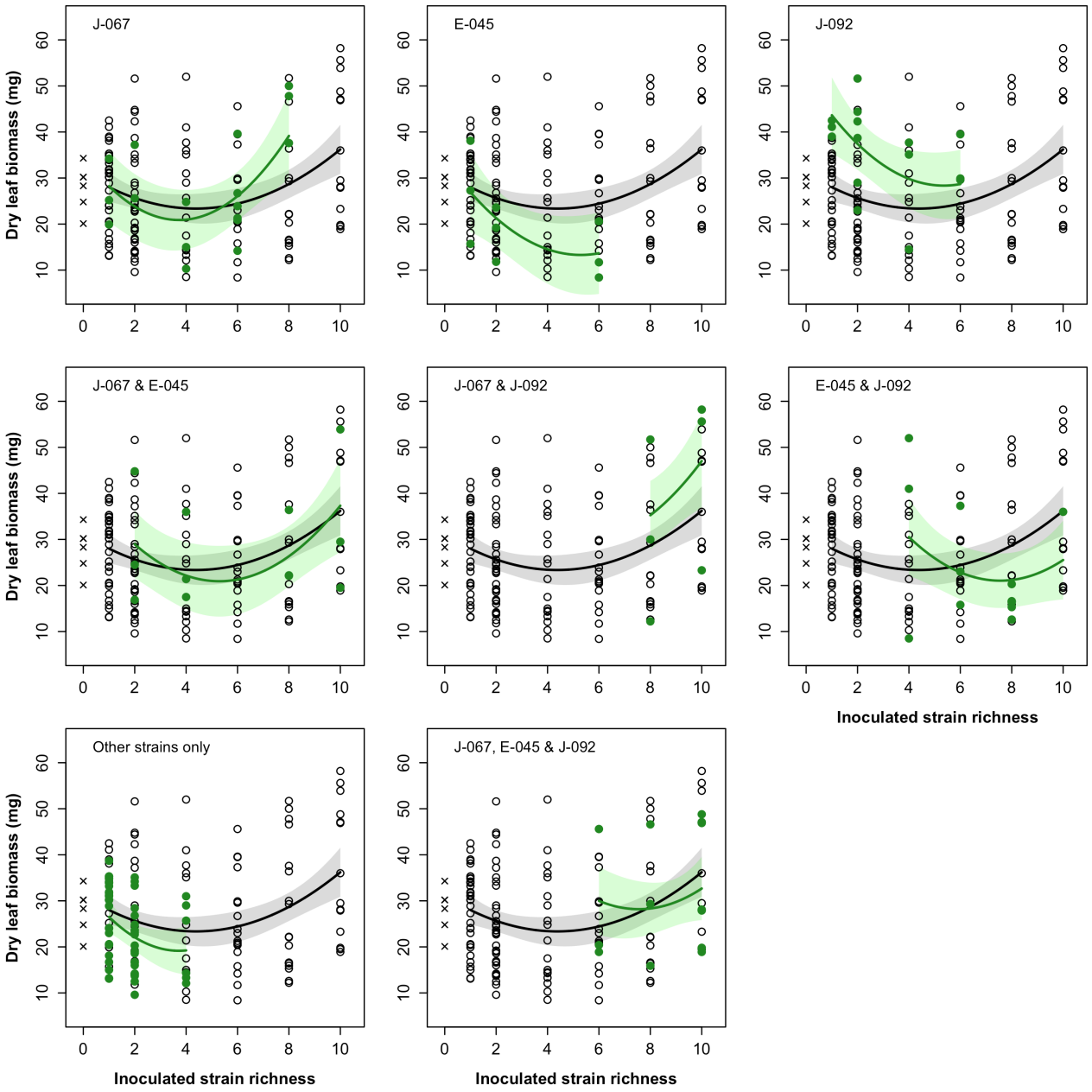


Figure 2.3 The effect of *Methylobacterium* inoculated taxonomic richness on *A. thaliana* leaf biomass was influenced by the presence (or absence) of strains J-067, E-045, and J-092. Model averaging was performed (Table 2.3b) to reveal how richness and those three high-impact strains interact to shape the observed diversity-productivity pattern. Eight panels illustrate the relation between strain richness and leaf biomass for samples (green points) containing a specific combination of strains (top left of each panel). Solid lines indicate fitted values of the averaged model based on fully averaged coefficients (green) and fitted values of the simple polynomial regression model (biomass ~ richness^2; black). Shaded areas indicate 95% confidence intervals. *n* = 137. "x" data points represent control plants; those samples were not used in the model but were included in the figure as a visual assessment of the effect of *Methylobacterium* inoculation on host growth.

2.5.4 The inoculated strain richness-leaf biomass relationship deviated from the alternative models

In communities composed of two to six *Methylobacterium* strains, our data-fitted model revealed that dry leaf biomass of *A. thaliana* was lower than expected if all strains had the same effect on biomass in mixed assemblages and in monocultures, given their relative abundance (Figure 2.4). At inoculated richnesses of two and four strains, the curve of the biomass response to richness was similar to the negative selection model predictions, but started to deviate from these predictions with a slope becoming positive beyond four strains. At a richness of eight strains, the observed biomass corresponded to the no diversity effect model, and when communities reached ten strains, most observed biomass values were higher than those expected by the no diversity effect model. The model we fitted on observed data did not match the positive selection model's response curve at any richness values.

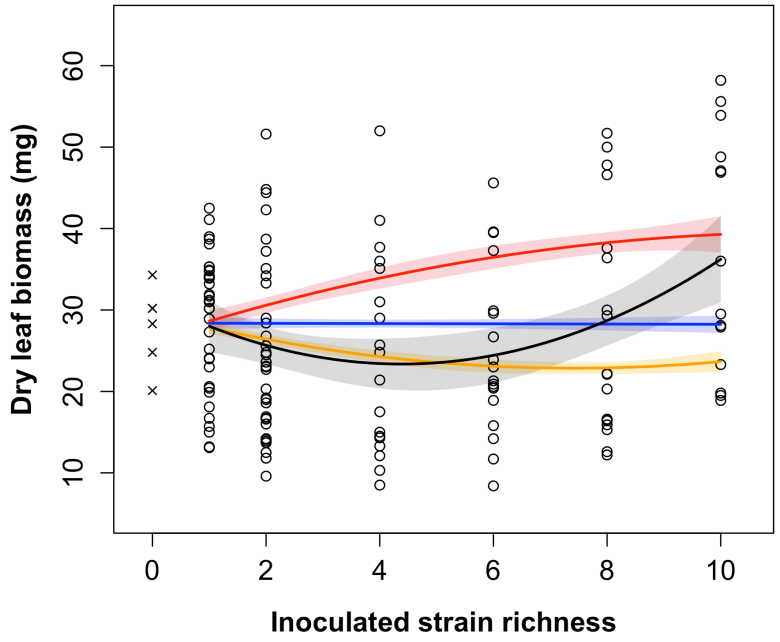


Figure 2.4 When two to six strains were inoculated, measured *A. thaliana* leaf biomass associated with mixed *Methylobacterium* communities was generally lower than expected if all strains were influencing plant growth as they do in monocultures; however, ten-strains assemblages often had a higher than expected influence on plant growth. Except at a ten-strains richness, leaf biomass predicted by the data-fitted model was consistently lower than expected if mixed communities' influence on plant growth was similar to the effect of the most growth-promoting strain in monoculture. Black lines represent predicted values of the regression model fitted on the observed data. Blue lines represent the no diversity effect model (Appendix B, Figure B.5a), red lines the positive selection model (Appendix B, Figure B.5b), and orange lines the negative selection model (Appendix B, Figure B.5c). Solid lines indicate fitted values of the different models for the range of richness values used in the experiment. Shaded areas indicate 95% confidence interval. *n* =

137. “x” data points represent control plants; those samples were not used in the models but were included as a visual assessment of the effect of *Methylobacterium* inoculation on host growth.

2.6 Discussion

2.6.1 Causality revealed: leaf bacterial diversity impacted individual plant productivity

We detected an effect of inoculated leaf bacterial diversity on plant growth in a fully controlled gnotobiotic experimental environment, indicating that foliar *Methylobacterium* taxonomic diversity causes variation in host plant productivity. However, we had predicted that functional redundancy among strains would result in a positive logarithmic relationship between inoculated strain richness and harvested leaf biomass, but the data portrayed an alternative pattern: increasing diversity caused an increase in plant growth, but only after an initial decrease at low strain richness compared with single strain effects (Figure 2.1). This pattern suggests an underlying ecological mechanism by which *Methylobacterium* diversity might have a small inherent negative effect on plant growth that can be overcome when diversity reaches a certain threshold. Direct or indirect inter-strain competition could explain this phenomenon. In a synthetic communities experience on *A. thaliana*, Carlström *et al.* (2019) demonstrated that when a *Methylobacterium* strain was dominant in a community, the other two inoculated *Methylobacterium* strains tended to be present in low abundance. Furthermore, Schäfer *et al.* (2022) found that most bacterial pairwise interactions in the phyllosphere were negative, and Becker *et al.* (2012) demonstrated that increased strain richness in the rhizosphere could lead to a decrease in plant protection against pathogens. However, Grilli *et al.* (2017) demonstrated that multi-species interactions can stabilize competition dynamics. Likewise, by assembling synthetic bacterial communities in the zebrafish gut, Sundarraman *et al.* (2020) revealed that strong pairwise competition was reduced in communities containing four to five strains, resulting in higher-than-predicted cell abundance of every strain.

In the light of those findings, we hypothesize that the observed initial negative effect of diversity on host biomass could be due to strong pairwise competition that ultimately reduced bacterial abundance and thus the potential for growth enhancement, for example by reducing the production of growth-producing chemical compounds such as phytohormones. Multi-species competition alleviation could then explain the diversity-driven increase in leaf biomass occurring when more than four strains were inoculated. The fact that all strains belonged to the same genus

could also have exacerbated the strong pairwise competition, since closely related organisms tend to share similar traits and niches (Cavender-Bares *et al*., 2009; Martiny *et al*., 2013).

2.6.2 *Methylobacterium* strains differed in their effect on plant growth and shaped bacterial diversity-host productivity relationship

Methylobacterium strains had varying influences on plant growth, validating our second hypothesis. We identified three strains exhibiting a high impact on *A. thaliana* vegetative growth (Table 2.2). Two of these strains, *Methylobacterium sp.* J-067 and *Methylobacterium sp.* J-092, had a higher-than-average positive impact on leaf biomass, notwithstanding community diversity. However, we had not predicted to find a strain, *Methylobacterium sp.* E-045, with a global detrimental effect on its host, since *Methylobacterium* is generally considered a beneficial growth-promoting genus – although in field inoculation experiments, Flores-Núñez *et al.* (2023) identified a *Methylobacterium* strain reducing the growth of *Agave tequilana*, and Sanjenbam *et al.* (2022) demonstrated that non host-native strains could reduce rice plant health. Furthermore, even if those three high-impact strains showed a general tendency to either be beneficial or detrimental to their host, a noticeable fraction of plant samples harbouring these strains in monocultures or in mixed communities had a leaf biomass opposite to the expected effect of its symbionts (Figure 2.2). The fact that the effect of a distinct inoculated community composition on plant growth is highly variable hints at the importance of stochastic events (i.e., drift) in population and community dynamics, succession and alternative states (Maignien *et al.*, 2014). Indeed, Le Moigne *et al.* (2023) have shown that communities developing from an identical initial bacterial source can result in different compositions over time. This could in turn lead to a variation in community functions and impacts on the host plant.

The way bacterial diversity drove host productivity was shaped in part by which *Methylobacterium* taxa were present in the communities (Figure 2.3), thus supporting our third hypothesis. By taking into account the presence of only three high-impact strains along with inoculated richness, our models were able to explain around the third of all the variance observed in host growth in our experiment (Table 2.3). Interestingly, even in the presence of beneficial strains J-067 and J-092, the biodiversity-productivity function showed an initial negative relation, implying that pairwise microbial interactions could overcome and restrain beneficial bacteria-host symbioses. In addition, our prediction that high-impact strains would steepen the regression slope was validated. At high richness values, when the beneficial strains J-067 and J-092 were inoculated in mixed communities, the associated plant biomass increased faster with increasing diversity. At low

richness values, the high-impact detrimental E-045 strain exacerbated the initial negative effect of diversity, in addition to neutralizing the effect of diversity on plant growth in rich assemblages. Furthermore, although the effects of J-067 and E-045 in monoculture were globally neutral, these strains behaved respectively as good and bad partners in mixed communities, corroborating a study by Garbeva *et al.* (2011) that showed that bacterial responses to competition are taxon-specific. This highlights the importance of considering how individual strains could interact with other bacteria in the phyllosphere before assessing their effects on the host. Furthermore, our finding that the biodiversity-ecosystem function relationship was dependent on composition in our study system was supported by a few studies showing the importance of composition on diversity-driven bacterial community functions (Armitage, 2016; Bell *et al*., 2005; Jousset *et al*., 2014).

2.6.3 The impact of *Methylobacterium* diversity on *A. thaliana* growth was plausibly driven by complementarity

Compared to most studies analyzing the effect of diversity of a group of microorganisms on the functions of these organisms' communities or abiotic environment (Bell *et al*., 2005; Bernhard and Kelly, 2016; Osburn *et al*., 2023; Philippot *et al*., 2013), our experiment investigated the effect of symbiont diversity on the functioning of their host. Without being able to measure the growth of the inoculated bacterial strains, we could not apply the classical equation from Loreau and Hector (2001) to measure the relative contribution of complementarity and selection to the biodiversity effect. However, by analyzing how host growth responded to increasing bacterial diversity in a conceptual framework based on alternative models (Figure B.4), we were able to qualitatively disentangle the mechanisms at play in different parts of the tested richness spectrum. The host growth response curve to diversity matched the negative selection model's prediction curve for two- and four-strain communities (Figure 2.4). This pattern could indicate negative species selection, although species selection alone could not explain the fact that the observed biomass-richness relationship clearly diverged from the negative selection model predictions when more than four strains were inoculated. If one or two particular host-detrimental strains were driving the biodiversity effect, it is difficult to conceive a diversity-dependent process that could relieve the bacterial assemblages of the dominance of these strains. On the other hand, complementarity, first through negative pairwise interactions then through positive multispecies competition alleviation, could explain the observed biomass response pattern to diversity at these relatively low richness values. In communities of six to ten strains, it was difficult to confidently separate complementarity versus selection effects, as the observed pattern did not fit any of the alternative selection models, nor did it go beyond their predictions to suggest complementarity. However, it

is worth noting that the general trend in our data did not resemble at all the predictions of the positive selection model, ruling out the possibility of pure positive species selection driving the diversity-productivity relationship.

The detection of three high-impact strains might suggest a certain underlying effect of species selection since the presence of those specific strains was associated with higher or lower than average plant biomass across richness values. However, out of those three strains, only J-092 showed an effect in monoculture (Figure 2.2), suggesting that J-067 and E-045 effects on their host might be mostly due to how they interact with other partners (positive or negative complementarity) rather than due to their intrinsic growth-promoting or growth-inhibiting characteristics. For example, J-067 might be a facilitator or community stabilizer, while E-045 might be a strong competitor. Furthermore, the three strains associated with the lowest plant biomass in monocultures and thus driving the negative selection model's predictions (J-068, J-048 and J-078) were not statistically associated with low biomass across the entire range of richness values. Finally, another observation supporting complementarity – and the importance of diversity *per se* – is the increase in productivity of plants when richness jumped from eight to ten strains, since all high-impact strains were already found interacting in communities at both of these diversity levels. Interestingly, only communities of ten strains showed a net positive biodiversity effect (Figure 2.4), and if the predicted biomass response to richness based on observed data was to be extrapolated over ten-strain richness, the values would be higher than expected by the positive selection model, indicating a positive complementarity effect. This highlighted that seemingly low-impact strains could actually be important actors driving diversity-productivity relationships. *A. thaliana* Col-0 grown in non-gnotobiotic conditions (i.e., with a wild microbiome) has a mean rosette dry biomass of 163.7 (± 52.0) mg at the corresponding growth stage at which we harvested the plants (Boyes *et al*., 2001), whereas the highest biomass we recorded in our experiment is just under 60 mg. This suggests that a reduced foliar bacterial diversity such as in our experiment could negatively impact plant growth, and most importantly that there is potential for increased biomass with increased bacterial richness. Overall, our experiment suggests that complementarity effect is most plausibly the key mechanism generally shaping the diversity-productivity relationship – a conclusion shared by Laforest-Lapointe *et al*. (2017a) – although it is likely that both complementarity and selection act in richness- and taxa-dependent ways to drive the overall effects of *Methylobacterium* diversity on plant host productivity.

2.7 Conclusion

Our study showed that increasing *Methylobacterium* diversity caused an increase in host vegetative growth, but only after reaching a certain richness threshold. Some *Methylobacterium* strains had beneficial or detrimental impacts on plant productivity, but their effect was highly variable in monocultures and in mixed assemblages. Strains also differed in their influence on the diversity-productivity relationship, their impact depending on associations with other strains. Finally, even if species selection likely contributes to explaining the net biodiversity effect we observed, our results suggest complementarity was the main ecological mechanism driving plant productivity in our experiment. What is especially interesting about our study is that even though some biodiversity-ecosystem functioning studies showed positive (Bell *et al*., 2005; Eisenhauer *et al*., 2012; Laforest-Lapointe *et al*., 2017a), neutral (Jiang, 2007) or negative (Becker *et al*., 2012) bacterial biodiversity effects, few, if any, have found that this relationship changes with varying values of microbial diversity. The complex plant productivity response to bacterial biodiversity in our study highlights the fact that there is a need to better understand competition, coexistence, and assembly dynamics in diverse microbial communities (Levine *et al*., 2017) – both in host-dependent and independent contexts –, and to identify the mechanisms (e.g., variation in phytohormones production) by which microbial interactions impact the host. This knowledge will be crucial to better understand plant-microbes symbioses in the context of global change (Perreault and Laforest-Lapointe, 2022) – a topic of special interest for applied sciences such as forestry, agriculture, and conservation.

2.8 Acknowledgements

We would like to thank François Ouellet for providing *A. thaliana* seeds and plant growth medium ingredients, as well as for his advice on growing *A. thaliana*; Geneviève Bourret for her help in the laboratory; Sylvain Dallaire for the maintenance of the growth chamber; Daniel Schönig for his help with the analyses; and Jean-Baptiste Leducq for his previous work on isolating and characterizing the *Methylobacterium* strains used in the experiment.

CHAPITRE 3
CONCLUSION

Ce 3e et dernier chapitre présente une synthèse des principaux résultats des travaux de recherche présentés dans les deux chapitres précédents en lien avec les objectifs et hypothèses initiales, ainsi qu'une discussion critique de leurs implications biologiques. J'y mets en lumière la contribution de mon mémoire au domaine de l'écologie et de l'évolution microbiennes, souligne les limites de mes travaux et propose des pistes à explorer afin de poursuivre de manière pertinente la recherche portant sur le genre bactérien *Methylobacterium* et, plus généralement, sur les interactions entre bactéries et plantes hôtes au sein de la phyllosphère.

3.1 L'espèce hôte oriente l'écologie et l'évolution du genre *Methylobacterium*

Mon premier chapitre a mis en lumière le rôle prépondérant de l'espèce hôte dans le façonnement du microbiome foliaire de la forêt tempérée nordique, et plus précisément sur l'écologie et l'évolution du genre *Methylobacterium* au sein de la phyllosphère. De manière globale, toutes mes hypothèses ont été validées. En effet, mes analyses ont démontré que l'espèce d'arbre ou d'arbuste hôte déterminait près de la moitié de la variation taxonomique observée au niveau du genre au sein des communautés bactériennes, et environ le tiers de la variation en termes d'espèces et de gènes parmi les communautés de *Methylobacterium*. Notamment, *Corylus cornuta*, la seule espèce arbustive étudiée, abritait le plus grand nombre d'espèces de *Methylobacterium*, dont plusieurs espèces uniquement présentes chez elle, et le genre y occupait une plus grande proportion des communautés bactériennes. L'influence de l'espèce hôte observée dans mon étude est en concordance avec la recherche antérieurement effectuée sur la phyllosphère des arbres tempérés (Laforest-Lapointe *et al.*, 2016a; Leducq *et al.*, 2022a) et avec le fait que chaque espèce hôte fournit aux bactéries une niche foliaire distincte en termes de morphologie et de phytochimie (de la Riva *et al.*, 2017; Gargallo-Garriga *et al.*, 2020; Nicotra *et al.*, 2011; Schweiger *et al.*, 2021). Les différences les plus marquées dans la structure taxonomique et génétique des communautés bactériennes, à tous les niveaux d'organisation écologique, ont été observées entre les espèces hôtes arborescentes et arbustive, ainsi qu'entre les espèces feuillues et conifériennes. Cela suggère que les pressions de sélection divergent fortement entre ces types d'hôtes et viennent définir deux grands axes de variation compositionnelle en termes de genres, d'espèces et de gènes. L'ensemble de mes analyses indique donc que le port de la plante, tout comme sa feuille, exercent une influence majeure sur

le microbiome de la phyllosphère, incluant sur l'évolution de ses populations de bactéries. D'autre part, en exposant de nombreuses associations entre gènes, fonctions et espèces hôtes, ainsi qu'en révélant des signaux de sélection positive spécifiques à l'hôte dans les populations de la phyllosphère, j'ai démontré qu'une partie substantielle du pangénome accessoire de *Methylobacterium* contribuait aux adaptations aux niches foliaires distinctes des espèces hôtes, tel qu'anticipé selon le modèle du pangénome adaptatif formulé par McInerney *et al.* (2017). Bien que je n'aie pas investigué directement les mécanismes d'évolution du pangénome, mes résultats suggèrent que le transfert horizontal de gènes entre populations occupant une même niche, ainsi que la duplication de gènes, ont probablement façonné en partie le pangénome en réponse aux conditions foliaires. Enfin, la convergence des patrons qui ont émergé de mes résultats à tous les niveaux d'analyses m'a amené à formuler deux hypothèses pour expliquer les différences dans la phyllosphère entre arbres et arbustes, et entre feuillus et conifères : l'exposition au soleil et la phénologie des feuilles.

3.2 Hypothèse de l'exposition au soleil

La première hypothèse formulée suite à l'interprétation des résultats de mon premier chapitre est que l'exposition au soleil joue un rôle clé dans la différenciation des communautés bactériennes foliaires entre les arbres et les arbustes au sein d'une forêt fermée. Plus les feuilles sont hautes dans la canopée forestière, plus elles reçoivent de rayonnement solaire. Conséquemment, les bactéries foliaires des arbres sont généralement plus exposées au soleil que celles des arbustes et des herbacées. Des variations dans la structure des communautés bactériennes de la phyllosphère en fonction du positionnement dans les strates verticales de la canopée ont déjà été mesurées (Herrmann *et al.*, 2021; Stone et Jackson, 2019), tout comme l'influence du rayonnement UV sur ces communautés (Carvalho et Castillo, 2018; Jacobs et Sundin, 2001; Truchado *et al.*, 2017). Le rayonnement solaire est une ressource énergétique inépuisable et précieuse dans un environnement pauvre en carbone comme la phyllosphère. Or, les bactéries foliaires exposées au soleil sont aussi périodiquement soumises à une température élevée, ainsi qu'au rayonnement ultraviolet qui entraîne la production de dérivés réactifs de l'oxygène et un stress oxydatif (Santos *et al.*, 2012, 2013). La lumière du soleil exerce donc une forte pression de sélection au sein de la phyllosphère. Elle favorise sur les feuilles des espèces hôtes les plus exposées (celles des arbres dans une forêt fermée) les bactéries capables d'assimiler efficacement l'énergie lumineuse tout en se protégeant de la chaleur et du rayonnement UV.

Plusieurs de mes résultats suggèrent en effet que le microbiome de la phyllosphère est fortement influencé par la quantité d'énergie lumineuse reçue. D'abord, les communautés de la phyllosphère diffèrent fortement selon qu'elles résident sur un arbre ou sur un arbuste. En effet, pour chacune de mes analyses en coordonnées principales, le premier axe (représentant la plus grande proportion de la variation compositionnelle) séparait nettement les communautés arborescentes des communautés arbustives. Les gènes enrichis chez les communautés de *Methylobacterium* résidant sur les espèces d'arbre (et donc les communautés les plus exposées au soleil) étaient principalement impliqués dans le transport et le métabolisme des acides aminés, la traduction et les ribosomes, ainsi que la production et la conversion de l'énergie. Mais surtout, plusieurs gènes enrichis étaient impliqués dans la photosynthèse, la phosphorylation oxydative et la réponse aux stress oxydatif et de chaleur. Ces résultats contrastent avec ceux obtenus pour l'espèce arbustive *C. cornuta*. Dans les communautés de *Methylobacterium* résidant sur les feuilles de *C. cornuta,* les gènes enrichis étaient principalement impliqués dans le mobilome, la transcription, la défense et la composition en protéines et lipides de la membrane cellulaire. Cet enrichissement suggère une plus grande pression de sélection provenant des interactions biotiques (bactéries à bactéries) que de l'exposition au soleil, en raison d'une probable immigration continuelle de microorganismes provenant du sol ou des précipitations.

3.3 Hypothèse de la phénologie des feuilles

La deuxième hypothèse formulée dans le cadre du premier chapitre est celle de la phénologie des feuilles. La distinction entre les communautés de la phyllosphère des conifères et des feuillus a déjà été soulignée dans de nombreuses analyses (Laforest-Lapointe *et al.*, 2016a; Lajoie et Kembel, 2021), sans que soit explicitement formulée une hypothèse globale en lien avec la phénologie foliaire. Dans la forêt tempérée du nord-est de l'Amérique du Nord, toutes les espèces de conifères sauf le mélèze laricin (*Larix laricina* [Du Roi] K. Koch) maintiennent leurs aiguilles durant la période hivernale, contrairement à toutes les espèces d'arbres feuillus et à la grande majorité des espèces arbustives feuillues (sauf plusieurs petits arbustes de la famille *Ericaceae*) qui perdent leurs feuilles. Mon hypothèse stipule que le microbiome de la phyllosphère est fortement influencé par la durée de persistance de l'habitat foliaire. Les communautés bactériennes des espèces de plantes à feuillage sempervirent, soit la majorité des conifères et des arbustes *Ericaceae*, persistent au travers des saisons (malgré une certaine variation compositionnelle). En conséquence, ces communautés sont soumises à de plus grandes fluctuations climatiques et vivent sur des feuilles en moyenne plus âgées. De plus, la persistance de la surface foliaire permet aux communautés d'atteindre des états successionnels plus avancés

dans lesquels les processus d'assemblage déterministes prennent plus d'importance que les processus stochastiques. Cela peut ainsi mener ainsi à une sélection homogène (Dini-Andreote *et al.*, 2015) des populations formant les communautés coniférienes. Pour leur part, les communautés bactériennes vivant à la surface des plantes à feuillage caduc (soit la quasi-totalité des espèces angiospermes) se désagrègent complètement chaque automne. Puis, bien que des sous-populations migrent vers des compartiments persistants tels que les bourgeons (Doronina *et al.*, 2004), les communautés se recomposent au débourrement printanier, un événement de colonisation foliaire massif et stochastique. En conséquence, les populations bactériennes qui s'établissent sur les feuillus sont soumises à de nouvelles interactions interspécifiques. La phénologie des feuilles des espèces hôtes impose donc des pressions de sélection fortement divergentes entre plantes à feuillage persistant et caduc.

L'ensemble de mes analyses soutient cette hypothèse sur l'influence de la phénologie foliaire. La composition en genres des communautés bactériennes était nettement différenciée entre les feuillus et les conifères, les communautés des conifères étant notamment plus homogènes que celles des feuillus. Au niveau de la composition taxonomique des communautés de *Methylobacterium*, les quelques espèces dominantes étaient différentes entre conifères et feuillus, tandis qu'au niveau de leur composition génique, la séparation était nettement marquée entre les deux types d'hôtes. Les gènes enrichis dans les communautés de *Methylobacterium* vivant sur les conifères étaient principalement impliqués dans les métabolismes des acides aminés et des lipides. Notamment, des gènes jouant des rôles clés dans les métabolismes des acides gras à longues chaînes et des glycérophospholipides étaient associés aux communautés coniférienes. L'enrichissement de ces gènes participe potentiellement à une réponse plus efficace (et à une meilleure survie) aux fluctuations saisonnières de température en remodelant la structure, et donc la flexibilité, de la membrane cellulaire. De plus, l'enrichissement des gènes impliqués dans l'oxydation de la méthylamine pourrait refléter une sécrétion potentiellement plus élevée de cette amine par les feuilles âgées (Shiraishi *et al.*, 2015). D'autre part, les gènes enrichis dans les communautés de *Methylobacterium* résidant à la surface des feuilles d'espèces décidues étaient surtout impliqués dans la signalisation, la transcription, la membrane cellulaire et la défense, suggérant une plus forte capacité à interagir avec les autres microorganismes colonisateurs. Enfin, de manière analogue aux communautés résidant sur les arbustes (mais découlant d'un type de dispersion différente au sein de la métacommunauté, c'est-à-dire d'une colonisation printanière massive plutôt que d'une migration continuelle), l'importance des interactions biotiques pourrait

expliquer la plus grande proportion de gènes sélectionnés positivement chez les espèces feuillues que chez les conifères.

3.4 La biodiversité de la phyllosphère module la productivité de l'hôte

Mon deuxième chapitre a mis en lumière l'influence directe de la biodiversité foliaire – et plus précisément des taxons de *Methylobacterium* – sur la croissance des plantes hôtes. Mon étude vient supporter l'hypothèse de causalité dans l'augmentation conjointe de la diversité bactérienne sur les feuilles et de la productivité des écosystèmes (Laforest-Lapointe *et al.*, 2017a). Mes analyses ont validé chacune de mes trois hypothèses de travail, mais certains patrons de variation contraires à mes prédictions sont venus apportés des nuances importantes, qui ont servi à mieux comprendre les dynamiques écologiques entre les souches et leurs impacts sur la plante hôte. Une augmentation du nombre de souches (la richesse) dans les communautés foliaires causait une augmentation de la biomasse végétative de l'hôte *A. thaliana*, mais seulement après une réduction initiale de la biomasse des plantes abritant des communautés de deux ou quatre souches, et ce nonobstant l'identité des souches. Ce phénomène s'expliquerait possiblement par un effet d'allègement de la compétition interspécifique entraîné par la complexité croissante des interactions entre de multiples espèces (Grilli *et al.*, 2017; Sundarraman *et al.*, 2020), à partir de six souches dans le cas de mon expérience. De plus, j'ai détecté deux souches de *Methylobacterium* bénéfiques et une souche néfaste pour la croissance de *A. thaliana,* et ce malgré la grande variation de la croissance de l'hôte associée à des communautés de composition initialement identique, autant en monoculture qu'en mélange. La présence de ces souches dans les communautés de différentes richesses influençait substantiellement la nature de la relation entre biodiversité bactérienne et croissance végétale, démontrant ainsi clairement le rôle crucial de la composition des communautés dans le fonctionnement des écosystèmes diversifiés (Bell *et al.*, 2005; Jousset *et al.*, 2014; Langenheder *et al.*, 2010). Enfin, l'ensemble des analyses suggérait que, sans écarter une influence possible de l'effet de sélection des espèces dominantes, la complémentarité des niches était le principal mécanisme déterminant la relation ultimement positive entre la richesse des communautés bactériennes de la phyllosphère et la croissance de l'hôte.

3.5 Limites des travaux

Pour répondre aux objectifs du premier chapitre, j'ai choisi des approches analytiques qui, comme toute approche, viennent avec quelques limitations. Afin de pouvoir plus aisément analyser les communautés de *Methylobacterium* à l'échelle de l'espèce, j'ai sélectionné un seul génome de

référence pour représenter chaque espèce, entraînant une représentation potentiellement incomplète de la diversité génique des espèces individuelles (Valiente-Mullor *et al.*, 2021). D'autre part, je n'étais pas en mesure de détecter les événements de transfert horizontal, de duplication et perte de gènes au sein des communautés et des populations naturelles. Bien que l'isolation de souches sur les différentes espèces hôtes aurait permis des analyses plus précises au niveau de l'évolution des populations et du pangénome (Shapiro *et al.*, 2012), l'approche par culture, contrairement à l'approche métagénomique, n'aurait pas permis de dresser un portrait exhaustif de la diversité des communautés bactériennes (Yashiro *et al.*, 2011; Youseif *et al.*, 2021). Dans les études subséquentes, une approche hybride pourrait être envisagée pour pallier aux biais inhérents aux deux approches (Diba *et al.*, 2023). Enfin, j'ai choisi d'opter pour un séquençage métagénomique très profond d'un petit nombre d'échantillons, plutôt que d'opter pour un séquençage superficiel d'un grand nombre d'échantillons. Cela a limité le pouvoir statistique et explique possiblement l'absence de détection de gènes différentiellement abondants entre *A. balsamea* et *T. occidentalis*, ainsi qu'entre *A. saccharum* et *F. grandifolia*.

En ce qui concerne l'expérience du deuxième chapitre, mon interprétation de la relation entre biodiversité et productivité est quelque peu limitée par le fait que je n'ai pas séquencé les communautés bactériennes des feuilles à la fin de la période de croissance végétative de la plante hôte. En effet, mes analyses sont basées sur la richesse et la composition des communautés synthétiques inoculées au début de la croissance de la plante. Or, ces communautés ont certainement évolué au fil du temps (Maignien *et al.*, 2014). Connaître l'abondance relative de chacune des souches dans sa communauté au moment de mesurer la biomasse végétale m'aurait permis d'établir des liens plus solides entre des processus écologiques tels que la compétition, la facilitation et la dérive, et la croissance de la plante hôte. D'autre part, bien que *Methylobacterium* soit un genre ubiquiste de la phyllosphère et une composante importante du microbiome naturel des feuilles de *A. thaliana* (Knief *et al.*, 2010), il aurait été intéressant d'incorporer aux communautés synthétiques des souches appartenant à d'autres groupes taxonomiques, pour augmenter le bassin de diversité écologique et génétique. Enfin, les communautés synthétiques, parce qu'elles abritent un nombre de taxons relativement bas en comparaison avec les communautés naturelles, ne contiennent parfois pas certaines espèces ou métabolismes clés (Vorholt *et al.*, 2017).

3.6 Importance de mes travaux

Methylobacterium est un genre bactérien omniprésent et écologiquement important au sein de la phyllosphère, parce qu'il peut stimuler la croissance de son hôte (Abanda-Nkpwatt *et al.*, 2006) et la protection contre les pathogènes (Madhaiyan *et al.*, 2004, 2006), participant ainsi au fonctionnement de l'écosystème, notamment au cycle du carbone (Yurimoto et Sakai, 2023). L'étude de son écologie et de son évolution est donc cruciale pour acquérir une compréhension holistique des dynamiques biologiques au sein des écosystèmes forestiers. Un des intérêts de mon mémoire réside dans la complémentarité des deux chapitres. Le premier s'intéresse à l'influence de l'espèce hôte sur le genre *Methylobacterium*, tandis qu'inversement le deuxième se focalise sur l'impact de la diversité de *Methylobacterium* sur la croissance de son hôte. De plus, le premier chapitre adopte une approche exploratrice et investigue les assemblages bactériens en nature, tandis que le deuxième est structuré autour d'hypothèses de recherche précises, testées en microcosmes en employant des communautés synthétiques. Un constat émerge des travaux complémentaires de mon mémoire : *Methylobacterium* et ses hôtes influencent mutuellement leurs trajectoires écologiques et évolutives.

Dans mon premier chapitre, en liant les gènes aux habitats, j'ai révélé pour la première fois une partie des bases génétiques responsables des adaptations de *Methylobacterium* à la surface foliaire de divers arbres et arbuste hôtes caractéristiques de la forêt tempérée nordique. De plus, j'ai pu examiner la contribution du pangénome accessoire aux adaptations à l'hôte, et ainsi évaluer qualitativement l'importance de la sélection adaptative dans le maintien de la diversité génétique au sein du pangénome de *Methylobacterium*. Ce faisant, j'ai démontré le rôle crucial que joue la biodiversité floristique dans l'écologie et l'évolution bactérienne. Enfin, j'ai élaboré deux hypothèses pouvant expliquer une grande partie de la variation taxonomique et génétique au sein des communautés bactériennes de la phyllosphère. J'espère que ces hypothèses serviront à guider la recherche future et qu'elles seront rigoureusement mises à l'épreuve afin de tenter de mieux comprendre les patrons de diversité microbienne observée dans nos forêts.

Dans le cadre de mon deuxième chapitre, j'ai révélé clairement la relation de cause à effet entre l'augmentation de la diversité bactérienne foliaire et l'augmentation de la croissance de la plante hôte. À ma connaissance, très peu d'études – voire aucune – ont démontré en conditions gnotobiotiques ce lien de causalité entre la diversité des symbiontes dans la phyllosphère et l'hôte. De plus, mes résultats ont apporté des nuances intéressantes au paradigme écologique classique liant biodiversité et fonctionnement de l'écosystème, notamment l'hypothèse de l'allègement de

la compétition par les interactions entre plusieurs espèces. Cette hypothèse devra néanmoins être testée expérimentalement, sur milieu de culture ou sur plante, entre souches de *Methylobacterium* ainsi qu'entre souches de lignées plus diversifiées. L'usage de communautés synthétiques pour tester des hypothèses écologiques est justement une avenue au potentiel élevé pour améliorer notre compréhension des dynamiques écosystémiques complexes impliquant les microorganismes (Vorholt *et al.*, 2017).

J'ai également découvert deux souches de *Methylobacterium* démontrant un fort potentiel pour accroître la croissance des végétaux, dont une appartenant potentiellement à une espèce clé de voute de la phyllosphère, *Methylobacterium sp. 018. M. sp. 018* était de loin l'espèce de *Methylobacterium* la plus abondante dans la phyllosphère de notre forêt d'étude, notamment sur les hôtes feuillus. Bien que la souche dont le génome a été choisi pour représenter *M. sp. 018* dans le pangénome, E-025, n'ait pas été utilisée dans les communautés synthétiques, une autre souche appartenant à l'espèce *M. sp. 018*, J-092, s'est avérée être la souche démontrant l'effet bénéfique le plus marqué sur la croissance de *A. thaliana*, autant en monoculture qu'en communautés diversifiées.

En conclusion, une meilleure compréhension des dynamiques écoévolutives entre bactéries et plantes ne peut que favoriser les avancées en recherche appliquée. Les changements climatiques appellent à accroître la résilience des forêts et des agroécosystèmes. Dans ce contexte, mes travaux, je l'espère, fourniront aux chercheurs en ingénierie du microbiome végétal des pistes pour optimiser la santé des arbres en pépinière, en plantation, ou en migration assistée (Argüelles-Moyao et Galicia, 2024; Busby *et al.*, 2022), et pour développer des approches agricoles plus écosystémiques, basées sur l'inoculation bactérienne, pour augmenter le rendement des cultures dans les champs (Bajpai *et al.*, 2022; de Souza *et al.*, 2020).

3.7 Pistes de recherche à explorer

Plusieurs hypothèses ont été formulées dans le cadre de mes travaux et devraient être spécifiquement mises à l'épreuve dans un cadre expérimental dédié. En ce qui concerne les hypothèses sur l'influence de la lumière du soleil et de la phénologie des feuilles pour expliquer la structure des communautés bactériennes, une plus grande diversité d'espèces d'arbres et d'arbustes devra être considérée afin de décortiquer les différents facteurs d'influence. Il serait intéressant d'inclure, par exemple, le conifère au feuillage caduc *Larix laricina*, l'arbuste conifèrien *Taxus canadensis* et des espèces angiospermes arbustives au feuillage persistent de la famille

des *Ericaceae*. De plus, une attention particulière devrait être portée à *C. cornuta* dans les études futures, pour examiner si l'espèce en soi possède des caractéristiques qui la distingue clairement de toutes les autres espèces arbustives. Au-delà des gènes, l'analyse du transcriptome (Scheublin *et al.*, 2014) ou du protéome (Delmotte *et al.*, 2009), en fournissant un portrait de l'expression des gènes et des protéines dans les cellules bactériennes vivantes, permettrait une analyse encore plus précise des structures cellulaires, des métabolismes et des fonctions clés associées aux différentes niches écologiques de la phyllosphère.

Peu d'études se sont penchées sur la phyllosphère des arbustes en forêt. Mon étude montre que leur rôle comme pont entre les microbiomes de la haute canopée et du sol mérite d'être investigué pour mieux comprendre les dynamiques de dispersion au sein des métacommunautés forestières, ainsi que les implications écologiques et évolutives de cette dispersion. Est-ce que les arbustes sont principalement des hôtes transitoires servant à la circulation entre compartiments écologiques? Sont-ils d'importants réservoirs (sources) de diversité taxonomique et génétique bactérienne, ou agissent-ils plutôt comme des puits? La migration des microorganismes du sol et de la rhizosphère – deux des compartiments écologiques les plus riches en espèces (Thompson *et al.*, 2017) – vers les feuilles a été démontrée expérimentalement (Chi *et al*., 2005; Massoni *et al*., 2021; Zarraonaindia *et al*., 2015), mais les impacts de cette migration n'ont été que peu abordés. D'autre part, une migration des microorganismes des strates supérieures de la canopée vers les arbustes pourrait avoir lieu par l'intermédiaire des précipitations (Bittar *et al.*, 2018), mais le rôle de la pluie dans le déplacement des bactéries et leur colonisation des strates inférieures est encore à clarifier (Stone et Jackson, 2019, 2021). Nonobstant la provenance des microorganismes, une immigration continuelle sur les feuilles des arbustes pourrait entraîner une plus grande diversité et une plus forte intensité d'interactions biotiques microbes à microbes, et ainsi venir moduler les pressions de sélection agissant chez les communautés et populations bactériennes de ces plantes hôtes. Certains de nos résultats suggèrent d'ailleurs une importance accrue des interactions biotiques chez *C. cornuta* en comparaison avec les espèces d'arbres hôtes*,* tels que la prévalence des gènes enrichis ou sous sélection positive impliqués dans le mobilome, la transcription, la défense, et la composition en protéines et lipides de la membrane cellulaire. Pour mieux comprendre l'écologie de la phyllosphère des arbustes, il apparaît donc primordial d'investiguer la migration entre compartiments écologiques en étudiant les changements spatio-temporels dans la composition taxonomique des microbiomes des différentes strates de la canopée, de la pluie, du sol et de la rhizosphère. Enfin, les feuilles des jeunes arbres sont potentiellement soumises aux mêmes dynamiques migratoires.

Conséquemment, pour distinguer l'effet du gradient vertical (et d'une immigration potentiellement accrue) de l'effet du type de plante (arbre ou arbuste), il serait intéressant d'observer comment se rapproche ou s'éloigne le microbiome d'une espèce d'arbre de celui d'une espèce d'arbuste évolutivement rapprochée (ex. *Betula alleghaniensis* et *C. cornuta*, deux espèces de la famille *Betulaceae*) au fil de sa croissance. Est-ce que le jeune arbre possède un microbiome foliaire typique des arbres (par exemple, riche en gènes conférant un meilleur *fitness* à grande exposition au rayonnement solaire), même si ses feuilles sont basses dans la canopée?

Puisque la forêt tempérée est principalement caractérisée par sa forte saisonnalité, que la phytochimie change au courant de la saison de croissance (Dorokhov *et al.*, 2018), et que les communautés bactériennes de la phyllosphère fluctuent dans le temps (Leducq *et al.*, 2022a; Redford et Fierer, 2009), il serait également intéressant d'analyser si certains gènes ou protéines sont différentiellement enrichies au sein des communautés entre les saisons.

Je recommande aussi vivement d'étudier plus en détail l'espèce *Methylobacterium sp. 018*, pour laquelle de nombreuses souches ont été isolées de la phyllosphère. Une analyse de son pangénome, une caractérisation biochimique des souches, et des expériences de coculture ou d'inoculation sur plantes permettraient de mieux comprendre le rôle joué par cette espèce et ses gènes au sein des communautés microbiennes et de ses associations symbiotiques avec les différentes espèces de plantes. En ce qui concerne le pangénome du genre *Methylobacterium*, il serait intéressant d'étudier en détail les mécanismes ayant façonné la structure du pangénome accessoire – transfert horizontal et provenance, duplication, perte de gènes –, particulièrement en ce qui concerne les gènes démontrant un potentiel adaptatif aux différents types d'hôtes. Pour les analyses plus fines de ces dynamiques écoévolutives au sein des génomes, il serait intéressant de se concentrer sur les espèces largement répandues qui occupent diverses niches et pour lesquelles plusieurs souches ont été isolées, plutôt que sur le pangénome du genre dans son entièreté.

La recherche sur l'écologie de la phyllosphère se focalise souvent sur les métabolismes du carbone et des nutriments, avec un accent sur l'étude des gènes et des protéines (Bashir *et al.*, 2022; Knief *et al.*, 2012a). Dans mon étude, plusieurs gènes impliqués dans les métabolismes des acides gras ont été identifiés parmi les gènes conférant potentiellement des adaptations à l'hôte. Une étude approfondie du lipidome des bactéries de la phyllosphère pourrait dévoiler des mécanismes clés d'adaptation aux différentes niches foliaires. D'autre part, il serait intéressant de

tenter d'établir le lien entre les éléments génétiques mobiles au sein des génomes de *Methylobacterium* et l'écologie de ses populations. J'ai révélé plusieurs gènes codant pour des transposases qui étaient soumis à la sélection positive en association avec différentes espèces hôtes. Les transposases servent à relocaliser les transposons (ou éléments transposables), un type d'élément génétique mobile, au sein d'un génome (Hickman et Dyda, 2015). Ces transposons peuvent jouer un rôle dans l'adaptation à des environnements nouveaux ou changeants (Casacuberta et González, 2013).

Finalement, mon approche expérimentale avec communautés synthétiques se démarque d'un grand nombre d'études portant sur la relation biodiversité-fonctionnement de l'écosystème en se penchant non pas sur la réponse d'un groupe d'organismes face à sa propre diversité (ex. l'influence de la diversité des plantes sur la productivité d'une prairie), mais plutôt sur la réponse d'une espèce hôte (une plante angiosperme) à la diversité d'un groupe de microorganismes symbiontes évolutivement très éloigné (les bactéries). Il en découle des patrons complexes dans la relation causale entre richesse bactérienne et croissance de l'hôte, qui mettent en lumière la pertinence de prendre en considération le microbiome foliaire dans l'étude des fonctions des écosystèmes terrestres. De plus – et surtout – mes travaux révèlent la nécessité d'étudier plus spécifiquement les dynamiques de compétition et de coexistence ainsi que le processus d'assemblage au sein des communautés bactériennes, à la fois sur l'hôte (*in vivo*) et à l'extérieur de l'hôte (*in vitro*). Il serait également fort important d'investiguer, dans ces deux contextes, l'effet de la diversité bactérienne sur le métabolome microbien et végétal, en mettant l'accent sur la production et la sécrétion de molécules antimicrobiennes et de phytohormones qui possiblement sous-tendent les variations observées. Ces connaissances permettraient à la fois d'isoler l'influence de l'hôte sur la structuration des communautés microbiennes, et d'identifier les mécanismes biochimiques faisant le pont entre les réponses microbienne et végétale à la diversité des communautés de bactéries foliaires. D'autre part, l'utilisation de l'organisme modèle *A. thaliana* est très pertinente dans le cadre de certaines expériences avec des communautés synthétiques, notamment lorsqu'on cherche à évaluer l'effet de son génotype sur ses symbiontes (Bodenhausen *et al.*, 2014). Or, du fait de sa taille, il peut-être difficile de cultiver de très nombreux individus simultanément, ce qui limite le nombre de traitements (et de réplicats) pouvant être administrés dans une expérience. L'utilisation de plantes hôtes alternatives, comme la petite lentille d'eau (*Lemna minor*), est une avenue prometteuse pour tester les interactions entre les microorganismes de la phyllosphère et leur hôte, comme l'effet d'allègement de la compétition par les interactions multiples, car il est possible de cultiver des milliers d'échantillons à la fois.

APPENDICE A

SUPPLEMENTARY MATERIAL – CHAPTER 1

Table A.1 Number of metagenomic read pairs (raw and quality filtered) obtained from sequencing, and number and percentage of reads that mapped to *the Methylobacterium pangenome*.

sample	raw pairs	filtered pairs	filtered pairs (%)	mapped reads	mapped reads (%)
ABBA_M11	77366671	75813593	97,99	649356	0,42825829
ABBA_M12	87141269	85384508	97,98	2278952	1,33452312
ABBA_M13	88780323	86880987	97,86	1806157	1,03944319
ABBA_M14	123574867	121167854	98,05	2718638	1,12184788
ABBA_M15	81172304	79497890	97,94	1169192	0,7353604
ACSA_M21	67857059	66571885	98,11	1248284	0,93754593
ACSA_M22	82728655	81098799	98,03	598867	0,36922063
ACSA_M23	54301257	53372573	98,29	275189	0,25780001
ACSA_M24	64832058	63664639	98,2	989884	0,77742057
ACSA_M25	83924385	82203757	97,95	1621315	0,98615627
COCO_M01	75942855	74412420	97,98	6193356	4,16150691
COCO_M02	76928927	75377338	97,98	5915567	3,92396916
COCO_M03	87544660	85697936	97,89	7938674	4,63177666
COCO_M04	84214348	82424196	97,87	1813444	1,10006775
COCO_M05	92352501	90309976	97,79	6246367	3,45829291
FAGR_M16	93003886	91137163	97,99	2251260	1,2350944
FAGR_M17	95570792	93755464	98,1	2420157	1,29067518
FAGR_M18	105350083	103332477	98,08	489432	0,2368239
FAGR_M19	90247775	88512584	98,08	1540790	0,87037906
FAGR_M20	72907395	71672261	98,31	1213583	0,84661973
THOC_M06	85461053	83664129	97,9	863490	0,51604553
THOC_M07	88637669	86818511	97,95	1669054	0,96123164
THOC_M08	84190622	82383296	97,85	1809544	1,09824691
THOC_M09	96083679	94188691	98,03	4635647	2,46082993
THOC_M10	86909907	85020898	97,83	3352641	1,97165702

ABBA, *Abies balsamea*; ACSA, *Acer saccharum*; COCO, *Corylus cornuta*; FAGR, *Fagus grandifolia*; THOC, *Thuja occidentalis*.

Table A.2 Pearson *r* coefficient was used to test for correlation between the number of reads mapped to *Methylobacterium*'s pangenome and the number of quality-filtered read pairs, for each of the host species.

host species	***r***
A. balsamea	0.7325609
T. occidentalis	0.6956963
A. saccharum	0.5913624
F. grandifolia	-0.3021538
C. cornuta	0.0258877

Table A.3 Comparison of *Methylobacterium* relative abundance in bacterial communities between *C. cornuta* and all four tree host species (Welch t-test), and between conifer and broadleaf trees (Student t-test).

comparison	***df***	***t***	***p***
shrub – trees	4.2772	4.6028	0.01706
conifer trees – broadleaf trees	18	1.376	0.37140

df, degrees of freedom; *t*, t-statistic; *p*, *p*–value (Bonferroni adjusted).

Table A.4 Comparison of bacterial genera community composition between *C. cornuta* and all four tree host species, and between conifer and broadleaf trees, using PERMANOVAs and multivariate homogeneity of variances tests (betadisper).

	PERMANOVA				betadisper		
comparison	***df***	***F***	***p****	***R^2***	***df***	***F***	***p***
shrub – trees	1	7.4635	0.001	0.245	1	0.2004	0.6586
conifer trees – broadleaf trees	1	5.4734	0.001	0.2332	1	8.0953	0.01074

df, degrees of freedom; *p*, *p*-value; * Bonferroni adjusted *p*-value.

Table A.5 Comparison of *Methylobacterium* species richness between *C. cornuta* and all four tree host species, and between conifer and broadleaf trees, using Welch t-tests.

comparison	***df***	***t***	***p***
shrub – trees	5.2691	3.2931	0.0401
conifer trees – broadleaf trees	15.006	1.3754	0.3784

df, degrees of freedom; *t*, t-statistic; *p*, p–value (Bonferroni adjusted).

Table A.6 Comparison of *Methylobacterium* species community composition between *C. cornuta* and all four tree host species, and between conifer and broadleaf trees, using PERMANOVAs and multivariate homogeneity of variances tests (betadisper).

	PERMANOVA				betadisper		
comparison	***df***	***F***	***p****	***R^2***	***df***	***F***	***p***
shrub – trees	1	5.2303	0.001	0.1921	1	3.8377	0.06291
conifer trees – broadleaf trees	1	2.5358	0.048	0.1298	1	0.1969	0.6628

df, degrees of freedom; *p*, p-value; * Bonferroni adjusted p-value.

Table A.7 Standardized effect size of mean pairwise distance (MPD) calculated for all communities (samples). The negative *z*-scores (mpd.obs.z) indicate that *Methylobacterium* communities were composed of phylogenetically closely related species.

sample	ntaxa	mpd.obs	mpd.rand.mean	mpd.rand.sd	mpd.obs.rank	mpd.obs.z	mpd.obs.p
ABBA_M11	3	0.1684	0.3603	0.0842	20	-2.2786	0.02
ABBA_M12	9	0.2525	0.3587	0.034	15	-3.12	0.015
ABBA_M13	5	0.2552	0.3618	0.0552	53	-1.9322	0.053
ABBA_M14	9	0.2454	0.3603	0.0339	7	-3.3915	0.007
ABBA_M15	6	0.1791	0.3608	0.0483	6	-3.7614	0.006
ACSA_M21	3	0.3208	0.3595	0.0886	285	-0.4361	0.285
ACSA_M22	2	0.3532	0.356	0.1374	405	-0.0206	0.405
ACSA_M24	5	0.2534	0.3589	0.0554	55	-1.9042	0.055
ACSA_M25	10	0.2159	0.36	0.0303	1	-4.7547	0.001
COCO_M01	14	0.308	0.3591	0.0206	19	-2.4838	0.019
COCO_M02	14	0.3179	0.3592	0.0205	32	-2.0208	0.032
COCO_M03	14	0.2415	0.3599	0.0213	1	-5.5689	0.001
COCO_M04	7	0.229	0.3607	0.0424	15	-3.1112	0.015
COCO_M05	19	0.3679	0.358	0.012	778	0.8299	0.778
FAGR_M16	11	0.2857	0.3598	0.0275	17	-2.6902	0.017
FAGR_M17	9	0.2465	0.3605	0.0342	9	-3.3303	0.009
FAGR_M18	1	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
FAGR_M19	6	0.2335	0.3592	0.0465	17	-2.7054	0.017
FAGR_M20	5	0.2552	0.3618	0.0552	53	-1.9322	0.053
THOC_M06	6	0.2446	0.3616	0.0463	32	-2.5267	0.032
THOC_M07	8	0.2591	0.3603	0.0382	22	-2.6465	0.022
THOC_M08	10	0.2467	0.3599	0.031	7	-3.6573	0.007
THOC_M09	12	0.2831	0.3586	0.026	10	-2.9034	0.01
THOC_M10	10	0.243	0.36	0.031	6	-3.7734	0.006

ntaxa, number of *Methylobacterium* species in sample; mpd.obs, observed mean pairwise distance (mpd) between species within a community; mpd.rand.mean, mean mpd in null communities; mpd.rand.sd, standard deviation of mpd in null communities; mpd.obs.rank, rank of observed mpd compared to null communities; mpd.obs.z, standardized effect size of mpd compared to null communities; mpd.obs.p, *p*-value of observed mpd compared to null communities.

Table A.8 Comparison of gene clusters composition of *Methylobacterium* communities between *C. cornuta* and all four tree host species, and between conifer and broadleaf trees, using PERMANOVAs and multivariate homogeneity of variances tests (betadisper).

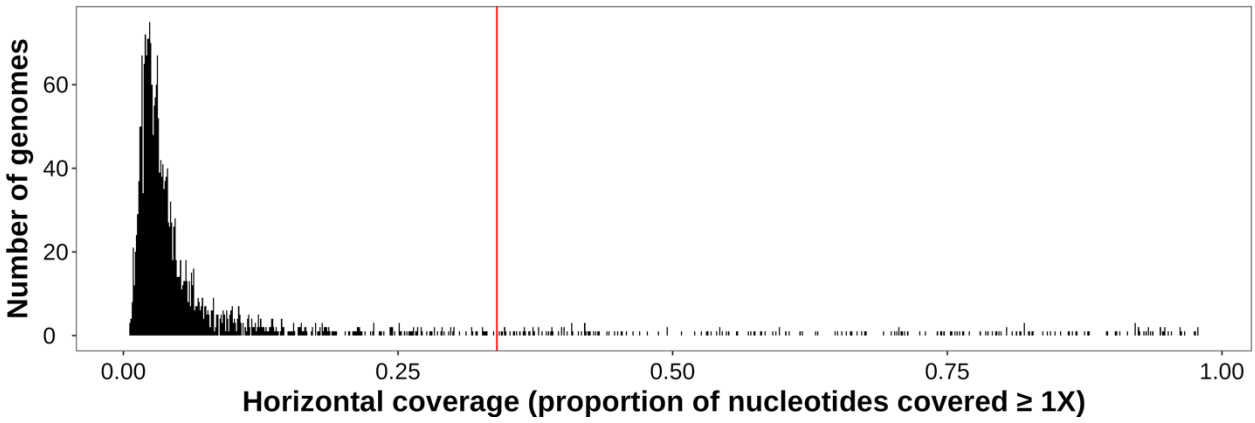
	PERMANOVA				betadisper		
comparison	***df***	***F***	***p****	R^2	***df***	***F***	***p***
shrub – trees	1	3.0322	0.004	0.12617	1	11.562	0.0027
conifer trees – broadleaf trees	1	2.0041	0.006	0.11131	1	0.7255	0.4069

df, degrees of freedom; *p*, *p*-value; * Bonferroni adjusted *p*-value.

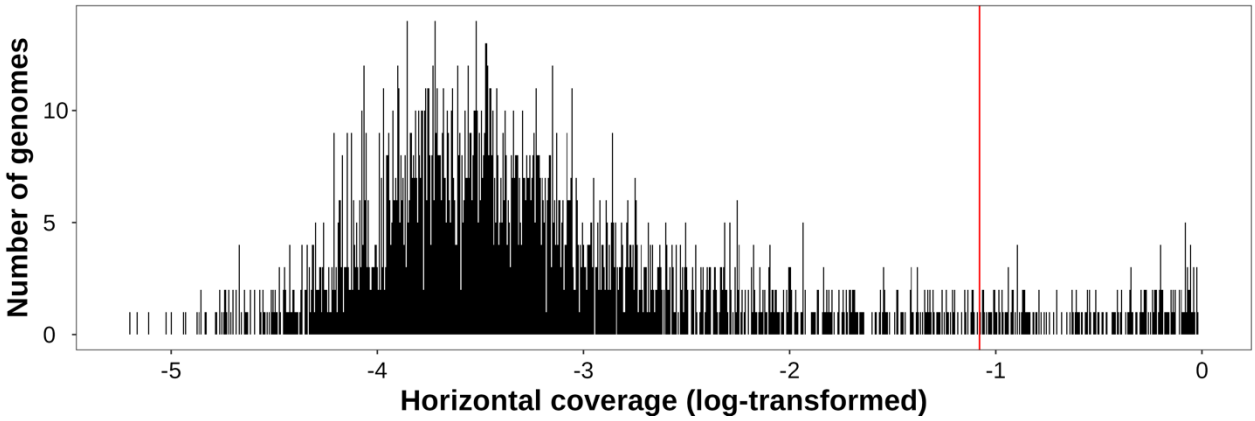
Table A.9 Post-hoc Chi–squared test for pairwise comparison of selection types between *Methylobacterium* species.

comparison	***df***	***χ2***	***p***
M. sp. 018 – *M. sp*. 021	2	546.4	< 0.001
M. sp. 018 – *M. sp.* 022	2	9.6804	0.023718
M. sp. 021 – *M. sp*. 022	2	105.71	< 0.001

df, degrees of freedom; *χ2*, Chi–squared test statistic;
p, *p*–value (Bonferroni adjusted).



(a)



(b)

Figure A.1 Distribution histograms of horizontal coverage values for all 104 genomes in all 25 samples (*n* = 2600). Log-normal distribution (a) and log-transformed normal distribution (b). Red lines represent values with a *z*-score of 1.96 calculated on the log-transformed normal distribution.

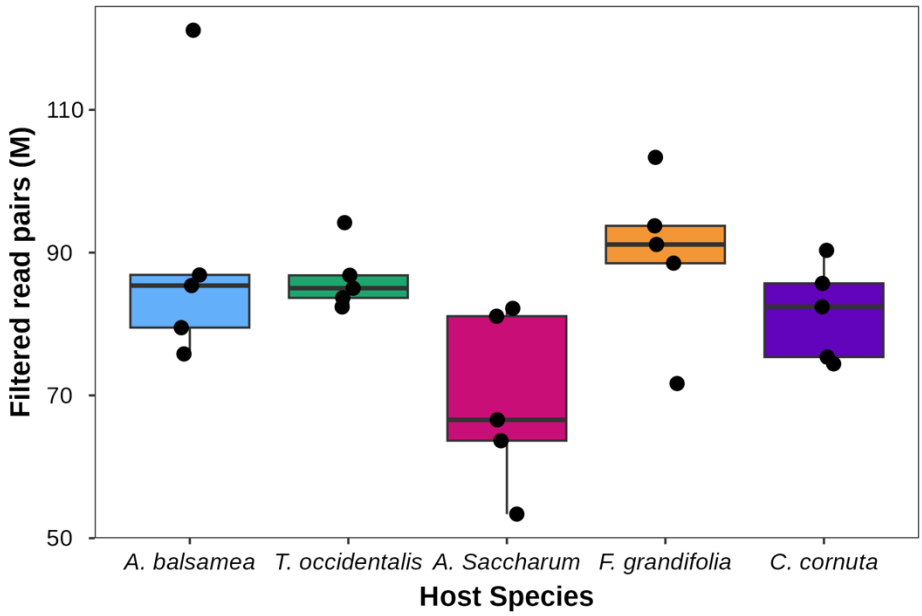


Figure A.2 The number of filtered read pairs (in millions) obtained from metagenomic shotgun sequencing did not significantly differ among the five host species (ANOVA, $p = 0.063$), although there was a tendency for *A. saccharum* samples to yield fewer reads.

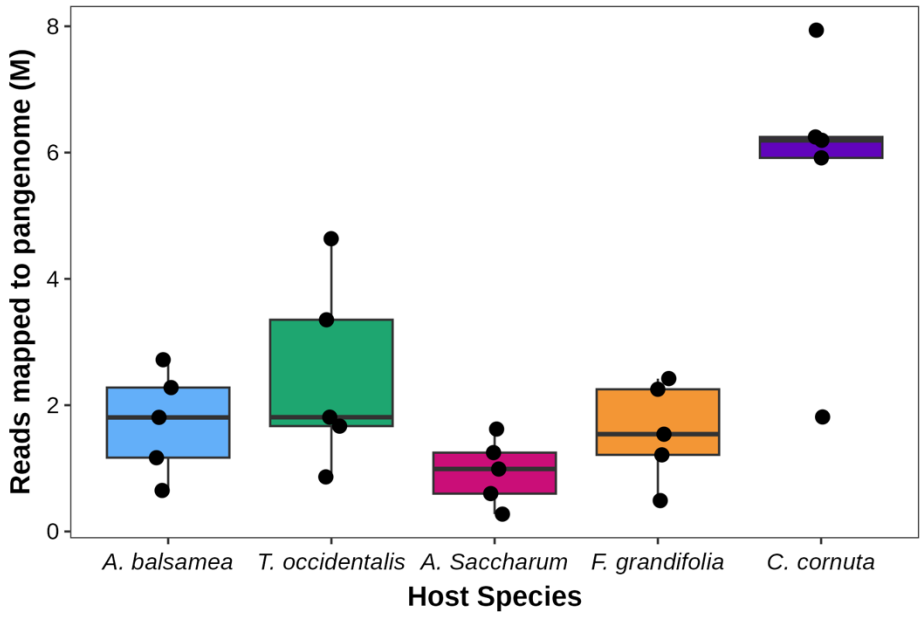


Figure A.3 A higher amount of filtered single reads mapped to the *Methylobacterium* pangenome in *C. cornuta* samples (Kruskal-Wallis, $p = 0.015$) than in samples from the other four host species.

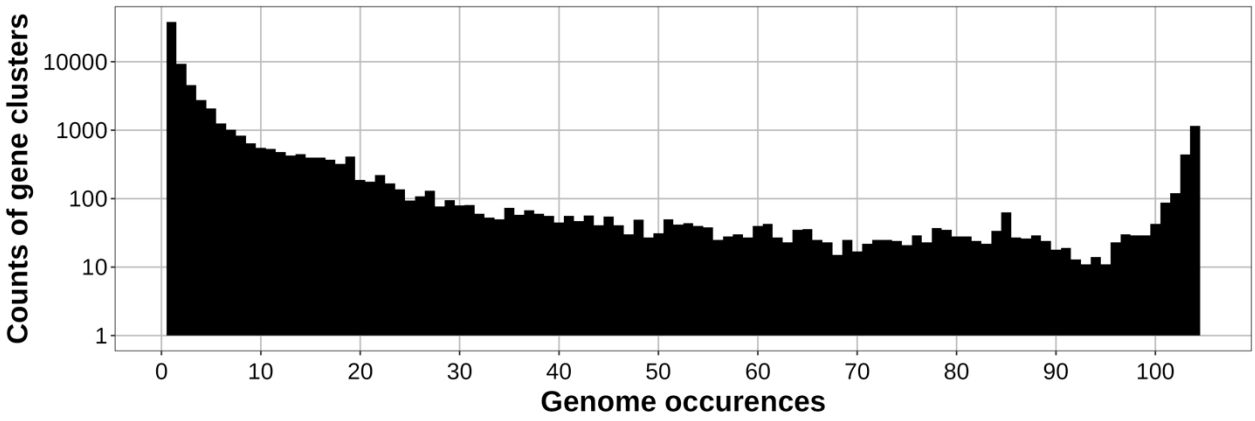


Figure A.4 Distribution of gene cluster counts by genome occurrences within the pangenome of *Methylobacterium*. This figure illustrates how many different gene clusters are present in a given number of genomes. For example, around 1000 gene clusters are found in all 104 genomes composing the pangenome.

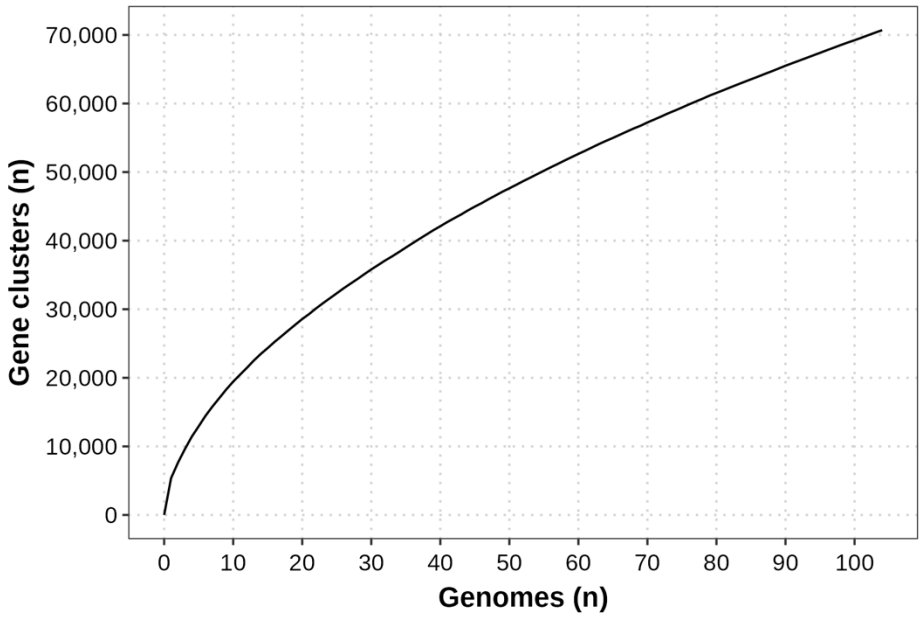


Figure A.5 *Methylobacterium* had an open pangenome: the number of new gene clusters increased continuously with increasing number of genomes considered in the pangenome, without reaching a plateau.The curve shows a power regression following Heaps law (Heap's law test, α = 0.376). We used the R package *micropan* (Snipen et Liland, 2015) to analyze the *Methylobacterium* pangenome openness. We used the function 'rarefaction' (100 permutations) to generate the rarefaction curve illustrating the cumulative number of distinct gene clusters in the pangenome with increasing genome number. We used the function 'heaps' (100 permutations) to test if the pangenome was open ($\alpha < 1$) or closed ($\alpha > 1$), according to Tettelin *et al.* (2008) who suggested that an open pangenome follows a Heaps' law. Heaps' law was originally formulated as the power law regression between the number of distinct words in a text with increasing text length (Heaps, 1978).

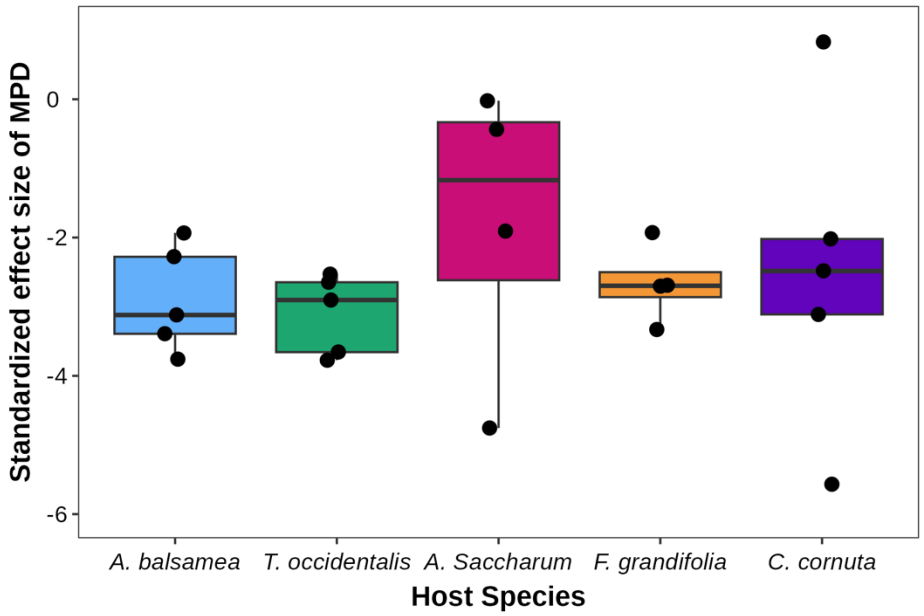


Figure A.6 The standardized effect size of mean pairwise distance (SES_{MPD}) – a measure of community phylogenetic dispersion expressed as *z*-scores – of *Methylobacterium* species within communities did not vary among the five host species (ANOVA, $p = 0.729$).

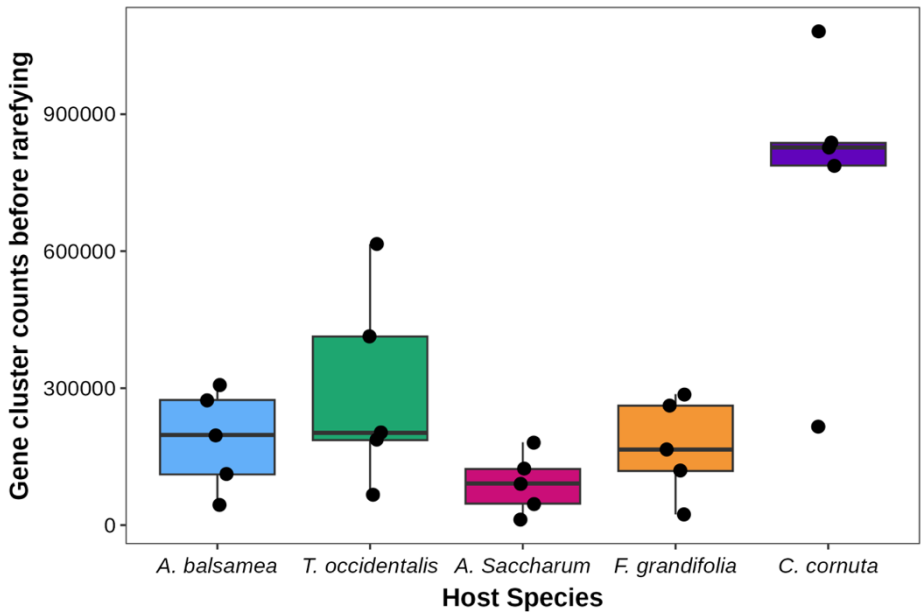


Figure A.7 Gene cluster counts before rarefying and filtering varied among host species (ANOVA, $p < 0{,}001$, adj. $R^2 = 0.591$), with *C. cornuta* samples containing a higher number of counts.

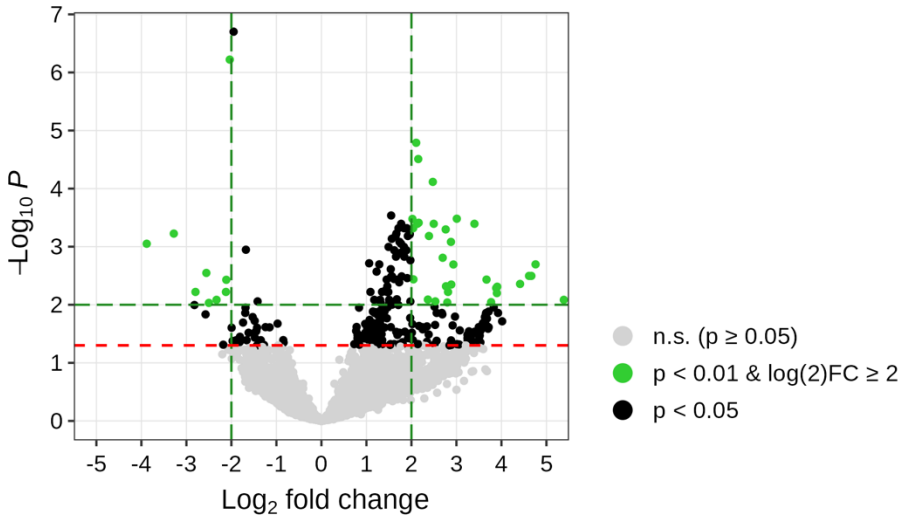


Figure A.8 Volcano plot highlighting the fraction of differentially abundant gene clusters (coloured dots) between tree host species and the shrub species *C. cornuta* that has a *p*-value < 0.01 and an absolute log(2) fold change ≥ 2 (*n* = analysis performed with DESeq2). Dashed green lines indicate the $p < 0.01$ & log(2)FC ≥ 2 threshold, and the red dashed line separates significant ($p < 0.05$) DAGCs (above, black or green) from nonsignificant DAGCs (below, gray). Positive log(2)FC indicates gene clusters associated with trees; negative log(2)FC indicates gene clusters associated with *C. cornuta*.

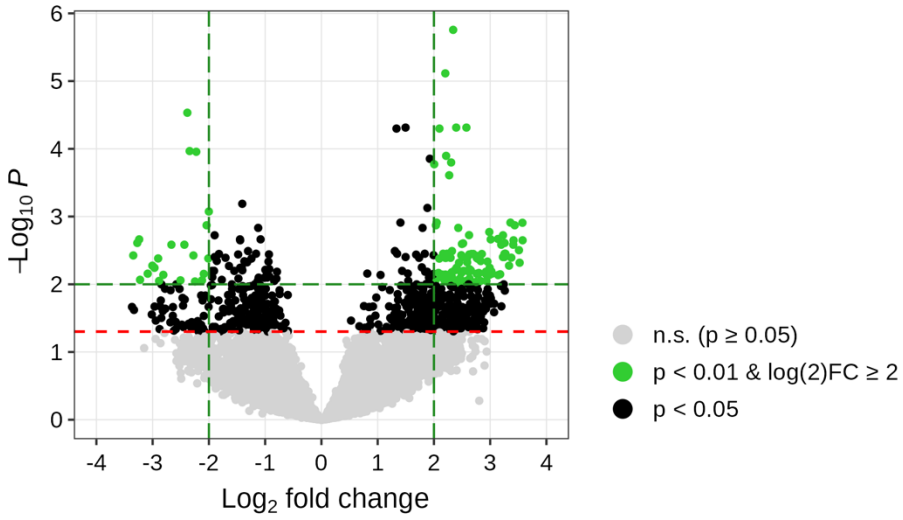


Figure A.9 Volcano plot highlighting the fraction of differentially abundant gene clusters (coloured dots) between broadleaf and conifer host species that has a *p*-value < 0.01 and an absolute log(2) fold change ≥ 2 (*n* = analysis performed with DESeq2). Dashed green lines indicate the $p < 0.01$ & log(2)FC ≥ 2 threshold, and the red dashed line separates significant ($p < 0.05$) DAGCs (above, black or green) from nonsignificant DAGCs (below, gray). Positive log(2)FC indicates gene clusters associated with broadleaves; negative log(2)FC indicates gene clusters associated with conifers.

APPENDICE B

SUPPLEMENTARY MATERIAL – CHAPTER 2

Table B.1 *Methylobacterium* strains (Leducq *et al.*, 2022a, 2022b) used in the experiment.

ID	strain	clade	forest	host sp.	sampling date	T°C iso.
A	E-046	A	MSH	ACSA	18-06-27	20
B	J-078	D	MSH	FAGR	18-10-18	30
C	J-088	A	SBL	ACSA	18-09-20	20
D	J-059	A	SBL	FAGR	18-07-16	20
E	J-043	B	MSH	FAGR	18-10-18	20
F	J-067	A	MSH	FAGR	18-08-06	30
G	J-048	A	SBL	FAGR	18-07-16	30
H	J-076	A	MSH	FAGR	18-10-18	30
I	E-045	D	MSH	ACSA	18-06-27	20
J	J-092	A	SBL	ACSA	18-07-16	20
K	E-005	A	MSH	FAGR	18-09-07	30
L	J-068	D	MSH	FAGR	18-08-06	30

ID: coding letters used in synthetic communities design and R script; forest: forest of isolation (MSH: Réserve Gault at Mont-Saint-Hilaire, QC, CAN; SBL: Station de Biologie des Laurentides, QC, CAN); host sp.: tree species from which the strain was isolated (ACSA: Acer saccharum; FAGR: Fagus grandifolia); sampling date: YY-MM-DD; T°C iso.: temperature at which the strain was isolated.

Table B.2 Synthetic community compositions and experimental design. 48 distinct community compositions were constructed based on 12 strains. Each letter (A to L) represent a *Methylobacterium* strain (Appendix B, Table B.1).

Strain richness	1	2	4	6	8	10
Community composition	A	CE	ABCD	ADFGHL	ADFGHIKL	ADEFGHIJKL
	B	BD	EFGH	CDEGHJ	ABDEFHKL	ABCDEFIJKL
	C	GJ	IJKL	BCEIKJ	ABCFGIJL	BCDEGHIJKL
	D	HL	ADGJ	ABEHIK	ACDEGHIJ	ABCFGHIJKL
	E	AK	BEHK	CDFGKL	BCDEHIJK	ABCDEFGHIK
	F	FI	CFIL	ABFIJL	BCEFGJKL	ABCDEFGHJL
	G	AL				
	H	BK				
	I	CJ				
	J	DI				
	K	EH				
	L	FG				
n	12	12	6	6	6	6

Table B.3 Saturated model equations used for model exploration in the context of our 2nd hypothesis (a) and 3rd hypothesis (b).

ID	Saturated global model
a	dry leaf biomass ~ E-046 + J-078 + J-088 + J-059 + J-043 + J-067 + J-048 + J-076 + E-045 + J-092 + E-005 + J-068
b	dry leaf biomass ~ strain richness + J-067 + E-045 + J-092 + strain richness:J-067 + strain richness:E-045 + strain richness:J-092 + J-067:E-045 + J-067:J-092 + E-045:J-092 + strain richness:J-067:E-045 + strain richness:J-067:J-092 + strain richness:E-045:J-092 + J-067:E-045:J-092 + strain richness:J-067:E-045:J-092.

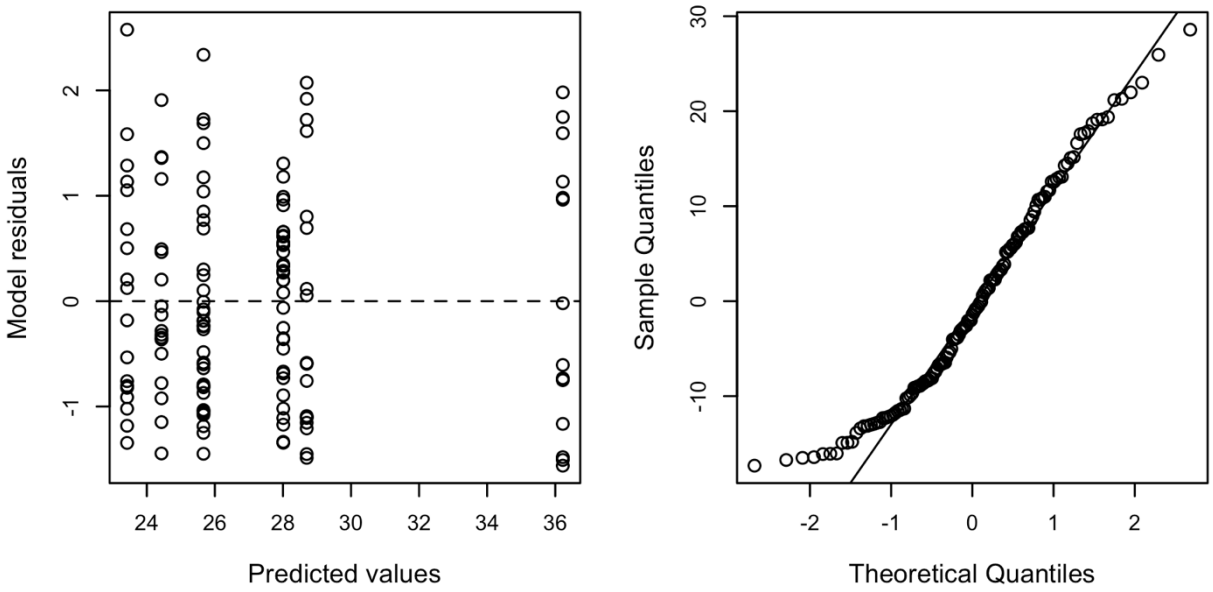


Figure B.1 The assumptions of homoscedasticity and normality of residuals for the second-degree polynomial regression model were met (Model ID 1, Table 2.1). The model can be used to draw inferences on the effect of *Methylobacterium* strain richness at the moment of inoculation on *A. thaliana* dry leaf biomass at the time of harvest. Left panel shows the homogeneity in the variance of scaled residuals across the response variable values. Right panel compares the model's residuals quantiles (*y*-axis) to the quantiles of a normal distribution (*x*-axis).

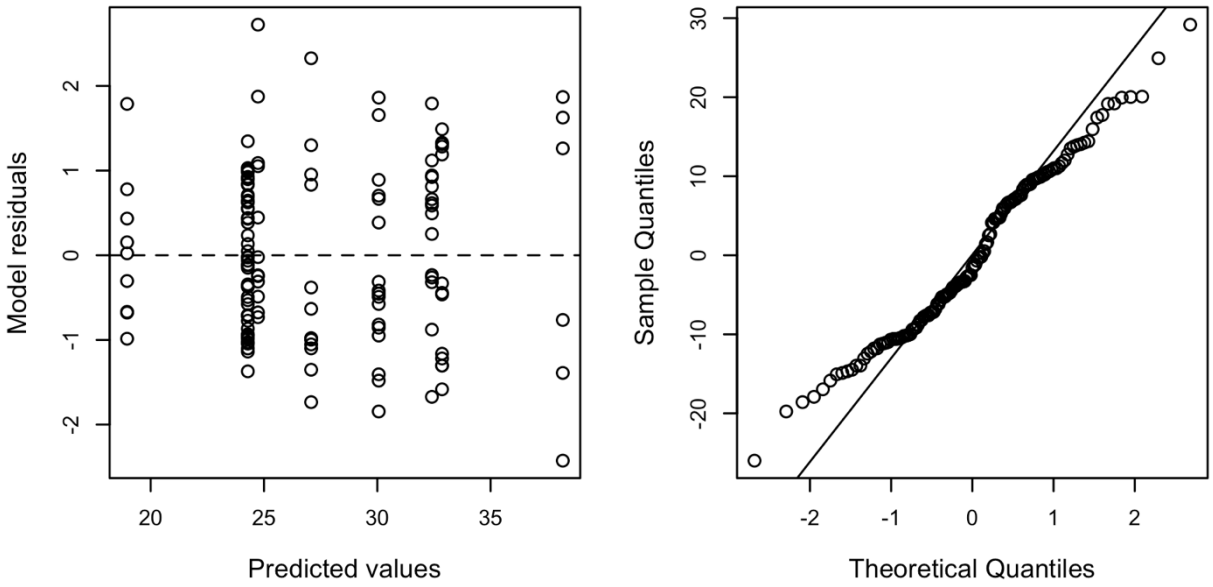


Figure B.2 The assumptions of homoscedasticity and normality of residuals were met for the model ID 1 of Table 2.2 (leaf biomass ~ J-067 + E-045 + J-092). This model could be used to draw inferences on the effect of the included *Methylobacterium* strains on *A. thaliana* dry leaf biomass at the time of harvest. Left panel shows the homogeneity in the variance of scaled residuals across the response variable values. Right panel compares the model's residuals quantiles (*y*-axis) to the quantiles of a normal distribution (*x*-axis).

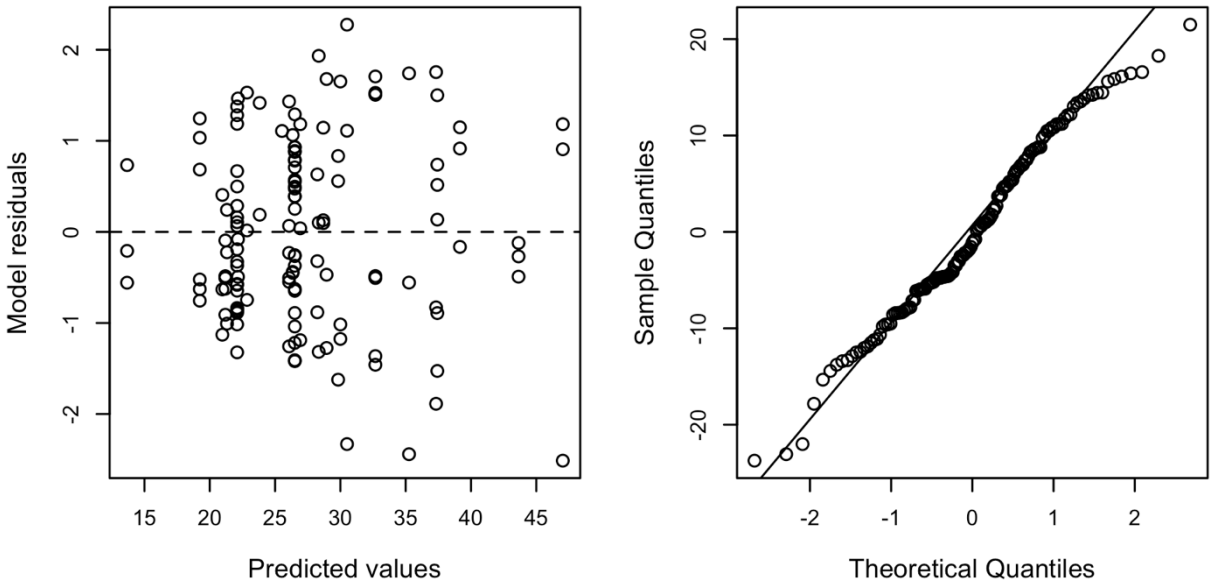


Figure B.3 The assumptions of homoscedasticity and normality of residuals were met for the averaged model including polynomial strain richness, J-067, E-045, J-092, and some of their interactions (Table 2.3). This averaged model can be used to draw inferences on the effect of these three *Methylobacterium* strains and their interactions, on the effect of inoculated strain richness on *A. thaliana* dry leaf biomass at the time of harvest. Left panel shows the homogeneity in the variance of scaled residuals across the response variable values. Right panel compares the model's residuals quantiles (*y*-axis) to the quantiles of a normal distribution (*x*-axis).

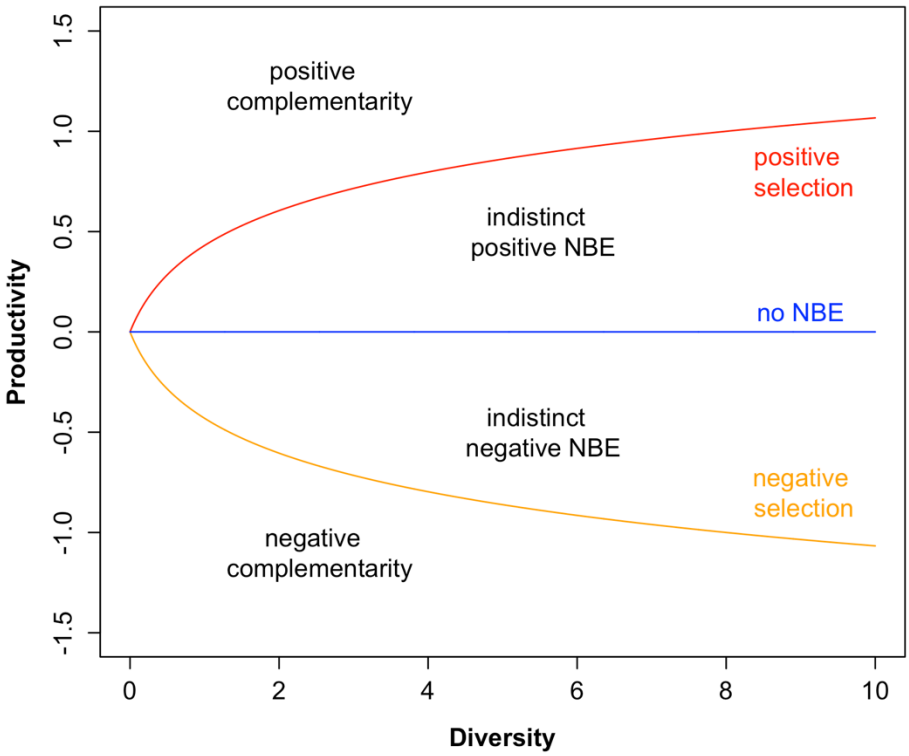


Figure B.4 Conceptual framework to test whether complementarity or selection drives the diversity-productivity relationship. Predictions of a model built on observed data that would closely follow the positive selection (red) or the negative selection (orange) models predictions would be a strong argument in favour of positive or negative selection respectively. Measured biomass values higher than those predicted by the positive selection model would represent positive complementarity, while values under those predicted by the negative selection model would represent negative complementarity. Values between the mean and the positive selection model, and those between the mean and the negative selection model, would respectively represent indistinct positive or negative biodiversity effect. Values following the no diversity effect model (blue) would imply no net biodiversity effect (NBE).

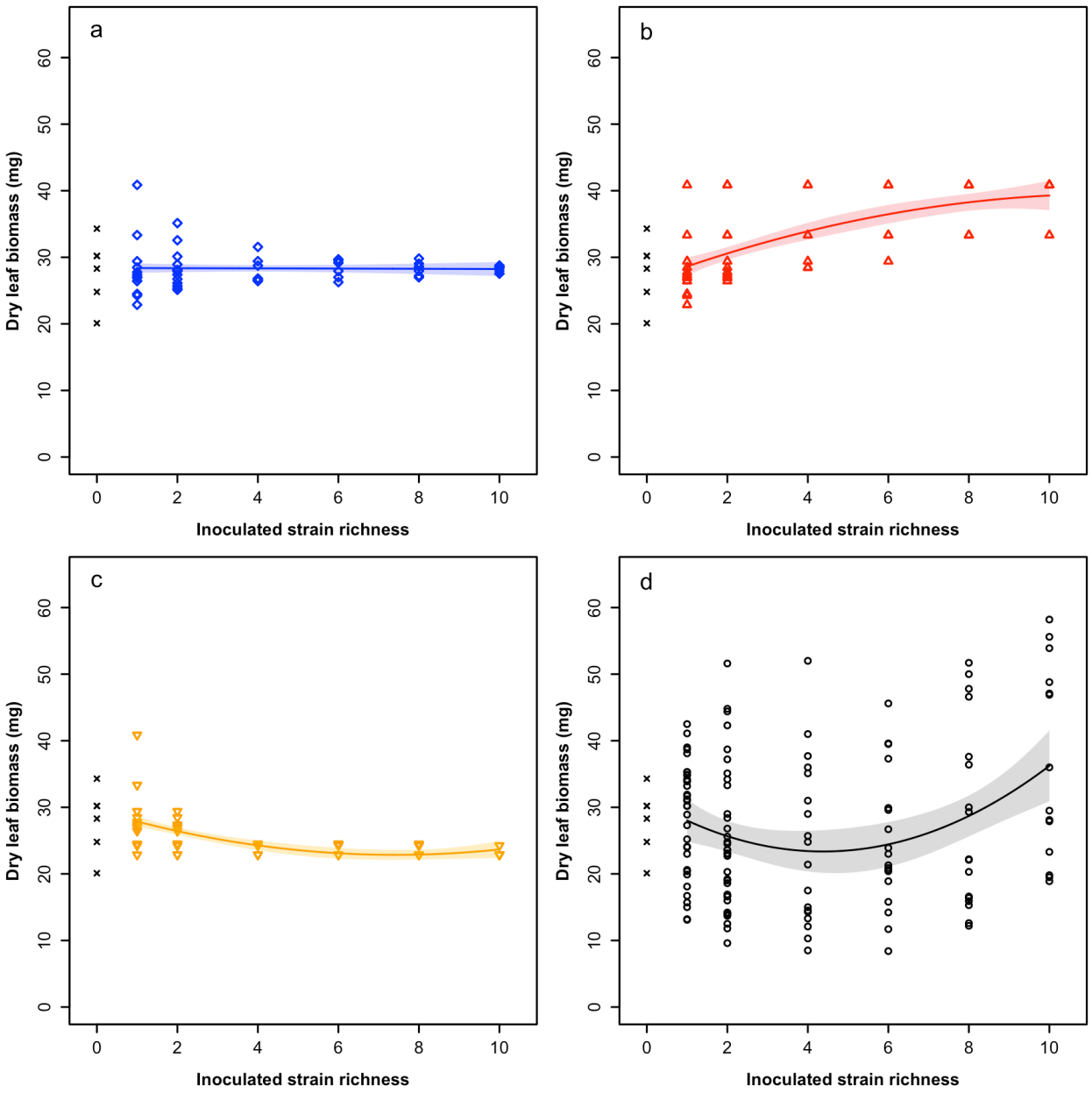


Figure B.5 (a) Linear regression "no diversity effect model" ($y = -0.02x + 28.39$) representing the leaf biomass response to inoculated strain richness assuming that, for every mixed bacterial community composed of equally abundant inoculated strains, all strains' effect on biomass corresponds to their mean effect in monoculture. (b) Second-degree polynomial regression "positive selection model" ($y = -0.10x^2 + 2.24x + 26.47$) representing the biomass response to inoculated richness assuming that the effect of mixed communities on biomass corresponds to the effect of the occurring strain that has the highest mean effect in monoculture. (c) Second-degree polynomial regression "negative selection model" ($y = 0.12x^2 - 1.80x + 29.55$) representing the biomass response to inoculated richness assuming that the effect of mixed communities on biomass corresponds to the effect of the occurring strain that has the lowest mean effect in monoculture. (d) Second-degree polynomial regression model fitted on observed data ($y = 0.41x^2 - 3.56x + 31.17$). In all panels, solid lines indicate fitted values of the models for the range of richness values used in the experiment. Shaded areas indicate 95% confidence intervals. n =

137. “x” data points represent control plants; those samples were not used in the models but were included as a visual assessment of the effect of *Methylobacterium* inoculation on host growth.

RÉFÉRENCES

Abanda-Nkpwatt, D., Müsch, M., Tschiersch, J., Boettner, M. et Schwab, W. (2006). Molecular interaction between Methylobacterium extorquens and seedlings: growth promotion, methanol consumption, and localization of the methanol emission site. *Journal of Experimental Botany, 57*(15), 4025-4032. https://doi.org/10.1093/jxb/erl173

Agler, M. T., Ruhe, J., Kroll, S., Morhenn, C., Kim, S.-T., Weigel, D. et Kemen, E. M. (2016). Microbial Hub Taxa Link Host and Abiotic Factors to Plant Microbiome Variation. *PLOS Biology, 14*(1), e1002352. https://doi.org/10.1371/journal.pbio.1002352

Aktas, M., Wessel, M., Hacker, S., Klüsener, S., Gleichenhagen, J. et Narberhaus, F. (2010). Phosphatidylcholine biosynthesis and its significance in bacteria interacting with eukaryotic cells. *European Journal of Cell Biology, 89*(12), 888-894. https://doi.org/10.1016/j.ejcb.2010.06.013

Alessa, O., Ogura, Y., Fujitani, Y., Takami, H., Hayashi, T., Sahin, N. et Tani, A. (2021). Comprehensive Comparative Genomics and Phenotyping of Methylobacterium Species. *Frontiers in Microbiology, 12*, 740610. https://doi.org/10.3389/fmicb.2021.740610

Alneberg, J., Bennke, C., Beier, S., Bunse, C., Quince, C., Ininbergs, K., Riemann, L., Ekman, M., Jürgens, K., Labrenz, M., Pinhassi, J. et Andersson, A. F. (2020). Ecosystem-wide metagenomic binning enables prediction of ecological niches from genomes. *Communications Biology, 3*(1), 1-10. https://doi.org/10.1038/s42003-020-0856-x

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology, 26*(1), 32-46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x

Anderson, M. J. (2006). Distance-Based Tests for Homogeneity of Multivariate Dispersions. *Biometrics, 62*(1), 245-253. https://doi.org/10.1111/j.1541-0420.2005.00440.x

Anderson, M. J., Ellingsen, K. E. et McArdle, B. H. (2006). Multivariate dispersion as a measure of beta diversity. *Ecology Letters, 9*(6), 683-693. https://doi.org/10.1111/j.1461-0248.2006.00926.x

Andersson, D. I. et Hughes, D. (2009). Gene Amplification and Adaptive Evolution in Bacteria. *Annual Review of Genetics, 43*, 167-195. https://doi.org/10.1146/annurev-genet-102108-134805

Andreani, N. A., Hesse, E. et Vos, M. (2017). Prokaryote genome fluidity is dependent on effective population size. *The ISME Journal, 11*(7), 1719-1721. https://doi.org/10.1038/ismej.2017.36

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S. et Ogata, H. (2020). KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics, 36*(7), 2251-2252. https://doi.org/10.1093/bioinformatics/btz859

Araya, M. A., Valenzuela, T., Inostroza, N. G., Maruyama, F., Jorquera, M. A. et Acuña, J. J. (2020). Isolation and Characterization of Cold-Tolerant Hyper-ACC-Degrading Bacteria from the Rhizosphere, Endosphere, and Phyllosphere of Antarctic Vascular Plants. *Microorganisms, 8*(11), 1788. https://doi.org/10.3390/microorganisms8111788

Ares-Arroyo, M., Coluzzi, C., Sousa, J. A. M. de et Rocha, E. P. C. (2024). Hijackers, hitchhikers, or co-drivers? The mysteries of mobilizable genetic elements. *PLOS Biology, 22*(8), e3002796. https://doi.org/10.1371/journal.pbio.3002796

Argüelles-Moyao, A. et Galicia, L. (2024). Assisted migration and plant invasion: importance of belowground ecology in conifer forest tree ecosystems. *Canadian Journal of Forest Research, 54*(1), 110-121. https://doi.org/10.1139/cjfr-2023-0016

Armitage, D. W. (2016). Time-variant species pools shape competitive dynamics and biodiversity–ecosystem function relationships. *Proceedings of the Royal Society B, 283*(1838), 20161437. https://doi.org/10.1098/rspb.2016.1437

Aydogan, E. L., Budich, O., Hardt, M., Choi, Y. H., Jansen-Willems, A. B., Moser, G., Müller, C., Kämpfer, P. et Glaeser, S. P. (2020). Global warming shifts the composition of the abundant bacterial phyllosphere microbiota as indicated by a cultivation-dependent and -independent study of the grassland phyllosphere of a long-term warming field experiment. *FEMS Microbiology Ecology, 96*(8), fiaa087. https://doi.org/10.1093/femsec/fiaa087

Bach, E., Rangel, C. P., Ribeiro, I. D. A. et Passaglia, L. M. P. (2022). Pangenome analyses of Bacillus pumilus, Bacillus safensis, and Priestia megaterium exploring the plant-associated features of bacilli strains isolated from canola. *Molecular Genetics and Genomics, 297*(4), 1063-1079. https://doi.org/10.1007/s00438-022-01907-0

Bajpai, A., Mahawar, H., Dubey, G., Atoliya, N., Parmar, R., Devi, M. H., Kollah, B. et Mohanty, S. R. (2022). Prospect of pink pigmented facultative methylotrophs in mitigating abiotic stress and climate change. *Journal of Basic Microbiology, 62*(8), 889-899. https://doi.org/10.1002/jobm.202200087

Baker, C. M., Chitrakar, R., Obulareddy, N., Panchal, S., Williams, P. et Melotto, M. (2010). Molecular battles between plant and pathogenic bacteria in the phyllosphere*. Brazilian Journal of Medical and Biological Research, 43*, 698-704. https://doi.org/10.1590/S0100-879X2010007500060

Baltrus, D. A. (2013). Exploring the costs of horizontal gene transfer. *Trends in Ecology & Evolution, 28*(8), 489-495. https://doi.org/10.1016/j.tree.2013.04.002

Banerjee, S., Schlaeppi, K. et van der Heijden, M. G. A. (2018). Keystone taxa as drivers of microbiome structure and functioning. *Nature Reviews Microbiology, 16*(9), 567-576. https://doi.org/10.1038/s41579-018-0024-1

Bartoń, K. (2023). *MuMIn: Multi-Model Inference* (v1.47.5). https://CRAN.R-project.org/package=MuMIn

Baselga, A. (2013). Separating the two components of abundance-based dissimilarity: balanced changes in abundance vs. abundance gradients. *Methods in Ecology and Evolution, 4*(6), 552-557. https://doi.org/10.1111/2041-210X.12029

Bashir, I., War, A. F., Rafiq, I., Reshi, Z. A., Rashid, I. et Shouche, Y. S. (2022). Phyllosphere microbiome: Diversity and functions. *Microbiological Research, 254*, 126888. https://doi.org/10.1016/j.micres.2021.126888

Batool, F., Rehman, Y. et Hasnain, S. (2016). Phylloplane associated plant bacteria of commercially superior wheat varieties exhibit superior plant growth promoting abilities. *Frontiers in Life Science, 9*(4), 313-322. https://doi.org/10.1080/21553769.2016.1256842

Beattie, G. A. (2011). Water Relations in the Interaction of Foliar Bacterial Pathogens with Plants. *Annual Review of Phytopathology, 49*(1), 533-555. https://doi.org/10.1146/annurev-phyto-073009-114436

Beatty, J. T., Overmann, J., Lince, M. T., Manske, A. K., Lang, A. S., Blankenship, R. E., Van Dover, C. L., Martinson, T. A. et Plumley, F. G. (2005). An obligately photosynthetic bacterial anaerobe from a deep-sea hydrothermal vent. *Proceedings of the National Academy of Sciences, 102*(26), 9306-9310. https://doi.org/10.1073/pnas.0503674102

Becker, J., Eisenhauer, N., Scheu, S. et Jousset, A. (2012). Increasing antagonistic interactions cause bacterial communities to collapse at high diversity. *Ecology Letters, 15*(5), 468-474. https://doi.org/10.1111/j.1461-0248.2012.01759.x

Bell, T., Newman, J. A., Silverman, B. W., Turner, S. L. et Lilley, A. K. (2005). The contribution of species richness and composition to bacterial services. *Nature, 436*(7054), 1157-1160. https://doi.org/10.1038/nature03891

Bernhard, A. E. et Kelly, J. J. (2016). Editorial: Linking Ecosystem Function to Microbial Diversity. *Frontiers in Microbiology, 7*. https://doi.org/10.3389/fmicb.2016.01041

Bittar, T. B., Pound, P., Whitetree, A., Moore, L. D. et Van Stan II, J. T. (2018). Estimation of Throughfall and Stemflow Bacterial Flux in a Subtropical Oak-Cedar Forest. *Geophysical Research Letters, 45*(3), 1410-1418. https://doi.org/10.1002/2017GL075827

Bobay, L.-M. et Ochman, H. (2018). Factors driving effective population size and pan-genome evolution in bacteria. *BMC Evolutionary Biology, 18*, 153. https://doi.org/10.1186/s12862-018-1272-4

Bodenhausen, N., Bortfeld-Miller, M., Ackermann, M. et Vorholt, J. A. (2014). A Synthetic Community Approach Reveals Plant Genotypes Affecting the Phyllosphere Microbiota. *PLOS Genetics, 10*(4), e1004283. https://doi.org/10.1371/journal.pgen.1004283

Boeuf, D., Eppley, J. M., Mende, D. R., Malmstrom, R. R., Woyke, T. et DeLong, E. F. (2021). Metapangenomics reveals depth-dependent shifts in metabolic potential for the ubiquitous marine bacterial SAR324 lineage. *Microbiome, 9*(1), 172. https://doi.org/10.1186/s40168-021-01119-5

Boucher, Y., Cordero, O. X., Takemura, A., Hunt, D. E., Schliep, K., Bapteste, E., Lopez, P., Tarr, C. L. et Polz, M. F. (2011). Local Mobile Gene Pools Rapidly Cross Species Boundaries To Create Endemicity within Global Vibrio cholerae Populations. *mBio, 2*(2), e00335-10. https://doi.org/10.1128/mbio.00335-10

Bousfield, I. J. et Green, P. N. (1985). Reclassification of Bacteria of the Genus Protomonas Urakami and Komagata 1984 in the Genus Methylobacterium (Patt, Cole, and Hanson) Emend. Green and Bousfield 1983. *International Journal of Systematic Bacteriology, 35*(2), 209-209. https://doi.org/10.1099/00207713-35-2-209

Boyes, D. C., Zayed, A. M., Ascenzi, R., McCaskill, A. J., Hoffman, N. E., Davis, K. R. et Görlach, J. (2001). Growth Stage–Based Phenotypic Analysis of Arabidopsis: A Model for High Throughput Functional Genomics in Plants. *The Plant Cell, 13*, 1499-1510. https://doi.org/10.1105/TPC.010011

Brandl, M., Clark, E. M. et Lindow, S. E. (1996). Characterization of the indole-3-acetic acid (IAA) biosynthetic pathway in an epiphytic strain of Erwinia herbicola and IAA production in vitro. *Canadian Journal of Microbiology, 42*(6), 586-592. https://doi.org/10.1139/m96-079

Breitwieser, F. P., Lu, J. et Salzberg, S. L. (2019). A review of methods and databases for metagenomic classification and assembly. *Briefings in Bioinformatics, 20*(4), 1125-1136. https://doi.org/10.1093/bib/bbx120

Bringel, F. et Couée, I. (2015). Pivotal roles of phyllosphere microorganisms at the interface between plant functioning and atmospheric trace gas dynamics. *Frontiers in Microbiology, 6*, 486. https://doi.org/10.3389/fmicb.2015.00486

Brockhurst, M. A., Chapman, T., King, K. C., Mank, J. E., Paterson, S. et Hurst, G. D. D. (2014). Running with the Red Queen: the role of biotic conflicts in evolution. *Proceedings of the Royal Society B: Biological Sciences, 281*(1797), 20141382. https://doi.org/10.1098/rspb.2014.1382

Brockhurst, M. A., Harrison, E., Hall, J. P. J., Richards, T., McNally, A. et MacLean, C. (2019). The Ecology and Evolution of Pangenomes. *Current Biology, 29*(20), R1094-R1103. https://doi.org/10.1016/j.cub.2019.08.012

Brooker, R. W., George, T. S., Homulle, Z., Karley, A. J., Newton, A. C., Pakeman, R. J. et Schöb, C. (2021). Facilitation and biodiversity–ecosystem function relationships in crop production systems and their role in sustainable farming. *Journal of Ecology, 109*(5), 2054-2067. https://doi.org/10.1111/1365-2745.13592

Buchfink, B., Xie, C. et Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods, 12*(1), 59-60. https://doi.org/10.1038/nmeth.3176

Buck, M., Mehrshad, M. et Bertilsson, S. (2022). mOTUpan: a robust Bayesian approach to leverage metagenome-assembled genomes for core-genome estimation. *NAR Genomics and Bioinformatics, 4*(3), lqac060. https://doi.org/10.1093/nargab/lqac060

Busby, P. E., Newcombe, G., Neat, A. S. et Averill, C. (2022). Facilitating Reforestation Through the Plant Microbiome: Perspectives from the Phyllosphere. *Annual Review of Phytopathology, 60*(1), 337-356. https://doi.org/10.1146/annurev-phyto-021320-010717

Cao, Y.-R., Wang, Q., Jin, R.-X., Tang, S.-K., Jiang, Y., He, W.-X., Lai, H.-X., Xu, L.-H. et Jiang, C.-L. (2011). Methylobacterium soli sp. nov. a methanol-utilizing bacterium isolated from the forest soil. *Antonie van Leeuwenhoek, 99*(3), 629-634. https://doi.org/10.1007/s10482-010-9535-0

Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., Narwani, A., Mace, G. M., Tilman, D., Wardle, D. A., Kinzig, A. P., Daily, G. C., Loreau, M., Grace, J. B., Larigauderie, A., Srivastava, D. S. et Naeem, S. (2012). Biodiversity loss and its impact on humanity. *Nature, 486*(7401), 59-67. https://doi.org/10.1038/nature11148

Carlström, C. I., Field, C. M., Bortfeld-Miller, M., Müller, B., Sunagawa, S. et Vorholt, J. A. (2019). Synthetic microbiota reveal priority effects and keystone strains in the Arabidopsis phyllosphere. *Nature Ecology & Evolution, 3*(10), 1445-1454. https://doi.org/10.1038/s41559-019-0994-z

Carr, V. R., Shkoporov, A., Hill, C., Mullany, P. et Moyes, D. L. (2021). Probing the Mobilome: Discoveries in the Dynamic Microbiome. *Trends in Microbiology, 29*(2), 158-170. https://doi.org/10.1016/j.tim.2020.05.003

Carrión, O., McGenity, T. J. et Murrell, J. C. (2020). Molecular Ecology of Isoprene-Degrading Bacteria. *Microorganisms, 8*(7), 967. https://doi.org/10.3390/microorganisms8070967

Carvalho, S. D. et Castillo, J. A. (2018). Influence of Light on Plant–Phyllosphere Interaction. *Frontiers in Plant Science, 9*, 1482. https://doi.org/10.3389/fpls.2018.01482

Casacuberta, E. et González, J. (2013). The impact of transposable elements in environmental adaptation. *Molecular Ecology, 22*(6), 1503-1517. https://doi.org/10.1111/mec.12170

Cavender-Bares, J., Kozak, K. H., Fine, P. V. A. et Kembel, S. W. (2009). The merging of community ecology and phylogenetic biology. *Ecology Letters, 12*(7), 693-715. https://doi.org/10.1111/j.1461-0248.2009.01314.x

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics, 10*(3), 195-205. https://doi.org/10.1038/nrg2526

Chen, W.-M., Cai, C.-Y., Li, Z.-H., Young, C.-C. et Sheu, S.-Y. (2019). Methylobacterium oryzihabitans sp. nov., isolated from water sampled from a rice paddy field. *International Journal of Systematic and Evolutionary Microbiology, 69*(12), 3843-3850. https://doi.org/10.1099/ijsem.0.003693

Chew, A. G. M. et Bryant, D. A. (2007). Chlorophyll Biosynthesis in Bacteria: The Origins of Structural and Functional Diversity. *Annual Review of Microbiology, 61*, 113-129. https://doi.org/10.1146/annurev.micro.61.080706.093242

Chi, F., Shen, S.-H., Cheng, H.-P., Jing, Y.-X., Yanni, Y. G. et Dazzo, F. B. (2005). Ascending Migration of Endophytic Rhizobia, from Roots to Leaves, inside Rice Plants and Assessment of Benefits to Rice Growth Physiology. *Applied and Environmental Microbiology, 71*(11), 7271-7278. https://doi.org/10.1128/AEM.71.11.7271-7278.2005

Cho, A., Cho, Y.-J., Kim, S. et Kim, O.-S. (2020). Draft Genome Sequence of the Psychrotolerant Bacterium Methylobacterium sp. Strain BTF04, Isolated from Freshwater in Antarctica. *Microbiology Resource Announcements, 9*(21), e00171-20. https://doi.org/10.1128/MRA.00171-20

Choudoir, M. J., Narayanan, A., Rodriguez-Ramos, D., Simoes, R., Efroni, A., Sondrini, A. et DeAngelis, K. M. (2023). Pangenomes reveal genomic signatures of microbial adaptation to chronic soil warming [prépublication]. *bioRxiv*. https://doi.org/10.1101/2023.03.16.532972

Christodoulou, D., Link, H., Fuhrer, T., Kochanowski, K., Gerosa, L. et Sauer, U. (2018). Reserve Flux Capacity in the Pentose Phosphate Pathway Enables Escherichia coli's Rapid Response to Oxidative Stress. *Cell Systems, 6*(5), 569-578.e7. https://doi.org/10.1016/j.cels.2018.04.009

Chwastek, G., Surma, M. A., Rizk, S., Grosser, D., Lavrynenko, O., Rucińska, M., Jambor, H. et Sáenz, J. (2020). Principles of Membrane Adaptation Revealed through Environmentally Induced Bacterial Lipidome Remodeling. *Cell Reports, 32*(12). https://doi.org/10.1016/j.celrep.2020.108165

Coleman, M. L. et Chisholm, S. W. (2010). Ecosystem-specific selection pressures revealed through comparative population genomics. *Proceedings of the National Academy of Sciences, 107*(43), 18634-18639. https://doi.org/10.1073/pnas.1009480107

Copeland, J. K., Yuan, L., Layeghifard, M., Wang, P. W. et Guttman, D. S. (2015). Seasonal Community Succession of the Phyllosphere Microbiome. *Molecular Plant-Microbe Interactions, 28*(3), 274-285. https://doi.org/10.1094/MPMI-10-14-0331-FI

Corpe, W. A. et Rheem, S. (1989). Ecology of the methylotrophic bacteria on living leaf surfaces. *FEMS Microbiology Ecology, 5*(4), 243-249. https://doi.org/10.1111/j.1574-6968.1989.tb03698.x

Craig, L., Pique, M. E. et Tainer, J. A. (2004). Type IV pilus structure and bacterial pathogenicity. *Nature Reviews Microbiology, 2*(5), 363-378. https://doi.org/10.1038/nrmicro885

Csotonyi, J. T., Swiderski, J., Stackebrandt, E. et Yurkov, V. (2010). A new environment for aerobic anoxygenic phototrophic bacteria: biological soil crusts. *Environmental Microbiology Reports, 2*(5), 651-656. https://doi.org/10.1111/j.1758-2229.2010.00151.x

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M. et Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience, 10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

DeBellis, T., Kembel, S. W. et Lessard, J.-P. (2019). Shared mycorrhizae but distinct communities of other root-associated microbes on co-occurring native and invasive maples. *PeerJ, 7*, e7295. https://doi.org/10.7717/peerj.7295

De Bellis, T., Laforest-Lapointe, I., Solarik, K. A., Gravel, D. et Kembel, S. W. (2022). Regional variation drives differences in microbial communities associated with sugar maple across a latitudinal range. *Ecology, 103*(8), e3727. https://doi.org/10.1002/ecy.3727

de la Riva, E. G., Marañón, T., Violle, C., Villar, R. et Pérez-Ramos, I. M. (2017). Biogeochemical and Ecomorphological Niche Segregation of Mediterranean Woody Species along a Local Gradient. *Frontiers in Plant Science, 8*, 1242. https://doi.org/10.3389/fpls.2017.01242

De Mandal, S. et Jeon, J. (2023). Phyllosphere Microbiome in Plant Health and Disease. *Plants, 12*(19), 3481. https://doi.org/10.3390/plants12193481

Delmont, T. O. et Eren, A. M. (2018). Linking pangenomes and metagenomes: the Prochlorococcus metapangenome. *PeerJ, 6*, e4320. https://doi.org/10.7717/peerj.4320

Delmotte, N., Knief, C., Chaffron, S., Innerebner, G., Roschitzki, B., Schlapbach, R., von Mering, C. et Vorholt, J. A. (2009). Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. *Proceedings of the National Academy of Sciences, 106*(38), 16428-16433. https://doi.org/10.1073/pnas.0905240106

de Matos, J. P., Ribeiro, D. F., da Silva, A. K., de Paula, C. H., Cordeiro, I. F., Lemes, C. G. de C., Sanchez, A. B., Rocha, L. C. M., Garcia, C. C. M., Almeida, N. F., Alves, R. M., de Abreu, V. A. C., Varani, A. M. et Moreira, L. M. (2024). Diversity and potential functional role of phyllosphere-associated actinomycetota isolated from cupuassu (Theobroma grandiflorum) leaves: implications for ecosystem dynamics and plant defence strategies. *Molecular Genetics and Genomics, 299*(1), 73. https://doi.org/10.1007/s00438-024-02162-1

de Souza, R. S. C., Armanhi, J. S. L. et Arruda, P. (2020). From Microbiome to Traits: Designing Synthetic Microbial Communities for Improved Crop Resiliency. *Frontiers in Plant Science, 11*, 1179. https://doi.org/10.3389/fpls.2020.01179

Diba, F., Hoque, M. N., Rahman, M. S., Haque, F., Rahman, K. Md. J., Moniruzzaman, Md., Khan, M., Hossain, M. A. et Sultana, M. (2023). Metagenomic and culture-dependent approaches unveil active microbial community and novel functional genes involved in arsenic mobilization and detoxification in groundwater. *BMC Microbiology, 23*(1), 241. https://doi.org/10.1186/s12866-023-02980-0

Dini-Andreote, F., Stegen, J. C., Van Elsas, J. D. et Salles, J. F. (2015). Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proceedings of the National Academy of Sciences, 112*(11). https://doi.org/10.1073/pnas.1414261112

Domingo-Sananes, M. R. et McInerney, J. O. (2021). Mechanisms That Shape Microbial Pangenomes. *Trends in Microbiology, 29*(6), 493-503. https://doi.org/10.1016/j.tim.2020.12.004

Dong, X., Peng, Y., Wang, M., Woods, L., Wu, W., Wang, Y., Xiao, X., Li, J., Jia, K., Greening, C., Shao, Z. et Hubert, C. R. J. (2023). Evolutionary ecology of microbial populations inhabiting deep sea sediments associated with cold seeps. *Nature Communications, 14*(1), 1127. https://doi.org/10.1038/s41467-023-36877-3

Dörken, V. M. et Lepetit, B. (2018). Morpho-anatomical and physiological differences between sun and shade leaves in Abies alba Mill. (Pinaceae, Coniferales): a combined approach. *Plant, Cell & Environment, 41*(7), 1683-1697. https://doi.org/10.1111/pce.13213

Dorokhov, Y. L., Sheshukova, E. V. et Komarova, T. V. (2018). Methanol in Plant Life. *Frontiers in Plant Science, 9*, 1623. https://doi.org/10.3389/fpls.2018.01623

Doronina, N. V., Ivanova, E. G., Suzina, N. E. et Trotsenko, Yu. A. (2004). Methanotrophs and methylobacteria are found in woody plant tissues within the winter period. *Microbiology, 73*(6), 702-709. https://doi.org/10.1007/s11021-005-0012-0

Doronina, N. V., Sokolov, A. P. et Trotsenko, Y. A. (1996). Isolation and initial characterization of aerobic chloromethane-utilizing bacteria. *FEMS Microbiology Letters, 142*(2-3), 179-183. https://doi.org/10.1111/j.1574-6968.1996.tb08427.x

Douglas, G. M. et Shapiro, B. J. (2021). Genic Selection Within Prokaryotic Pangenomes. *Genome Biology and Evolution, 13*(11), evab234. https://doi.org/10.1093/gbe/evab234

Dourado, M. N., Andreote, F. D., Dini-Andreote, F., Conti, R., Araújo, J. M. et Araújo, W. L. (2012). Analysis of 16S rRNA and mxaF genes revealing insights into Methylobacterium niche-specific plant association. *Genetics and Molecular Biology, 35*(1), 142-148.

Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guénard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N. et Wagner, H. H. (2024). *adespatial: Multivariate Multiscale Spatial Analysis*. https://CRAN.R-project.org/package=adespatial

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research, 32*(5), 1792-1797. https://doi.org/10.1093/nar/gkh340

Eisenhauer, N., Scheu, S. et Jousset, A. (2012). Bacterial Diversity Stabilizes Community Productivity. *PLOS One, 7*(3), e34517. https://doi.org/10.1371/journal.pone.0034517

Eller, A. S. D., de Gouw, J., Graus, M. et Monson, R. K. (2012). Variation among different genotypes of hybrid poplar with regard to leaf volatile organic compound emissions. *Ecological Applications, 22*(7), 1865-1875. https://doi.org/10.1890/11-2273.1

Engelberg-Kulka, H., Hazan, R. et Amitai, S. (2005). mazEF: a chromosomal toxin-antitoxin module that triggers programmed cell death in bacteria. *Journal of Cell Science, 118*(19), 4327-4332. https://doi.org/10.1242/jcs.02619

Environnement et Ressources naturelles. (2024). *Données des stations pour le calcul des normales climatiques au Canada de 1981 à 2010, Saint-Hyppolyte, QC*. https://climat.meteo.gc.ca/climate_normals/results_1981_2010_f.html?stnID=5630&autofwd=1

Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L. et Delmont, T. O. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ, 3*, e1319. https://doi.org/10.7717/peerj.1319

Ereshefsky, M. (2010). Microbiology and the species problem. *Biology & Philosophy, 25*(4), 553-568. https://doi.org/10.1007/s10539-010-9211-9

Esser, D. S., Leveau, J. H. J., Meyer, K. M. et Wiegand, K. (2015). Spatial scales of interactions among bacteria and between bacteria and the leaf surface. *FEMS Microbiology Ecology, 91*(3), fiu034. https://doi.org/10.1093/femsec/fiu034

Fairman, J. W., Noinaj, N. et Buchanan, S. K. (2011). The structural biology of β-barrel membrane proteins: a summary of recent reports. *Current Opinion in Structural Biology, 21*(4), 523-531. https://doi.org/10.1016/j.sbi.2011.05.005

FAO. (2020). *Global Forest Resources Assessment 2020*. https://doi.org/10.4060/ca9825en

Farré-Armengol, G., Filella, I., Llusia, J. et Peñuelas, J. (2016). Bidirectional Interaction between Phyllospheric Microbiotas and Plant Volatile Emissions. *Trends in Plant Science, 21*(10), 854-860. https://doi.org/10.1016/j.tplants.2016.06.005

Faust, K. et Raes, J. (2012). Microbial interactions: from networks to models. *Nature Reviews Microbiology, 10*(8), 538-550. https://doi.org/10.1038/nrmicro2832

Fayet, O., Ziegelhoffer, T. et Georgopoulos, C. (1988). The groES and groEL heat shock gene products of Escherichia coli are essential for bacterial growth at all temperatures. *Journal of Bacteriology,* 171(3), 1379-1385. https://doi.org/10.1128/jb.171.3.1379-1385.1989

Ferrera, I., Sánchez, O., Kolářová, E., Koblížek, M. et Gasol, J. M. (2017). Light enhances the growth rates of natural populations of aerobic anoxygenic phototrophic bacteria. *The ISME Journal, 11*(10), 2391-2393. https://doi.org/10.1038/ismej.2017.79

Flores-Núñez, V. M., Camarena-Pozos, D. A., Chávez-González, J. D., Alcalde-Vázquez, R., Vázquez-Sánchez, M. N., Hernández-Melgar, A. G., Xool-Tamayo, J., Moreno-Ulloa, A. et Martínez, L. P. P. (2023). Synthetic Communities Increase Microbial Diversity and Productivity of Agave tequilana Plants in the Field. *Phytobiomes Journal, 7*(4), 435-448. https://doi.org/10.1094/PBIOMES-01-23-0001-R

Francez-Charlot, A., Frunzke, J., Reichen, C., Ebneter, J. Z., Gourion, B. et Vorholt, J. A. (2009). Sigma factor mimicry involved in regulation of general stress response. *Proceedings of the National Academy of Sciences, 106*(9), 3467-3472. https://doi.org/10.1073/pnas.0810291106

Franzetti, A., Gandolfi, I., Bestetti, G., Padoa Schioppa, E., Canedoli, C., Brambilla, D., Cappelletti, D., Sebastiani, B., Federici, E., Papacchini, M. et Ambrosini, R. (2020). Plant-microorganisms interaction promotes removal of air pollutants in Milan (Italy) urban area. *Journal of Hazardous Materials, 384*, 121021. https://doi.org/10.1016/j.jhazmat.2019.121021

Freeman, B. C., Chen, C. et Beattie, G. A. (2010). Identification of the trehalose biosynthetic loci of Pseudomonas syringae and their contribution to fitness in the phyllosphere. *Environmental Microbiology, 12*(6), 1486-1497. https://doi.org/10.1111/j.1462-2920.2010.02171.x

Frelich, L. E., Montgomery, R. A. et Oleksyn, J. (2015). Northern Temperate Forests. Dans K. S.-H. Peh, R. T. Corlett et Y. Bergeron (dir.), *Routledge Handbook of Forest Ecology*. Routledge.

Fry, S. C. (1989). Cellulases, hemicelluloses and auxin-stimulated growth: a possible relationship. *Physiologia Plantarum, 75*(4), 532-536. https://doi.org/10.1111/j.1399-3054.1989.tb05620.x

Fürnkranz, M., Wanek, W., Richter, A., Abell, G., Rasche, F. et Sessitsch, A. (2008). Nitrogen fixation by phyllosphere bacteria associated with higher plants and their colonizing epiphytes of a tropical lowland rainforest of Costa Rica. *The ISME Journal, 2*(5), 561-570. https://doi.org/10.1038/ismej.2008.14

Gallego, V., García, M. T. et Ventosa, A. (2005a). Methylobacterium hispanicum sp. nov. and Methylobacterium aquaticum sp. nov., isolated from drinking water. *International Journal of Systematic and Evolutionary Microbiology, 55*(1), 281-287. https://doi.org/10.1099/ijs.0.63319-0

Gallego, V., García, M. T. et Ventosa, A. (2005b). Methylobacterium isbiliense sp. nov., isolated from the drinking water system of Sevilla, Spain. *International Journal of Systematic and Evolutionary Microbiology, 55*(6), 2333-2337. https://doi.org/10.1099/ijs.0.63773-0

Gallego, V., García, M. T. et Ventosa, A. (2005c). Methylobacterium variabile sp. nov., a methylotrophic bacterium isolated from an aquatic environment. *International Journal of Systematic and Evolutionary Microbiology, 55*(4), 1429-1433. https://doi.org/10.1099/ijs.0.63597-0

Galperin, M. Y., Wolf, Y. I., Makarova, K. S., Vera Alvarez, R., Landsman, D. et Koonin, E. V. (2021). COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Research, 49*, D274-D281. https://doi.org/10.1093/nar/gkaa1018

Gao, Y., Jin, Y.-J., Li, H.-D. et Chen, H.-J. (2005). Volatile Organic Compounds and Their Roles in Bacteriostasis in Five Conifer Species. *Journal of Integrative Plant Biology, 47*(4), 499-507. https://doi.org/10.1111/j.1744-7909.2005.00081.x

Garbeva, P., Silby, M. W., Raaijmakers, J. M., Levy, S. B. et Boer, W. de. (2011). Transcriptional and antagonistic responses of Pseudomonas fluorescens Pf0-1 to phylogenetically different bacterial competitors. *The ISME journal, 5*(6), 973-985. https://doi.org/10.1038/ismej.2010.196

Gargallo-Garriga, A., Sardans, J., Granda, V., Llusià, J., Peguero, G., Asensio, D., Ogaya, R., Urbina, I., Van Langenhove, L., Verryckt, L. T., Chave, J., Courtois, E. A., Stahl, C., Grau, O., Klem, K., Urban, O., Janssens, I. A. et Peñuelas, J. (2020). Different “metabolomic niches” of the highly diverse tree species of the French Guiana rainforests. *Scientific Reports, 10*(1), 6937. https://doi.org/10.1038/s41598-020-63891-y

Gaube, P., Marchenko, P., Müller, C., Schweiger, R., Tenhaken, R., Keller, A. et Junker, R. R. (2023). Inter- and intraspecific phytochemical variation correlate with epiphytic flower and leaf bacterial communities. *Environmental Microbiology, 25*(9), 1624-1643. https://doi.org/10.1111/1462-2920.16382

Giacomini, J. J., Torres-Morales, J., Dewhirst, F. E., Borisy, G. G. et Mark Welch, J. L. (2023). Site Specialization of Human Oral Veillonella Species. *Microbiology Spectrum, 11*(1), e04042-22. https://doi.org/10.1128/spectrum.04042-22

Givnish, T. J. (1987). Comparative Studies of Leaf Form: Assessing the Relative Roles of Selective Pressures and Phylogenetic Constraints. *New Phytologist, 106*(s1), 131-160. https://doi.org/10.1111/j.1469-8137.1987.tb04687.x

Gogarten, J. P. et Townsend, J. P. (2005). Horizontal gene transfer, genome innovation and evolution. *Nature Reviews Microbiology, 3*(9), 679-687. https://doi.org/10.1038/nrmicro1204

Good, N. M., Moore, R. S., Suriano, C. J. et Martinez-Gomez, N. C. (2019). Contrasting in vitro and in vivo methanol oxidation activities of lanthanide-dependent alcohol dehydrogenases XoxF1 and ExaF from Methylobacterium extorquens AM1. *Scientific Reports, 9*(1), 4248. https://doi.org/10.1038/s41598-019-41043-1

Good, N. M., Vu, H. N., Suriano, C. J., Subuyuj, G. A., Skovran, E. et Martinez-Gomez, N. C. (2016). Pyrroloquinoline Quinone Ethanol Dehydrogenase in Methylobacterium extorquens AM1 Extends Lanthanide-Dependent Metabolism to Multicarbon Substrates. *Journal of Bacteriology, 198*(22), 3109-3118. https://doi.org/10.1128/JB.00478-16

Gourion, B., Francez-Charlot, A. et Vorholt, J. A. (2008). PhyR Is Involved in the General Stress Response of Methylobacterium extorquens AM1. *Journal of Bacteriology, 190*(3), 1027-1035. https://doi.org/10.1128/JB.01483-07

Gourion, B., Rossignol, M. et Vorholt, J. A. (2006). A proteomic study of Methylobacterium extorquens reveals a response regulator essential for epiphytic growth. *Proceedings of the National Academy of Sciences, 103*(35), 13186-13191. https://doi.org/10.1073/pnas.0603530103

Goychuk, A., Kulbanska, I., Shvets, M., Pasichnyk, L., Patyka, V., Kalinichenko, A. et Degtyareva, L. (2023). Bacterial Diseases of Bioenergy Woody Plants in Ukraine. *Sustainability, 15*(5), 4189. https://doi.org/10.3390/su15054189

Green, P. N. (2006). Methylobacterium. Dans M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer et E. Stackebrandt (dir.), *The Prokaryotes* (p. 257-265). Springer New York. https://doi.org/10.1007/0-387-30745-1_14

Green, P. N. et Ardley, J. K. (2018). Review of the genus Methylobacterium and closely related organisms: a proposal that some Methylobacterium species be reclassified into a new genus, Methylorubrum gen. nov. *International Journal of Systematic and Evolutionary Microbiology, 68*(9), 2727-2748. https://doi.org/10.1099/ijsem.0.002856

Grilli, J., Barabás, G., Michalska-Smith, M. J. et Allesina, S. (2017). Higher-order interactions stabilize dynamics in competitive network models. *Nature, 548*(7666), 210-213. https://doi.org/10.1038/nature23273

Grossi, C. E. M., Fantino, E., Serral, F., Zawoznik, M. S., Fernandez Do Porto, D. A. et Ulloa, R. M. (2020). Methylobacterium sp. 2A Is a Plant Growth-Promoting Rhizobacteria That Has the Potential to Improve Potato Crop Yield Under Adverse Conditions. *Frontiers in Plant Science, 11*, 71. https://doi.org/10.3389/fpls.2020.00071

Großkopf, T. et Soyer, O. S. (2014). Synthetic microbial communities. *Current Opinion in Microbiology, 18*, 72-77. https://doi.org/10.1016/j.mib.2014.02.002

Gunasekera, T. S. et Sundin, G. W. (2006). Role of nucleotide excision repair and photoreactivation in the solar UVB radiation survival of Pseudomonas syringae pv. syringae B728a. *Journal of Applied Microbiology, 100*(5), 1073-1083. https://doi.org/10.1111/j.1365-2672.2006.02841.x

Hauruseu, D. et Koblížek, M. (2012). Influence of light on carbon utilization in aerobic anoxygenic phototrophs. *Applied and Environmental Microbiology, 78*(20), 7414-7419. https://doi.org/10.1128/AEM.01747-12

Heaps, H. S. (1978). *Information Retrieval: Computational and Theoretical Aspects*. Academic Press.

Heaton, B. E., Herrou, J., Blackwell, A. E., Wysocki, V. H. et Crosson, S. (2012). Molecular Structure and Function of the Novel BrnT/BrnA Toxin-Antitoxin System of Brucella abortus. *The Journal of Biological Chemistry, 287*(15), 12098-12110. https://doi.org/10.1074/jbc.M111.332163

Hector, A., Schmid, B., Beierkuhnlein, C., Caldeira, M. C., Diemer, M., Dimitrakopoulos, P. G., Finn, J. A., Freitas, H., Giller, P. S., Good, J., Harris, R., Högberg, P., Huss-Danell, K., Joshi, J., Jumpponen, A., Körner, C., Leadley, P. W., Loreau, M., Minns, A., … Lawton, J. H. (1999). Plant Diversity and Productivity Experiments in European Grasslands. *Science, 286*(5442), 1123-1127. https://doi.org/10.1126/science.286.5442.1123

Helfrich, E. J. N., Vogel, C. M., Ueoka, R., Schäfer, M., Ryffel, F., Müller, D. B., Probst, S., Kreuzer, M., Piel, J. et Vorholt, J. A. (2018). Bipartite interactions, antibiotic production and biosynthetic potential of the Arabidopsis leaf microbiome. *Nature Microbiology, 3*(8), 909-919. https://doi.org/10.1038/s41564-018-0200-0

Herrmann, M., Geesink, P., Richter, R. et Küsel, K. (2021). Canopy Position Has a Stronger Effect than Tree Species Identity on Phyllosphere Bacterial Diversity in a Floodplain Hardwood Forest. *Microbial Ecology, 81*(1), 157-168. https://doi.org/10.1007/s00248-020-01565-y

Hickman, A. B. et Dyda, F. (2015). Mechanisms of DNA Transposition. *Microbiology spectrum, 3*(2), MDNA3-0034-2014. https://doi.org/10.1128/microbiolspec.MDNA3-0034-2014

Hinchliffe, P., Greene, N. P., Paterson, N. G., Crow, A., Hughes, C. et Koronakis, V. (2014). Structure of the periplasmic adaptor protein from a major facilitator superfamily (MFS) multidrug efflux pump. *FEBS Letters, 588*(17), 3147-3153. https://doi.org/10.1016/j.febslet.2014.06.055

Hisano, M., Searle, E. B. et Chen, H. Y. H. (2018). Biodiversity as a solution to mitigate climate change impacts on the functioning of forest ecosystems. *Biological Reviews, 93*(1), 439-456. https://doi.org/10.1111/brv.12351

Holtappels, D., Rickus, G. E. J., Morgan, T., Rezende, R. R. de, Koskella, B. et Alfenas-Zerbini, P. (2024). Evidence for the key roles of the Pseudomonas syringae mobilome in shaping biotic interactions [prépublication]. *bioRxiv.* https://doi.org/10.1101/2024.03.19.585818

Hooper, D. U., Chapin III, F. S., Ewel, J. J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J. H., Lodge, D. M., Loreau, M., Naeem, S., Schmid, B., Setälä, H., Symstad, A. J., Vandermeer, J. et Wardle, D. A. (2005). Effects of Biodiversity on Ecosystem Functioning: A Consensus of Current Knowledge*. Ecological Monographs, 75*(1), 3-35. https://doi.org/10.1890/04-0922

Hördt, A., López, M. G., Meier-Kolthoff, J. P., Schleuning, M., Weinhold, L.-M., Tindall, B. J., Gronow, S., Kyrpides, N. C., Woyke, T. et Göker, M. (2020). Analysis of 1,000+ Type-Strain Genomes Substantially Improves Taxonomic Classification of Alphaproteobacteria. *Frontiers in Microbiology, 11*, 468. https://doi.org/10.3389/fmicb.2020.00468

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W. et Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics, 11*(1), 119. https://doi.org/10.1186/1471-2105-11-119

Hyun, J. C., Monk, J. M. et Palsson, B. O. (2022). Comparative pangenomics: analysis of 12 microbial pathogen pangenomes reveals conserved global structures of genetic and functional diversity. *BMC Genomics, 23*(1), 7. https://doi.org/10.1186/s12864-021-08223-8

Ijaq, J., Chandra, D., Ray, M. K. et Jagannadham, M. V. (2022). Investigating the Functional Role of Hypothetical Proteins From an Antarctic Bacterium Pseudomonas sp. Lz4W: Emphasis on Identifying Proteins Involved in Cold Adaptation. *Frontiers in Genetics, 13*, 825269. https://doi.org/10.3389/fgene.2022.825269

Innerebner, G., Knief, C. et Vorholt, J. A. (2011). Protection of Arabidopsis thaliana against Leaf-Pathogenic Pseudomonas syringae by Sphingomonas Strains in a Controlled Model System. *Applied and Environmental Microbiology, 77*(10), 3202-3210. https://doi.org/10.1128/AEM.00133-11

Ishak, S., Rondeau-Leclaire, J., Faticov, M., Roy, S. et Laforest-Lapointe, I. (2024). Boreal moss-microbe interactions are revealed through metagenome assembly of novel bacterial species. *Scientific Reports, 14*(1), 22168. https://doi.org/10.1038/s41598-024-73045-z

Ivanova, E. G., Doronina, N. V., Shepelyakovskaya, A. O., Laman, A. G., Brovko, F. A. et Trotsenko, Y. A. (2000). Facultative and Obligate Aerobic Methylobacteria Synthesize Cytokinins. *Microbiology, 69*(6), 646-651. https://doi.org/10.1023/A:1026693805653

Ivanova, E. G., Doronina, N. V. et Trotsenko, Y. A. (2001). Aerobic Methylobacteria Are Capable of Synthesizing Auxins. *Microbiology, 70*(4), 392-397. https://doi.org/10.1023/A:1010469708107

Jacobs, J. L., Carroll, T. L. et Sundin, G. W. (2005). The Role of Pigmentation, Ultraviolet Radiation Tolerance, and Leaf Colonization Strategies in the Epiphytic Survival of Phyllosphere Bacteria. *Microbial Ecology, 49*(1), 104-113. https://doi.org/10.1007/s00248-003-1061-4

Jacobs, J. L. et Sundin, G. W. (2001). Effect of Solar UV-B Radiation on a Phyllosphere Bacterial Community. *Applied and Environmental Microbiology, 67*(12), 5488-5496. https://doi.org/10.1128/AEM.67.12.5488-5496.2001

Jacques, M.-A., Arlat, M., Boulanger, A., Boureau, T., Carrère, S., Cesbron, S., Chen, N. W. G., Cociancich, S., Darrasse, A., Denancé, N., Fischer-Le Saux, M., Gagnevin, L., Koebnik, R., Lauber, E., Noël, L. D., Pieretti, I., Portier, P., Pruvost, O., Rieux, A., … Vernière, C. (2016). Using Ecology, Physiology, and Genomics to Understand Host Specificity in Xanthomonas. *Annual Review of Phytopathology, 54*, 163-187. https://doi.org/10.1146/annurev-phyto-080615-100147

Jager, E. S. de, Wehner, F. C. et Korsten, L. (2001). Microbial Ecology of the Mango Phylloplane. *Microbial Ecology, 42*(2), 201-207. https://doi.org/10.1007/s002480000106

Janakiev, T., Dimkić, I., Bojić, S., Fira, D., Stanković, S. et Berić, T. (2020). Bacterial communities of plum phyllosphere and characterization of indigenous antagonistic Bacillus thuringiensis R3/3 isolate. *Journal of Applied Microbiology, 128*(2), 528-543. https://doi.org/10.1111/jam.14488

Jaspers, E. et Overmann, J. (2004). Ecological Significance of Microdiversity: Identical 16S rRNA Gene Sequences Can Be Found in Bacteria with Highly Divergent Genomes and Ecophysiologies. *Applied and Environmental Microbiology, 70*(8), 4831-4839. https://doi.org/10.1128/AEM.70.8.4831-4839.2004

Jessup, C. M., Kassen, R., Forde, S. E., Kerr, B., Buckling, A., Rainey, P. B. et Bohannan, B. J. M. (2004). Big questions, small worlds: microbial model systems in ecology. *Trends in Ecology & Evolution, 19*(4), 189-197. https://doi.org/10.1016/j.tree.2004.01.008

Ji, P. et Wilson, M. (2002). Assessment of the Importance of Similarity in Carbon Source Utilization Profiles between the Biological Control Agent and the Pathogen in Biological Control of Bacterial Speck of Tomato. *Applied and Environmental Microbiology, 68*(9), 4383-4389. https://doi.org/10.1128/AEM.68.9.4383-4389.2002

Jiang, L. (2007). Negative selection effects suppress relationships between bacterial diversity and ecosystem functioning. *Ecology, 88*(5), 1075-1085. https://doi.org/10.1890/06-1556

Jørgensen, M. G., Pandey, D. P., Jaskolska, M. et Gerdes, K. (2009). HicA of Escherichia coli Defines a Novel Family of Translation-Independent mRNA Interferases in Bacteria and Archaea. *Journal of Bacteriology, 191*(4), 1191-1199. https://doi.org/10.1128/jb.01013-08

Jourand, P., Giraud, E., Béna, G., Sy, A., Willems, A., Gillis, M., Dreyfus, B. et de Lajudie, P. (2004). Methylobacterium nodulans sp. nov., for a group of aerobic, facultatively methylotrophic, legume root-nodule-forming and nitrogen-fixing bacteria. *International Journal of Systematic and Evolutionary Microbiology, 54*(6), 2269-2273. https://doi.org/10.1099/ijs.0.02902-0

Jousset, A., Becker, J., Chatterjee, S., Karlovsky, P., Scheu, S. et Eisenhauer, N. (2014). Biodiversity and species identity shape the antifungal activity of bacterial communities. *Ecology, 95*(5), 1184-1190. https://doi.org/10.1890/13-1215.1

Kachlany, S. C., Planet, P. J., Bhattacharjee, M. K., Kollia, E., DeSalle, R., Fine, D. H. et Figurski, D. H. (2000). Nonspecific Adherence by Actinobacillus actinomycetemcomitans Requires Genes Widespread in Bacteria and Archaea. *Journal of Bacteriology, 182*(21), 6169-6176. https://doi.org/10.1128/jb.182.21.6169-6176.2000

Kałużna, M., Fischer-Le Saux, M., Pothier, J. F., Jacques, M., Obradović, A., Tavares, F. et Stefani, E. (2021). Xanthomonas arboricola pv. juglandis and pv. corylina: Brothers or distant relatives? Genetic clues, epidemiology, and insights for disease management. *Molecular Plant Pathology, 22*(12), 1481-1499. https://doi.org/10.1111/mpp.13073

Karamanoli, K., Menkissoglu-Spiroudi, U., Bosabalidis, A. M., Vokou, D. et Constantinidou, H.-I. A. (2005). Bacterial colonization of the phyllosphere of nineteen plant species and antimicrobial activity of their leaf secondary metabolites against leaf associated bacteria. *Chemoecology, 15*(2), 59-67. https://doi.org/10.1007/s00049-005-0297-5

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., Blomberg, S. P. et Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics, 26*(11), 1463-1464. https://doi.org/10.1093/bioinformatics/btq166

Kembel, S. W. et Mueller, R. C. (2014). Plant traits and taxonomy drive host associations in tropical phyllosphere fungal communities. *Botany, 92*(4), 303-311. https://doi.org/10.1139/cjb-2013-0194

Kembel, S. W., O'Connor, T. K., Arnold, H. K., Hubbell, S. P., Wright, S. J. et Green, J. L. (2014). Relationships between phyllosphere bacterial communities and plant functional traits in a neotropical forest. *Proceedings of the National Academy of Sciences, 111*(38), 13715-13720. https://doi.org/10.1073/pnas.1216057111

Kettler, G. C., Martiny, A. C., Huang, K., Zucker, J., Coleman, M. L., Rodrigue, S., Chen, F., Lapidus, A., Ferriera, S., Johnson, J., Steglich, C., Church, G. M., Richardson, P. et Chisholm, S. W. (2007). Patterns and Implications of Gene Gain and Loss in the Evolution of Prochlorococcus. *PLOS Genetics, 3*(12), e231. https://doi.org/10.1371/journal.pgen.0030231

Kiefl, E., Esen, O. C., Miller, S. E., Kroll, K. L., Willis, A. D., Rappé, M. S., Pan, T. et Eren, A. M. (2023). Structure-informed microbial population genetics elucidate selective pressures that shape protein evolution. *Science Advances, 9*(8), eabq4632. https://doi.org/10.1126/sciadv.abq4632

Kim, M., Singh, D., Lai-Hoe, A., Go, R., Abdul Rahim, R., A.N., A., Chun, J. et Adams, J. M. (2012). Distinctive Phyllosphere Bacterial Communities in Tropical Trees. *Microbial Ecology, 63*(3), 674-681. https://doi.org/10.1007/s00248-011-9953-1

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press.

Knief, C., Delmotte, N., Chaffron, S., Stark, M., Innerebner, G., Wassmann, R., von Mering, C. et Vorholt, J. A. (2012a). Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. *The ISME Journal, 6*(7), 1378-1390. https://doi.org/10.1038/ismej.2011.192

Knief, C., Dengler, V., Bodelier, P. L. E. et Vorholt, J. A. (2012b). Characterization of Methylobacterium strains isolated from the phyllosphere and description of Methylobacterium longum sp. nov. *Antonie van Leeuwenhoek, 101*(1), 169-183. https://doi.org/10.1007/s10482-011-9650-6

Knief, C., Ramette, A., Frances, L., Alonso-Blanco, C. et Vorholt, J. A. (2010). Site and plant species are important determinants of the Methylobacterium community composition in the plant phyllosphere. *The ISME Journal, 4*(6), 719-728. https://doi.org/10.1038/ismej.2010.9

Knöppel, A., Lind, P. A., Lustig, U., Näsvall, J. et Andersson, D. I. (2014). Minor Fitness Costs in an Experimental Model of Horizontal Gene Transfer in Bacteria. *Molecular Biology and Evolution, 31*(5), 1220-1227. https://doi.org/10.1093/molbev/msu076

Koblížek, M. (2015). Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiology Reviews, 39*(6), 854-870. https://doi.org/10.1093/femsre/fuv032

Koenig, R. L., Morris, R. O. et Polacco, J. C. (2002). tRNA Is the Source of Low-Level trans-Zeatin Production in Methylobacterium spp. *Journal of Bacteriology, 184*(7), 1832-1842. https://doi.org/10.1128/JB.184.7.1832-1842.2002

Kollmann, J., Meyer, S. T., Bateman, R., Conradi, T., Gossner, M. M., de Souza Mendonça Jr., M., Fernandes, G. W., Hermann, J.-M., Koch, C., Müller, S. C., Oki, Y., Overbeck, G. E., Paterno, G. B., Rosenfield, M. F., Toma, T. S. P. et Weisser, W. W. (2016). Integrating ecosystem functions into restoration ecology - recent advances and future directions. *Restoration Ecology, 24*(6), 722-730. https://doi.org/10.1111/rec.12422

Kolodziejek, A. M., Sinclair, D. J., Seo, K. S., Schnider, D. R., Deobald, C. F., Rohde, H. N., Viall, A. K., Minnich, S. S., Hovde, C. J., Minnich, S. A. et Bohach, G. A. (2007). Phenotypic characterization of OmpX, an Ail homologue of Yersinia pestis KIM. *Microbiology, 153*(9), 2941-2951. https://doi.org/10.1099/mic.0.2006/005694-0

Kończak, B., Wiesner-Sękala, M. et Ziembińska-Buczyńska, A. (2024). The European trees phyllosphere characteristics and its potential in air bioremediation. *Environmental Pollution, 349*, 123977. https://doi.org/10.1016/j.envpol.2024.123977

Kovaleva, J., Degener, J. E. et van der Mei, H. C. (2014). Methylobacterium and Its Role in Health Care-Associated Infection. *Journal of Clinical Microbiology, 52*(5), 1317-1321. https://doi.org/10.1128/JCM.03561-13

Krueger, J. H. et Walker, G. C. (1984). groEL and dnaK genes of Escherichia coli are induced by UV irradiation and nalidixic acid in an htpR+-dependent fashion. *Proceedings of the National Academy of Sciences, 81*(5), 1499-1503. https://doi.org/10.1073/pnas.81.5.1499

Kuennecke, B. H. (2008). *Temperate Forest Biomes*. Bloomsbury Publishing USA.

Kumar, J., Babele, P. K., Singh, D. et Kumar, A. (2016). UV-B Radiation Stress Causes Alterations in Whole Cell Protein Profile and Expression of Certain Genes in the Rice Phyllospheric Bacterium Enterobacter cloacae. *Frontiers in Microbiology, 7*, 1440. https://doi.org/10.3389/fmicb.2016.01440

Kumar, P., Lee, J.-H., Beyenal, H. et Lee, J. (2020). Fatty Acids as Antibiofilm and Antivirulence Agents. *Trends in Microbiology, 28*(9), 753-768. https://doi.org/10.1016/j.tim.2020.03.014

La Porta, N., Hietala, A. M. et Baldi, P. (2023). Bacterial diseases in forest trees. Dans F. O. Asiegbu et A. Kovalchuk (dir.), *Forest Microbiology* (vol. 3, p. 139-166). Academic Press. https://doi.org/10.1016/B978-0-443-18694-3.00001-8

Laforest-Lapointe, I., Messier, C. et Kembel, S. W. (2016a). Host species identity, site and time drive temperate tree phyllosphere bacterial community structure. *Microbiome, 4*(1), 27. https://doi.org/10.1186/s40168-016-0174-1

Laforest-Lapointe, I., Messier, C. et Kembel, S. W. (2016b). Tree phyllosphere bacterial communities: exploring the magnitude of intra- and inter-individual variation among host species. *PeerJ, 4*, e2367. https://doi.org/10.7717/peerj.2367

Laforest-Lapointe, I., Paquette, A., Messier, C. et Kembel, S. W. (2017a). Leaf bacterial diversity mediates plant diversity and ecosystem function relationships. *Nature, 546*(7656), 145-147. https://doi.org/10.1038/nature22399

Laforest-Lapointe, I., Messier, C. et Kembel, S. W. (2017b). Tree Leaf Bacterial Community Structure and Diversity Differ along a Gradient of Urban Intensity. *mSystems, 2*(6), e00087-17. https://doi.org/10.1128/mSystems.00087-17

Laforest-Lapointe, I. et Whitaker, B. K. (2019). Decrypting the phyllosphere microbiota: progress and challenges. *American Journal of Botany, 106*(2), 171-173. https://doi.org/10.1002/ajb2.1229

Lajoie, G. et Kembel, S. W. (2021). Plant-bacteria associations are phylogenetically structured in the phyllosphere. *Molecular Ecology, 30*(21), 5572-5587. https://doi.org/10.1111/mec.16131

Lajoie, G., Maglione, R. et Kembel, S. W. (2020). Adaptive matching between phyllosphere bacteria and their tree hosts in a neotropical forest. *Microbiome, 8*(1), 70. https://doi.org/10.1186/s40168-020-00844-7

Lambais, M. R., Crowley, D. E., Cury, J. C., Büll, R. C. et Rodrigues, R. R. (2006). Bacterial Diversity in Tree Canopies of the Atlantic Forest. *Science, 312*(5782), 1917-1917. https://www.jstor.org/stable/3846432

Langenheder, S., Bulling, M. T., Solan, M. et Prosser, J. I. (2010). Bacterial Biodiversity-Ecosystem Functioning Relations Are Modified by Environmental Complexity. *PLOS One, 5*(5), e10834. https://doi.org/10.1371/journal.pone.0010834

Langmead, B. et Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods, 9*(4), 357-359. https://doi.org/10.1038/nmeth.1923

Last, F. T. (1955). Seasonal incidence of Sporobolomyces on cereal leaves. *Transactions of the British Mycological Society, 38*(3), 221-239. https://doi.org/10.1016/S0007-1536(55)80069-1

Le Moigne, A., Randegger, F., Gupta, A., Petchey, O. L. et Pernthaler, J. (2023). Stochasticity causes high β-diversity and functional divergence of bacterial assemblages in closed systems. *Ecology, 104*(4), e4005. https://doi.org/10.1002/ecy.4005

Leducq, J.-B., Seyer-Lamontagne, É., Condrain-Morel, D., Bourret, G., Sneddon, D., Foster, J. A., Marx, C. J., Sullivan, J. M., Shapiro, B. J. et Kembel, S. W. (2022a). Fine-Scale Adaptations to Environmental Variation and Growth Strategies Drive Phyllosphere Methylobacterium Diversity. *mBio, 13*(1), e03175-21. https://doi.org/10.1128/mbio.03175-21

Leducq, J.-B., Sneddon, D., Santos, M., Condrain-Morel, D., Bourret, G., Martinez-Gomez, N. C., Lee, J. A., Foster, J. A., Stolyar, S., Shapiro, B. J., Kembel, S. W., Sullivan, J. M. et Marx, C. J. (2022b). Comprehensive Phylogenomics of Methylobacterium Reveals Four Evolutionary Distinct Groups and Underappreciated Phyllosphere Diversity. *Genome Biology and Evolution, 14*(8), evac123. https://doi.org/10.1093/gbe/evac123

Lee, G. M., Scott-Nevros, Z. K., Lee, S.-M. et Kim, D. (2022). Pan-genome Analysis Reveals Comparative Genomic Features of Central Metabolic Pathways in Methylorubrum extorquens. *Biotechnology and Bioprocess Engineering, 28*, 990-1004. https://doi.org/10.1007/s12257-022-0154-1

Lee, H. S., Madhaiyan, M., Kim, C. W., Choi, S. J., Chung, K. Y. et Sa, T. M. (2006). Physiological enhancement of early growth of rice seedlings (Oryza sativa L.) by production of phytohormone of N2-fixing methylotrophic isolates. *Biology and Fertility of Soils, 42*(5), 402-408. https://doi.org/10.1007/s00374-006-0083-8

Lee, M.-C. et Marx, C. J. (2012). Repeated, Selection-Driven Genome Reduction of Accessory Genes in Experimental Populations. *PLOS Genetics, 8*(5), e1002651. https://doi.org/10.1371/journal.pgen.1002651

Legein, M., Smets, W., Vandenheuvel, D., Eilers, T., Muyshondt, B., Prinsen, E., Samson, R. et Lebeer, S. (2020). Modes of Action of Microbial Biocontrol in the Phyllosphere. *Frontiers in Microbiology, 11*, 1619. https://doi.org/10.3389/fmicb.2020.01619

Legendre, P. (2014). Interpreting the replacement and richness difference components of beta diversity. *Global Ecology and Biogeography, 23*(11), 1324-1334. https://doi.org/10.1111/geb.12207

Lenth, R. V. (2022). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (v1.8.2). https://CRAN.R-project.org/package=emmeans

Leveau, J. H. J. et Tech, J. J. (2011). Grapevine microbiomics: bacterial diversity on grape leaves and berries revealed by high-throughput sequence analysis of 16S rRNA amplicons. *Acta Horticulturae, 905*, 31-42. https://doi.org/10.17660/ActaHortic.2011.905.2

Levine, J. M., Bascompte, J., Adler, P. B. et Allesina, S. (2017). Beyond pairwise mechanisms of species coexistence in complex communities. *Nature, 546*(7656), 56-64. https://doi.org/10.1038/nature22898

Li, Y., Wang, T., Jing, H. et Xiao, Y. (2024a). Evolutionary ecology of denitrifying methanotrophic NC10 bacteria in the deep-sea biosphere. *Molecular Ecology, 33*(12), e17372. https://doi.org/10.1111/mec.17372

Li, Y., Jin, L., Wu, M., Wang, B., Qu, N., Zhou, H., Chen, T., Liu, G., Yue, M. et Zhang, G. (2024b). Forest management positively reshapes the phyllosphere bacterial community and improves community stability. *Environment International, 186*, 108611. https://doi.org/10.1016/j.envint.2024.108611

Lindow, S. E. et Brandl, M. T. (2003). Microbiology of the Phyllosphere. *Applied and Environmental Microbiology, 69*(4), 1875-1883. https://doi.org/10.1128/AEM.69.4.1875-1883.2003

Lindsey, B. E., Rivero, L., Calhoun, C. S., Grotewold, E. et Brkljacic, J. (2017). Standardized Method for High-throughput Sterilization of Arabidopsis Seeds. *Journal of Visualized Experiments, 128*, e56587. https://doi.org/10.3791/56587

Liu, J., Sun, X., Zuo, Y., Hu, Q. et He, X. (2023). Plant species shape the bacterial communities on the phyllosphere in a hyper-arid desert. *Microbiological Research, 269*, 127314. https://doi.org/10.1016/j.micres.2023.127314

Loreau, M. (2010). Linking biodiversity and ecosystems: towards a unifying ecological theory. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*(1537), 49-60. https://doi.org/10.1098/rstb.2009.0155

Loreau, M. et Hector, A. (2001). Partitioning selection and complementarity in biodiversity experiments. *Nature, 412*(6842), 72-76. https://doi.org/10.1038/35083573

Love, M. I., Huber, W. et Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology, 15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lu, J., Breitwieser, F. P., Thielen, P. et Salzberg, S. L. (2017). Bracken: estimating species abundance in metagenomics data. *PeerJ Computer Science, 3*, e104. https://doi.org/10.7717/peerj-cs.104

Ma, A., Lv, D., Zhuang, X. et Zhuang, G. (2013). Quorum Quenching in Culturable Phyllosphere Bacteria from Tobacco. *International Journal of Molecular Sciences, 14*(7), 14607-14619. https://doi.org/10.3390/ijms140714607

Madhaiyan, M. et Poonguzhali, S. (2014). Methylobacterium pseudosasae sp. nov., a pink-pigmented, facultatively methylotrophic bacterium isolated from the bamboo phyllosphere. *Antonie van Leeuwenhoek, 105*(2), 367-376. https://doi.org/10.1007/s10482-013-0085-0

Madhaiyan, M., Poonguzhali, S., Kwon, S.-W. et Sa, T.-M. (2009). Methylobacterium phyllosphaerae sp. nov., a pink-pigmented, facultative methylotroph from the phyllosphere of rice. *International Journal of Systematic and Evolutionary Microbiology, 59*(1), 22-27. https://doi.org/10.1099/ijs.0.001693-0

Madhaiyan, M., Poonguzhali, S., Senthilkumar, M., Chung, H., Yang, J., Sundaram, S. et Sa, T. (2004). Growth promotion and induction of systemic resistance in rice cultivar Co-47 (Oryza sativa L.) by Methylobacterium spp. *Botanical Bulletin of Academia Sinica, 45*(4), 315-324.

Madhaiyan, M., Poonguzhali, S., Senthilkumar, M., Lee, J.-S. et Lee, K.-C. (2012). Methylobacterium gossipiicola sp. nov., a pink-pigmented, facultatively methylotrophic bacterium isolated from the cotton phyllosphere. *International Journal of Systematic and Evolutionary Microbiology, 62*(1), 162-167. https://doi.org/10.1099/ijs.0.030148-0

Madhaiyan, M., Suresh Reddy, B. V., Anandham, R., Senthilkumar, M., Poonguzhali, S., Sundaram, S. P. et Sa, T. (2006). Plant Growth–Promoting Methylobacterium Induces Defense Responses in Groundnut (Arachis hypogaea L.) Compared with Rot Pathogens. *Current Microbiology, 53*(4), 270-276. https://doi.org/10.1007/s00284-005-0452-9

Maeng, S., Kim, D.-U., Lim, S., Lee, B.-H., Lee, K., Kim, M., Srinivasan, S. et Bai, J. (2021). Methylobacterium radiodurans sp. nov., a novel radiation-resistant Methylobacterium. *Archives of Microbiology, 203*(6), 3435-3442. https://doi.org/10.1007/s00203-021-02293-8

Maignien, L., DeForce, E. A., Chafee, M. E., Eren, A. M. et Simmons, S. L. (2014). Ecological Succession and Stochastic Variation in the Assembly of Arabidopsis thaliana Phyllosphere Communities. *mBio, 5*(1), e00682-13. https://doi.org/10.1128/mbio.00682-13

Malaterre, C. (2017). The challenges of microbial diversity. Dans J. Garson, Plutynski et S. Sarkar (dir.), *The Routledge Handbook of Philosophy of Biodiversity* (p. 139-152). Routledge.

Maleki, F., Khosravi, A., Nasser, A., Taghinejad, H. et Azizian, M. (2016). Bacterial Heat Shock Protein Activity. *Journal of Clinical and Diagnostic Research, 10*(3), BE01-BE03. https://doi.org/10.7860/JCDR/2016/14568.7444

Martino, M. E., Bayjanov, J. R., Caffrey, B. E., Wels, M., Joncour, P., Hughes, S., Gillet, B., Kleerebezem, M., van Hijum, S. A. F. T. et Leulier, F. (2016). Nomadic lifestyle of Lactobacillus plantarum revealed by comparative genomics of 54 strains isolated from different habitats. *Environmental Microbiology, 18*(12), 4974-4989. https://doi.org/10.1111/1462-2920.13455

Martiny, A. C., Treseder, K. et Pusch, G. (2013). Phylogenetic conservatism of functional traits in microorganisms. *The ISME Journal, 7*(4), 830-838. https://doi.org/10.1038/ismej.2012.160

Massoni, J., Bortfeld-Miller, M., Widmer, A. et Vorholt, J. A. (2021). Capacity of soil bacteria to reach the phyllosphere and convergence of floral communities despite soil microbiota variation. *Proceedings of the National Academy of Sciences, 118*(41), e2100150118. https://doi.org/10.1073/pnas.2100150118

Matsui, K. (2006). Green leaf volatiles: hydroperoxide lyase pathway of oxylipin metabolism. *Current Opinion in Plant Biology, 9*(3), 274-280. https://doi.org/10.1016/j.pbi.2006.03.002

McInerney, J. O., McNally, A. et O'Connell, M. J. (2017). Why prokaryotes have pangenomes. *Nature Microbiology, 2*(4), 17040. https://doi.org/10.1038/nmicrobiol.2017.40

McInerney, J. O., Whelan, F. J., Domingo-Sananes, M. R., McNally, A. et O'Connell, M. J. (2020). Pangenomes and Selection: The Public Goods Hypothesis. Dans H. Tettelin et D. Medini (dir.), *The Pangenome* (p. 151-167). Springer International Publishing. https://doi.org/10.1007/978-3-030-38281-0_7

Medini, D., Donati, C., Tettelin, H., Masignani, V. et Rappuoli, R. (2005). The microbial pan-genome. *Current Opinion in Genetics & Development, 15*(6), 589-594. https://doi.org/10.1016/j.gde.2005.09.006

Mercier, J. et Lindow, S. E. (2000). Role of Leaf Surface Sugars in Colonization of Plants by Bacterial Epiphytes. *Applied and Environmental Microbiology, 66*(1), 369-374. https://doi.org/10.1128/aem.66.1.369-374.2000

Meyer, K. M. et Leveau, J. H. J. (2012). Microbiology of the phyllosphere: a playground for testing ecological concepts. *Oecologia, 168*(3), 621-629. https://doi.org/10.1007/s00442-011-2138-2

Meziti, A., Rodriguez-R, L. M., Hatt, J. K., Peña-Gonzalez, A., Levy, K. et Konstantinidis, K. T. (2021). The Reliability of Metagenome-Assembled Genomes (MAGs) in Representing Natural Populations: Insights from Comparing MAGs against Isolate Genomes Derived from the Same Fecal Sample. *Applied and Environmental Microbiology, 87*(6), e02593-20. https://doi.org/10.1128/AEM.02593-20

Minami, T., Anda, M., Mitsui, H., Sugawara, M., Kaneko, T., Sato, S., Ikeda, S., Okubo, T., Tsurumaru, H. et Minamisawa, K. (2016). Metagenomic Analysis Revealed Methylamine and Ureide Utilization of Soybean-Associated Methylobacterium. *Microbes and Environments, 31*(3), 268-278. https://doi.org/10.1264/jsme2.ME16035

Minder, A. C., De Rudder, K. E. E., Narberhaus, F., Fischer, H.-M., Hennecke, H. et Geiger, O. (2001). Phosphatidylcholine levels in Bradyrhizobium japonicum membranes are critical for an efficient symbiosis with the soybean host plant. *Molecular Microbiology, 39*(5), 1186-1198. https://doi.org/10.1111/j.1365-2958.2001.02325.x

Ministère des Ressources naturelles et des Forêts. (2023). *Carte écoforestière originale et résultats d'inventaire courants*. Données Québec. https://www.donneesquebec.ca/recherche/dataset/resultats-d-inventaire-et-carte-ecoforestiere

Minoche, A. E., Dohm, J. C. et Himmelbauer, H. (2011). Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biology, 12*(11), R112. https://doi.org/10.1186/gb-2011-12-11-r112

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D. et Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research, 49*, D412-D419. https://doi.org/10.1093/nar/gkaa913

Mitchell, M. K. et Ellermann, M. (2022). Long Chain Fatty Acids and Virulence Repression in Intestinal Bacterial Pathogens. *Frontiers in Cellular and Infection Microbiology, 12*, 928503. https://doi.org/10.3389/fcimb.2022.928503

Moreira, J. C. F., Brum, M., De Almeida, L. C., Barrera-Berdugo, S., De Souza, A. A., De Camargo, P. B., Oliveira, R. S., Alves, L. F., Rosado, B. H. P. et Lambais, M. R. (2021). Asymbiotic nitrogen fixation in the phyllosphere of the Amazon forest: Changing nitrogen cycle paradigms. *Science of The Total Environment, 773*, 145066. https://doi.org/10.1016/j.scitotenv.2021.145066

Mori, A. S., Lertzman, K. P. et Gustafsson, L. (2017). Biodiversity and ecosystem services in forest ecosystems: a research agenda for applied forest ecology. *Journal of Applied Ecology, 54*(1), 12-27. https://doi.org/10.1111/1365-2664.12669

Morohoshi, T., Someya, N. et Ikeda, T. (2009). Novel N -Acylhomoserine Lactone-Degrading Bacteria Isolated from the Leaf Surface of Solanum tuberosum and Their Quorum-Quenching Properties. *Bioscience, Biotechnology, and Biochemistry, 73*(9), 2124-2127. https://doi.org/10.1271/bbb.90283

Morris, C. et Kinkel, L. L. (2002). Fifty years of phyllosphere microbiology: significant contributions to research in related fields. *International Symposium on the Microbiology of Aerial Plant Surfaces, Berkeley, USA, August 3-7, 2000.*

Naeem, S., Knops, J. M. H., Tilman, D., Howe, K. M., Kennedy, T. et Gale, S. (2000). Plant Diversity Increases Resistance to Invasion in the Absence of Covarying Extrinsic Factors. *Oikos, 91*(1), 97-108. https://doi.org/10.1034/j.1600-0706.2000.910108.x

Naeem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H. et Woodfin, R. M. (1995). Empirical evidence that declining species diversity may alter the performance of terrestrial ecosystems. *Philosophical Transactions of the Royal Society B, 347*(1321), 249-262. https://doi.org/10.1098/rstb.1995.0025

Nakagawa, T., Mitsui, R., Tani, A., Sasa, K., Tashiro, S., Iwama, T., Hayakawa, T. et Kawai, K. (2012). A Catalytic Role of XoxF1 as La3+-Dependent Methanol Dehydrogenase in Methylobacterium extorquens Strain AM1. *PLOS One, 7*(11), e50480. https://doi.org/10.1371/journal.pone.0050480

Nemecek-Marshall, M., MacDonald, R. C., Franzen, J. J., Wojciechowski, C. L. et Fall, R. (1995). Methanol Emission from Leaves (Enzymatic Detection of Gas-Phase Methanol and Relation of Methanol Fluxes to Stomatal Conductance and Leaf Development). *Plant Physiology, 108*(4), 1359-1368. https://doi.org/10.1104/pp.108.4.1359

Nicotra, A. B., Leigh, A., Boyce, C. K., Jones, C. S., Niklas, K. J., Royer, D. L. et Tsukaya, H. (2011). The evolution and functional significance of leaf shape in the angiosperms. *Functional Plant Biology, 38*(7), 535-552. https://doi.org/10.1071/FP11057

Niehus, R., Mitri, S., Fletcher, A. G. et Foster, K. R. (2015). Migration and horizontal gene transfer divide microbial genomes into multiple niches. *Nature Communications, 6*(1), 8924. https://doi.org/10.1038/ncomms9924

Nissinen, R., Franz, O., Kovanen, S., Mäkelä, M., Kraft, V., Ketola, K., Liukkonen, A., Heikkilä-Huhta, K., Häkkänen, H. et Ihalainen, J. A. (2023). Aerobic anoxygenic phototrophic bacteria are ubiquitous in phyllo- and endosphere microbiomes of boreal and subarctic plants [prépublication]. *bioRxiv*. https://doi.org/10.1101/2023.02.19.529139

Nowell, R. W., Green, S., Laue, B. E. et Sharp, P. M. (2014). The Extent of Genome Flux and Its Role in the Differentiation of Bacterial Lineages. *Genome Biology and Evolution, 6*(6), 1514-1529. https://doi.org/10.1093/gbe/evu123

Ochman, H., Lawrence, J. G. et Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature, 405*, 299-304. https://doi.org/10.1038/35012500

Oeum, K., Suong, M., Uon, K., Jobert, L., Bellafiore, S., Comte, A., Thomas, E., Kuok, F. et Moulin, L. (2024). Comparison of plant microbiota in diseased and healthy rice reveals methylobacteria as health signatures with biocontrol capabilities*. Frontiers in Plant Science, 15*, 1468192. https://doi.org/10.3389/fpls.2024.1468192

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., Caceres, M. D., Durand, S., … Weedon, J. (2022, 11 octobre). *vegan: Community Ecology Package* (v2.6-4). https://cran.r-project.org/web/packages/vegan/index.html

Oliver, T. H., Heard, M. S., Isaac, N. J. B., Roy, D. B., Procter, D., Eigenbrod, F., Freckleton, R., Hector, A., Orme, C. D. L., Petchey, O. L., Proença, V., Raffaelli, D., Suttle, K. B., Mace, G. M., Martín-López, B., Woodcock, B. A. et Bullock, J. M. (2015). Biodiversity and Resilience of Ecosystem Functions. *Trends in Ecology & Evolution, 30*(11), 673-684. https://doi.org/10.1016/j.tree.2015.08.009

Omer, Z. S., Tombolini, R., Broberg, A. et Gerhardson, B. (2004). Indole-3-acetic acid production by pink-pigmented facultative methylotrophic bacteria. *Plant Growth Regulation, 43*(1), 93-96. https://doi.org/10.1023/B:GROW.0000038360.09079.ad

Osburn, E. D., Yang, G., Rillig, M. C. et Strickland, M. S. (2023). Evaluating the role of bacterial diversity in supporting soil ecosystem functions under anthropogenic stress. *ISME Communications, 3*(1), 66. https://doi.org/10.1038/s43705-023-00273-1

Osono, T. (2006). Role of phyllosphere fungi of forest trees in the development of decomposer fungal communities and decomposition processes of leaf litter. *Canadian Journal of Microbiology, 52*(8), 701-716. https://doi.org/10.1139/w06-023

Øvreås, L. et Curtis, T. P. (2011). Microbial diversity and ecology. Dans A. E. Magurran et B. J. McGill (dir.), *Biological Diversity: Frontiers in Measurement and Assessment* (p. 221-236). Oxford University Press.

Park, C., Jung, H. S., Park, S., Jeon, C. O. et Park, W. (2020). Dominance of Gas-Eating, Biofilm-Forming Methylobacterium Species in the Evaporator Cores of Automobile Air-Conditioning Systems. *mSphere, 5*(1), e00761-19. https://doi.org/10.1128/mSphere.00761-19

Patt, T. E., Cole, G. C., Bland, J. et Hanson, R. S. (1974). Isolation and Characterization of Bacteria That Grow on Methane and Organic Compounds as Sole Sources of Carbon and Energy. *Journal of Bacteriology, 120*(2), 955-964.

Patt, T. E., Cole, G. C. et Hanson, R. S. (1976). Methylobacterium, a New Genus of Facultatively Methylotrophic Bacteria. *International Journal of Systematic Bacteriology, 26*(2), 226-229. https://doi.org/10.1099/00207713-26-2-226

Pattnaik, S., Rajkumari, J., Paramanandham, P. et Busi, S. (2017). Indole Acetic Acid Production and Growth-Promoting Activity of Methylobacterium extorquens MP1 and Methylobacterium zatmanii MS4 in Tomato. *International Journal of Vegetable Science, 23*(4), 321-330. https://doi.org/10.1080/19315260.2017.1283381

Pedersen, K., Zavialov, A. V., Pavlov, M. Y., Elf, J., Gerdes, K. et Ehrenberg, M. (2003). The bacterial toxin RelE displays codon-specific cleavage of mRNAs in the ribosomal A site. *Cell, 112*(1), 131-140. https://doi.org/10.1016/s0092-8674(02)01248-5

Peredo, E. L. et Simmons, S. L. (2018). Leaf-FISH: Microscale Imaging of Bacterial Taxa on Phyllosphere. *Frontiers in Microbiology, 8*, 2669. https://doi.org/10.3389/fmicb.2017.02669

Pérez-Cobas, A. E., Gomez-Valero, L. et Buchrieser, C. (2020). Metagenomic approaches in microbial ecology: an update on whole-genome and marker gene sequencing analyses. *Microbial Genomics, 6*(8), e000409. https://doi.org/10.1099/mgen.0.000409

Perna, N. T., Plunkett, G., Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J., Kirkpatrick, H. A., Pósfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E. J., Davis, N. W., Lim, A., … Blattner, F. R. (2001). Genome sequence of enterohaemorrhagic Escherichia coli O157:H7. *Nature, 409*(6819), 529-533. https://doi.org/10.1038/35054089

Perreault, R. et Laforest-Lapointe, I. (2022). Plant-microbe interactions in the phyllosphere: facing challenges of the anthropocene. *The ISME Journal, 16*(2), 339-345. https://doi.org/10.1038/s41396-021-01109-3

Petersen, L., Bollback, J. P., Dimmic, M., Hubisz, M. et Nielsen, R. (2007). Genes under positive selection in Escherichia coli. *Genome Research, 17*(9), 1336-1343. https://doi.org/10.1101/gr.6254707

Philippot, L., Spor, A., Hénault, C., Bru, D., Bizouard, F., Jones, C. M., Sarr, A. et Maron, P.-A. (2013). Loss in microbial diversity affects nitrogen cycling in soil. *The ISME Journal, 7*(8), 1609-1619. https://doi.org/10.1038/ismej.2013.34

Pires, A. P. F., Srivastava, D. S., Marino, N. A. C., MacDonald, A. A. M., Figueiredo-Barros, M. P. et Farjalla, V. F. (2018). Interactive effects of climate change and biodiversity loss on ecosystem functioning. *Ecology, 99*(5), 1203-1213. https://doi.org/10.1002/ecy.2202

Polz, M. F., Alm, E. J. et Hanage, W. P. (2013). Horizontal Gene Transfer and the Evolution of Bacterial and Archaeal Population Structure. *Trends in genetics, 29*(3), 170-175. https://doi.org/10.1016/j.tig.2012.12.006

Qiu, D., Huang, L., Liu, X. et Lin, S. (2019). Flourishing deep-sea AAP bacteria detected by flow cytometric sorting and molecular analysis. *PLOS One, 14*(6), e0218753. https://doi.org/10.1371/journal.pone.0218753

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. et Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology, 35*(9), 833-844. https://doi.org/10.1038/nbt.3935

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Ralser, M., Wamelink, M. M., Kowald, A., Gerisch, B., Heeren, G., Struys, E. A., Klipp, E., Jakobs, C., Breitenbach, M., Lehrach, H. et Krobitsch, S. (2007). Dynamic rerouting of the carbohydrate flux is key to counteracting oxidative stress. *Journal of Biology, 6*(4), 10. https://doi.org/10.1186/jbiol61

Ranjan, R., Rani, A., Metwally, A., McGee, H. S. et Perkins, D. L. (2016). Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochemical and Biophysical Research Communications, 469*(4), 967-977. https://doi.org/10.1016/j.bbrc.2015.12.083

Redford, A. J., Bowers, R. M., Knight, R., Linhart, Y. et Fierer, N. (2010). The ecology of the phyllosphere: geographic and phylogenetic variability in the distribution of bacteria on tree leaves: Biogeography of phyllosphere bacterial communities. *Environmental Microbiology, 12*(11), 2885-2893. https://doi.org/10.1111/j.1462-2920.2010.02258.x

Redford, A. J. et Fierer, N. (2009). Bacterial Succession on the Leaf Surface: A Novel System for Studying Successional Dynamics. *Microbial Ecology, 58*(1), 189-198. https://doi.org/10.1007/s00248-009-9495-y

Remus-Emsermann, M. N. P. et Schlechter, R. O. (2018). Phyllosphere microbiology: at the interface between microbial individuals and the plant host. *New Phytologist, 218*(4), 1327-1333. https://doi.org/10.1111/nph.15054

Reynolds, G. T. et Lutz, R. A. (2001). Sources of light in the deep ocean. *Reviews of Geophysics, 39*(1), 123-136. https://doi.org/10.1029/1999RG000071

Ritpitakphong, U., Falquet, L., Vimoltust, A., Berger, A., Métraux, J.-P. et L'Haridon, F. (2016). The microbiome of the leaf surface of Arabidopsis protects against a fungal pathogen. *New Phytologist, 210*(3), 1033-1043. https://doi.org/10.1111/nph.13808

Robson, J., McKenzie, J. L., Cursons, R., Cook, G. M. et Arcus, V. L. (2009). The vapBC Operon from Mycobacterium smegmatis Is An Autoregulated Toxin–Antitoxin Module That Controls Growth via Inhibition of Translation. *Journal of Molecular Biology, 390*(3), 353-367. https://doi.org/10.1016/j.jmb.2009.05.006

Rodríguez-Rodríguez, J. C., Fenton, N. J., Bergeron, Y. et Kembel, S. W. (2023). Soil and tree phyllosphere microbial communities differ between coniferous and broadleaf deciduous boreal forests. *Plant and Soil, 488*(1-2), 233-253. https://doi.org/10.1007/s11104-023-05959-y

Romera, F. J., García, M. J., Lucena, C., Martínez-Medina, A., Aparicio, M. A., Ramos, J., Alcántara, E., Angulo, M. et Pérez-Vicente, R. (2019). Induced Systemic Resistance (ISR) and Fe Deficiency Responses in Dicot Plants. *Frontiers in Plant Science, 10*, 287. https://doi.org/10.3389/fpls.2019.00287

Roper, M. C. (2011). Pantoea stewartii subsp. stewartii: lessons learned from a xylem-dwelling pathogen of sweet corn. *Molecular Plant Pathology, 12*(7), 628-637. https://doi.org/10.1111/j.1364-3703.2010.00698.x

Röttjers, L. et Faust, K. (2018). From hairballs to hypotheses–biological insights from microbial networks. *FEMS Microbiology Reviews, 42*(6), 761-780. https://doi.org/10.1093/femsre/fuy030

Roudier, F., Gissot, L., Beaudoin, F., Haslam, R., Michaelson, L., Marion, J., Molino, D., Lima, A., Bach, L., Morin, H., Tellier, F., Palauqui, J.-C., Bellec, Y., Renne, C., Miquel, M., DaCosta, M., Vignard, J., Rochat, C., Markham, J. E., … Faure, J.-D. (2010). Very-Long-Chain Fatty Acids Are Involved in Polar Auxin Transport and Developmental Patterning in Arabidopsis. *The Plant Cell, 22*(2), 364-375. https://doi.org/10.1105/tpc.109.071209

Rui, B., Shen, T., Zhou, H., Liu, J., Chen, J., Pan, X., Liu, H., Wu, J., Zheng, H. et Shi, Y. (2010). A systematic investigation of Escherichia coli central carbon metabolism in response to superoxide stress. *BMC Systems Biology, 4*(1), 122. https://doi.org/10.1186/1752-0509-4-122

Ruinen, J. (1956). Occurrence of Beijerinckia Species in the 'Phyllosphere'. *Nature, 177*(4501), 220-221. https://doi.org/10.1038/177220a0

Ryan, R. P., Vorhölter, F.-J., Potnis, N., Jones, J. B., Van Sluys, M.-A., Bogdanove, A. J. et Dow, J. M. (2011). Pathogenomics of Xanthomonas: understanding bacterium–plant interactions. *Nature Reviews Microbiology, 9*(5), 344-355. https://doi.org/10.1038/nrmicro2558

Salka, I., Čuperová, Z., Mašín, M., Koblížek, M. et Grossart, H.-P. (2011). Rhodoferax-related pufM gene cluster dominates the aerobic anoxygenic phototrophic communities in German freshwater lakes: Diversity of AAPs in freshwater lakes. *Environmental Microbiology, 13*(11), 2865-2875. https://doi.org/10.1111/j.1462-2920.2011.02562.x

Sandhu, A., Halverson, L. J. et Beattie, G. A. (2007). Bacterial degradation of airborne phenol in the phyllosphere. *Environmental Microbiology, 9*(2), 383-392. https://doi.org/10.1111/j.1462-2920.2006.01149.x

Sanjenbam, P., Shivaprasad, P. V. et Agashe, D. (2022). Impact of Phyllosphere Methylobacterium on Host Rice Landraces. *Microbiology Spectrum, 10*(4), e00810-22. https://doi.org/10.1128/spectrum.00810-22

Santos, A. L., Gomes, N. C. M., Henriques, I., Almeida, A., Correia, A. et Cunha, Â. (2012). Contribution of reactive oxygen species to UV-B-induced damage in bacteria. *Journal of Photochemistry and Photobiology B: Biology, 117*, 40-46. https://doi.org/10.1016/j.jphotobiol.2012.08.016

Santos, A. L., Oliveira, V., Baptista, I., Henriques, I., Gomes, N. C. M., Almeida, A., Correia, A. et Cunha, Â. (2013). Wavelength dependence of biological damage induced by UV radiation on bacteria. *Archives of Microbiology, 195*(1), 63-74. https://doi.org/10.1007/s00203-012-0847-5

Sapp, A., Huguet-Tapia, J. C., Sánchez-Lamas, M., Antelo, G. T., Primo, E. D., Rinaldi, J., Klinke, S., Goldbaum, F. A., Bonomi, H. R., Christner, B. C. et Otero, L. H. (2018). Draft Genome Sequence of Methylobacterium sp. Strain V23, Isolated from Accretion Ice of the Antarctic Subglacial Lake Vostok. *Genome Announcements, 6*(10), e00145-18. https://doi.org/10.1128/genomeA.00145-18

Schäfer, M., Vogel, C. M., Bortfeld-Miller, M., Mittelviefhaus, M. et Vorholt, J. A. (2022). Mapping phyllosphere microbiota interactions in planta to establish genotype–phenotype relationships. *Nature Microbiology, 7*(6), 856-867. https://doi.org/10.1038/s41564-022-01132-w

Scheublin, T. R., Deusch, S., Moreno-Forero, S. K., Müller, J. A., van der Meer, J. R. et Leveau, J. H. J. (2014). Transcriptional profiling of Gram-positive rthrobacter in the phyllosphere: induction of pollutant degradation genes by natural plant phenolic compounds. *Environmental Microbiology, 16*(7), 2212-2225. https://doi.org/10.1111/1462-2920.12375

Schloissnig, S., Arumugam, M., Sunagawa, S., Mitreva, M., Tap, J., Zhu, A., Waller, A., Mende, D. R., Kultima, J. R., Martin, J., Kota, K., Sunyaev, S. R., Weinstock, G. M. et Bork, P. (2013). Genomic variation landscape of the human gut microbiome. *Nature, 493*(7430), 45-50. https://doi.org/10.1038/nature11711

Schloss, P. D. (2024). Rarefaction is currently the best approach to control for uneven sequencing effort in amplicon sequence analyses. *mSphere, 9*(2), e00354-23. https://doi.org/10.1128/msphere.00354-23

Schmidt, O., Schuenemann, V. J., Hand, N. J., Silhavy, T. J., Martin, J., Lupas, A. N. et Djuranovic, S. (2007). prlF and yhaV encode a new toxin-antitoxin system in Escherichia coli. *Journal of molecular biology, 372*(4), 894-905. https://doi.org/10.1016/j.jmb.2007.07.016

Scholz, M., Ward, D. V., Pasolli, E., Tolio, T., Zolfo, M., Asnicar, F., Truong, D. T., Tett, A., Morrow, A. L. et Segata, N. (2016). Strain-level microbial epidemiology and population genomics from shotgun metagenomics. *Nature Methods, 13*(5), 435-438. https://doi.org/10.1038/nmeth.3802

Schönherr, J. et Baur, P. (1996). Cuticle Permeability Studies: A Model for Estimating Leaching of Plant Metabolites to Leaf Surfaces. Dans C. E. Morris, P. C. Nicot et C. Nguyen (dir.), *Aerial plant surface microbiology: Proceedings of the sixth international symposium held in bandol, france, september 11-15, 1995* (p. 1-24). Springer.

Schreiber, L., Krimm, U., Knoll, D., Sayed, M., Auling, G. et Kroppenstedt, R. M. (2005). Plant–microbe interactions: identification of epiphytic bacteria and their ability to alter leaf surface permeability. *New Phytologist, 166*(2), 589-594. https://doi.org/10.1111/j.1469-8137.2005.01343.x

Schweiger, R., Castells, E., Da Sois, L., Martínez-Vilalta, J. et Müller, C. (2021). Highly Species-Specific Foliar Metabolomes of Diverse Woody Species and Relationships with the Leaf Economics Spectrum. *Cells, 10*(3), 644. https://doi.org/10.3390/cells10030644

Segal, G. et Ron, E. Z. (1998). Regulation of Heat-Shock Response in Bacteria. *Annals of the New York Academy of Sciences, 851*(1), 147-151. https://doi.org/doi:10.1111/j.1749-6632.1998.tb08988.x

Shakir, S., Zaidi, S. S.-A., de Vries, F. T. et Mansoor, S. (2021). Plant Genetic Networks Shaping Phyllosphere Microbial Community. *Trends in Genetics, 37*(4), 306-316. https://doi.org/10.1016/j.tig.2020.09.010

Shapiro, B. J. (2014). Signatures of Natural Selection and Ecological Differentiation in Microbial Genomes. Dans C. R. Landry et N. Aubin-Horth (dir.), *Ecological Genomics* (vol. 781, p. 339-359). Springer Netherlands. https://doi.org/10.1007/978-94-007-7347-9_17

Shapiro, B. J. (2017). The population genetics of pangenomes. *Nature Microbiology, 2*(12), 1574-1574. https://doi.org/10.1038/s41564-017-0066-6

Shapiro, B. J., Friedman, J., Cordero, O. X., Preheim, S. P., Timberlake, S. C., Szabó, G., Polz, M. F. et Alm, E. J. (2012). Population genomics of early events in the ecological differentiation of bacteria. *Science, 336*(6077), 48-51. https://doi.org/10.1126/science.1218198

Shapiro, B. J., Leducq, J.-B. et Mallet, J. (2016). What Is Speciation? *PLOS Genetics, 12*(3), e1005860. https://doi.org/10.1371/journal.pgen.1005860

Shapiro, B. J. et Polz, M. F. (2015). Microbial Speciation. *Cold Spring Harbor Perspectives in Biology, 7*(10), a018143. https://doi.org/10.1101/cshperspect.a018143

Shenhav, L. et Zeevi. (2020). Resource conservation manifests in the genetic code. *Science, 370*(6517), 683-687. https://doi.org/10.1126/science.aaz964

Sheppard, S. K., Guttman, D. S. et Fitzgerald, J. R. (2018). Population genomics of bacterial host adaptation. *Nature Reviews Genetics, 19*(9), 549-565. https://doi.org/10.1038/s41576-018-0032-z

Shiraishi, K., Oku, M., Kawaguchi, K., Uchida, D., Yurimoto, H. et Sakai, Y. (2015). Yeast nitrogen utilization in the phyllosphere during plant lifespan under regulation of autophagy. *Scientific Reports, 5*, 9719. https://doi.org/10.1038/srep09719

Siegenthaler, A., Skidmore, A. K., de Groot, G. A., Laros, I., Rousseau, M. et Duan, Y. (2024). Temperate tree microbiomes: divergent soil and phyllosphere microbial communities share few but dominant taxa. *Plant and Soil, 496*(1), 319-340. https://doi.org/10.1007/s11104-023-06364-1

Simon, M., Scheuner, C., Meier-Kolthoff, J. P., Brinkhoff, T., Wagner-Döbler, I., Ulbrich, M., Klenk, H.-P., Schomburg, D., Petersen, J. et Göker, M. (2017). Phylogenomics of Rhodobacteraceae reveals evolutionary adaptation to marine and non-marine habitats. *The ISME Journal, 11*(6), 1483-1499. https://doi.org/10.1038/ismej.2016.198

Skovran, E., Raghuraman, C. et Martinez-Gomez, N. C. (2019). Lanthanides in Methylotrophy. Dans L. Chistoserdova (dir.), *Methylotrophs and Methylotroph Communities* (p. 101-116). https://doi.org/10.21775/cimb.033.101

Slomka, S., Françoise, I., Hornung, G., Asraf, O., Biniashvili, T., Pilpel, Y. et Dahan, O. (2020). Experimental Evolution of Bacillus subtilis Reveals the Evolutionary Dynamics of Horizontal Gene Transfer and Suggests Adaptive and Neutral Effects. *Genetics, 216*(2), 543-558. https://doi.org/10.1534/genetics.120.303401

Smets, W., Wuyts, K., Oerlemans, E., Wuyts, S., Denys, S., Samson, R. et Lebeer, S. (2016). Impact of urban land use on the bacterial phyllosphere of ivy (Hedera sp.). *Atmospheric Environment, 147*, 376-383. https://doi.org/10.1016/j.atmosenv.2016.10.017

Smillie, C. S., Smith, M. B., Friedman, J., Cordero, O. X., David, L. A. et Alm, E. J. (2011). Ecology drives a global network of gene exchange connecting the human microbiome. *Nature, 480*(7376), 241-244. https://doi.org/10.1038/nature10571

Smith, J. A. et Magnuson, R. D. (2004). Modular Organization of the Phd Repressor/Antitoxin Protein. *Journal of Bacteriology, 186*(9), 2692-2698. https://doi.org/10.1128/JB.186.9.2692-2698.2004

Snipen, L. et Liland, K. H. (2015). micropan: an R-package for microbial pan-genomics. *BMC Bioinformatics, 16*(1), 79. https://doi.org/10.1186/s12859-015-0517-0

Sohrabi, R., Paasch, B. C., Liber, J. A. et He, S. Y. (2023). Phyllosphere Microbiome. *Annual Review of Plant Biology, 74*(1), 539-568. https://doi.org/10.1146/annurev-arplant-102820-032704

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*(9), 1312-1313. https://doi.org/10.1093/bioinformatics/btu033

Stewart, R. C., Roth, A. F. et Dahlquist, F. W. (1990). Mutations that affect control of the methylesterase activity of CheB, a component of the chemotaxis adaptation system in Escherichia coli. *Journal of Bacteriology, 172*(6), 3388-3399. https://doi.org/10.1128/jb.172.6.3388-3399.1990

Stiefel, P., Zambelli, T. et Vorholt, J. A. (2013). Isolation of Optically Targeted Single Bacteria by Application of Fluidic Force Microscopy to Aerobic Anoxygenic Phototrophs from the Phyllosphere. *Applied and Environmental Microbiology, 79*(16), 4895-4905. https://doi.org/10.1128/AEM.01087-13

Stockwell, V. O., Johnson, K. B., Sugar, D. et Loper, J. E. (2002). Antibiosis Contributes to Biological Control of Fire Blight by Pantoea agglomerans Strain Eh252 in Orchards. *Phytopathology, 92*(11), 1202-1209. https://doi.org/10.1094/PHYTO.2002.92.11.1202

Stone, B. W. G. et Jackson, C. R. (2019). Canopy position is a stronger determinant of bacterial community composition and diversity than environmental disturbance in the phyllosphere. *FEMS Microbiology Ecology, 95*(4), fiz032. https://doi.org/10.1093/femsec/fiz032

Stone, B. W. G. et Jackson, C. R. (2021). Seasonal Patterns Contribute More Towards Phyllosphere Bacterial Community Structure than Short-Term Perturbations. *Microbial Ecology, 81*(1), 146-156. https://doi.org/10.1007/s00248-020-01564-z

Stroud, J. T., Delory, B. M., Barnes, E. M., Chase, J. M., De Meester, L., Dieskau, J., Grainger, T. N., Halliday, F. W., Kardol, P., Knight, T. M., Ladouceur, E., Little, C. J., Roscher, C., Sarneel, J. M., Temperton, V. M., Van Steijn, T. L. H., Werner, C. M., Wood, C. W. et Fukami, T. (2024). Priority effects transcend scales and disciplines in biology. *Trends in Ecology & Evolution, 37*(7), 677-688. https://doi.org/10.1016/j.tree.2024.02.004

Sun, J. S., Mashruwala, A. A., Fei, C. et Bassler, B. L. (2021,). Bacterial LomR Induces the Vibriophage VP882 VqmA-Directed Quorum-Sensing Lysogeny-Lysis Transition [prépublication]. *bioRxiv*. https://doi.org/10.1101/2021.11.15.468771

Sundarraman, D., Hay, E. A., Martins, D. M., Shields, D. S., Pettinari, N. L. et Parthasarathy, R. (2020). Higher-Order Interactions Dampen Pairwise Competition in the Zebrafish Gut Microbiome. *mBio, 11*(5), e01667-20. https://doi.org/10.1128/mBio.01667-20

Tahon, G. et Willems, A. (2017). Isolation and characterization of aerobic anoxygenic phototrophs from exposed soils from the Sør Rondane Mountains, East Antarctica. *Systematic and Applied Microbiology, 40*(6), 357-369. https://doi.org/10.1016/j.syapm.2017.05.007

Tani, A. et Sahin, N. (2013). Methylobacteriumhaplocladii sp. nov. and Methylobacteriumbrachythecii sp. nov., isolated from bryophytes. *International Journal of Systematic and Evolutionary Microbiology, 63*(9), 3287-3292. https://doi.org/10.1099/ijs.0.048215-0

Tani, A., Sahin, N., Fujitani, Y., Kato, A., Sato, K. et Kimbara, K. (2015). Methylobacterium Species Promoting Rice and Barley Growth and Interaction Specificity Revealed with Whole-Cell Matrix-Assisted Laser Desorption / Ionization-Time-of-Flight Mass Spectrometry (MALDI-TOF/MS) Analysis. *PLOS One, 10*(6), e0129509. https://doi.org/10.1371/journal.pone.0129509

Tessler, M., Neumann, J. S., Afshinnekoo, E., Pineda, M., Hersch, R., Velho, L. F. M., Segovia, B. T., Lansac-Toha, F. A., Lemke, M., DeSalle, R., Mason, C. E. et Brugler, M. R. (2017). Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. *Scientific Reports, 7*(1), 6589. https://doi.org/10.1038/s41598-017-06665-3

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., Angiuoli, S. V., Crabtree, J., Jones, A. L., Durkin, A. S., DeBoy, R. T., Davidsen, T. M., Mora, M., Scarselli, M., Margarit y Ros, I., Peterson, J. D., Hauser, C. R., Sundaram, J. P., Nelson, W. C., … Fraser, C. M. (2005). Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial “pan-genome”. *Proceedings of the National Academy of Sciences, 102*(39), 13950-13955. https://doi.org/10.1073/pnas.0506758102

Tettelin, H. et Medini, D. (dir.). (2020). *The Pangenome: Diversity, Dynamics and Evolution of Genomes*. Springer International Publishing. https://doi.org/10.1007/978-3-030-38281-0

Tettelin, H., Riley, D., Cattuto, C. et Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology, 11*(5), 472-477. https://doi.org/10.1016/j.mib.2008.09.006

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., Prill, R. J., Tripathi, A., Gibbons, S. M., Ackermann, G., Navas-Molina, J. A., Janssen, S., Kopylova, E., Vázquez-Baeza, Y., González, A., Morton, J. T., Mirarab, S., Zech Xu, Z., Jiang, L., … Zhao, H. (2017). A communal catalogue reveals Earth’s multiscale microbial diversity. *Nature, 551*(7681), 457-463. https://doi.org/10.1038/nature24621

Tian, M., Yu, G., He, N. et Hou, J. (2016). Leaf morphological and anatomical traits from tropical to temperate coniferous forests: Mechanisms and influencing factors. *Scientific Reports, 6*(1), 19703. https://doi.org/10.1038/srep19703

Tilman, D. (1996). Biodiversity: Population Versus Ecosystem Stability. *Ecology, 77*(2), 350-363. https://doi.org/10.2307/2265614

Tilman, D., Isbell, F. et Cowles, J. M. (2014). Biodiversity and Ecosystem Functioning. *Annual Review of Ecology, Evolution, and Systematics, 45*(1), 471-493. https://doi.org/10.1146/annurev-ecolsys-120213-091917

Tilman, D., Reich, P. B. et Isbell, F. (2012). Biodiversity impacts ecosystem productivity as much as resources, disturbance, or herbivory. *Proceedings of the National Academy of Sciences, 109*(26), 10394-10397. https://doi.org/10.1073/pnas.1208240109

Tilman, D., Reich, P. B., Knops, J., Wedin, D., Mielke, T. et Lehman, C. (2001). Diversity and Productivity in a Long-Term Grassland Experiment. *Science, 294*(5543), 843-845. https://doi.org/10.1126/science.1060391

Timilsina, S., Potnis, N., Newberry, E. A., Liyanapathiranage, P., Iruegas-Bocardo, F., White, F. F., Goss, E. M. et Jones, J. B. (2020). Xanthomonas diversity, virulence and plant–pathogen interactions. *Nature Reviews Microbiology, 18*(8), 415-427. https://doi.org/10.1038/s41579-020-0361-8

Treangen, T. J. et Rocha, E. P. C. (2011). Horizontal Transfer, Not Duplication, Drives the Expansion of Protein Families in Prokaryotes. *PLOS Genetics, 7*(1), e1001284. https://doi.org/10.1371/journal.pgen.1001284

Truchado, P., Gil, M. I., Reboleiro, P., Rodelas, B. et Allende, A. (2017). Impact of solar radiation exposure on phyllosphere bacterial community of red-pigmented baby leaf lettuce. *Food Microbiology, 66*, 77-85. https://doi.org/10.1016/j.fm.2017.03.018

Tukey, H. B. (1970). The Leaching of Substances from Plants. *Annual Review of Plant Physiology, 21*(1), 305-324. https://doi.org/10.1146/annurev.pp.21.060170.001513

Urakami, T. et Komagata, K. (1984). Protomonas, a New Genus of Facultatively Methylotrophic Bacteria. *International Journal of Systematic Bacteriology, 34*(2), 188-201. https://doi.org/10.1099/00207713-34-2-188

Utter, D. R., Borisy, G. G., Eren, A. M., Cavanaugh, C. M. et Mark Welch, J. L. (2020). Metapangenomics of the oral microbiome provides insights into habitat adaptation and cultivar diversity. *Genome Biology, 21*, 293. https://doi.org/10.1186/s13059-020-02200-2

Valiente-Mullor, C., Beamud, B., Ansari, I., Francés-Cuesta, C., García-González, N., Mejía, L., Ruiz-Hueso, P. et González-Candelas, F. (2021). One is not enough: On the effects of reference genome for the mapping and subsequent analyses of short-reads. *PLOS Computational Biology, 17*(1), e1008678. https://doi.org/10.1371/journal.pcbi.1008678

Van Der Plas, F. (2019). Biodiversity and ecosystem functioning in naturally assembled communities. *Biological Reviews, 94*(4), 1220-1245. https://doi.org/10.1111/brv.12499

Van Dongen, S. (2008). Graph Clustering Via a Discrete Uncoupling Process. *SIAM Journal on Matrix Analysis and Applications, 30*(1), 121-141. https://doi.org/10.1137/040608635

Van Dongen, S. et Abreu-Goodger, C. (2012). Using MCL to extract clusters from networks. *Methods in Molecular Biology, 804*, 281-295. https://doi.org/10.1007/978-1-61779-361-5_15

Ventorino, V., Sannino, F., Piccolo, A., Cafaro, V., Carotenuto, R. et Pepe, O. (2014). Methylobacterium populi VP2: Plant Growth-Promoting Bacterium Isolated from a Highly Polluted Environment for Polycyclic Aromatic Hydrocarbon (PAH) Biodegradation. *The Scientific World Journal, 2014,* e931793. https://doi.org/10.1155/2014/931793

Vogel, C., Bodenhausen, N., Gruissem, W. et Vorholt, J. A. (2016). The Arabidopsis leaf transcriptome reveals distinct but also overlapping responses to colonization by phyllosphere commensals and pathogen infection with impact on plant health. *New Phytologist, 212*(1), 192-207. https://doi.org/10.1111/nph.14036

Vorholt, J. A. (2012). Microbial life in the phyllosphere. *Nature Reviews Microbiology, 10*(12), 828-840. https://doi.org/10.1038/nrmicro2910

Vorholt, J. A., Vogel, C., Carlström, C. I. et Müller, D. B. (2017). Establishing Causality: Opportunities of Synthetic Communities for Plant Microbiome Research. *Cell Host & Microbe, 22*(2), 142-155. https://doi.org/10.1016/j.chom.2017.07.004

Vos, M., Hesselman, M. C., Beek, T. A. te, Passel, M. W. J. van et Eyre-Walker, A. (2015). Rates of Lateral Gene Transfer in Prokaryotes: High but Why? *Trends in Microbiology, 23*(10), 598-605. https://doi.org/10.1016/j.tim.2015.07.006

Vos, M., Quince, C., Pijl, A. S., Hollander, M. de et Kowalchuk, G. A. (2012). A Comparison of rpoB and 16S rRNA as Markers in Pyrosequencing Studies of Bacterial Diversity. *PLOS One, 7*(2), e30600. https://doi.org/10.1371/journal.pone.0030600

Vu, H. N., Subuyuj, G. A., Vijayakumar, S., Good, N. M., Martinez-Gomez, N. C. et Skovran, E. (2016). Lanthanide-Dependent Regulation of Methanol Oxidation Systems in Methylobacterium extorquens AM1 and Their Contribution to Methanol Growth. *Journal of Bacteriology, 198*(8), 1250-1259. https://doi.org/10.1128/JB.00937-15

Wade, T. G., Riitters, K., Wickham, J. D. et Jones, K. B. (2003). Distribution and Causes of Global Forest Fragmentation. *Conservation Ecology, 7*(2), 7. https://doi.org/10.5751/ES-00530-070207

Wang, Y. et Chen, C. (2005). Mutation analysis of the flp operon in Actinobacillus actinomycetemcomitans. *Gene, 351*, 61-71. https://doi.org/10.1016/j.gene.2005.02.010

Wang, Z., Jiang, Y., Zhang, M., Chu, C., Chen, Y., Fang, S., Jin, G., Jiang, M., Lian, J.-Y., Li, Y., Liu, Y., Ma, K., Mi, X., Qiao, X., Wang, X., Wang, X., Xu, H., Ye, W., Zhu, L., … Kembel, S. W. (2023a). Diversity and biogeography of plant phyllosphere bacteria are governed by latitude-dependent mechanisms. *New Phytologist, 240*(4), 1534-1547. https://doi.org/10.1111/nph.19235

Wang, Z., Peng, D., Fu, C., Luo, X., Guo, S., Li, L. et Yin, H. (2023b). Pan-metagenome reveals the abiotic stress resistome of cigar tobacco phyllosphere microbiome. *Frontiers in Plant Science, 14*, 1248476. https://doi.org/10.3389/fpls.2023.1248476

Webber, J. B., Putnam, M., Serdani, M., Pscheidt, J. W., Wiman, N. G. et Stockwell, V. O. (2020). Characterization of isolates of Xanthomonas arboricola pv. corylina, the causal agent of bacterial blight, from Oregon hazelnut orchards. *Journal of Plant Pathology, 102*(3), 799-812. https://doi.org/10.1007/s42161-020-00505-6

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J. R., Vázquez-Baeza, Y., Birmingham, A., Hyde, E. R. et Knight, R. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome, 5*(1), 27. https://doi.org/10.1186/s40168-017-0237-y

Welch, R. A., Burland, V., Plunkett, G., Redford, P., Roesch, P., Rasko, D., Buckles, E. L., Liou, S.-R., Boutin, A., Hackett, J., Stroud, D., Mayhew, G. F., Rose, D. J., Zhou, S., Schwartz, D. C., Perna, N. T., Mobley, H. L. T., Donnenberg, M. S. et Blattner, F. R. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. *Proceedings of the National Academy of Sciences, 99*(26), 17020-17024. https://doi.org/10.1073/pnas.252529799

Wellner, S., Lodders, N. et Kämpfer, P. (2012). Methylobacterium cerastii sp. nov., isolated from the leaf surface of Cerastium holosteoides. *International Journal of Systematic and Evolutionary Microbiology, 62*(4), 917-924. https://doi.org/10.1099/ijs.0.030767-0

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York.

Wickham, H., Vaughan, D. et Girlich, M. (2023a). *tidyr: Tidy Messy Data, R package* (v1.3.0). https://CRAN.R-project.org/package=tidyr

Wickham, H., François, R., Henry, L., Müller, K. et Vaughan, D. (2023b*). dplyr: A Grammar of Data Manipulation* (v1.1.2). https://CRAN.R-project.org/package=dplyr

Wiedenbeck, J. et Cohan, F. M. (2011). Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiology Reviews, 35*(5), 957-976. https://doi.org/10.1111/j.1574-6976.2011.00292.x

Wilderman, P. J., Vasil, A. I., Martin, W. E., Murphy, R. C. et Vasil, M. L. (2002). Pseudomonas aeruginosa Synthesizes Phosphatidylcholine by Use of the Phosphatidylcholine Synthase Pathway. *Journal of Bacteriology, 184*(17), 4792-4799. https://doi.org/10.1128/JB.184.17.4792-4799.2002

Wilson, M. et Lindow, S. E. (1993). Interactions between the biological control agent Pseudomonas fluorescens A506 and Erwinia amylovora in pear blossoms. *Phytopathology, 83*, 117-123. https://doi.org/10.1094/Phyto-83-117

Wimley, W. C. (2003). The versatile β-barrel membrane protein. *Current Opinion in Structural Biology, 13*(4), 404-411. https://doi.org/10.1016/S0959-440X(03)00099-X

Wood, D. E., Lu, J. et Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology, 20*(1), 257. https://doi.org/10.1186/s13059-019-1891-0

Woods, L. C., Gorrell, R. J., Taylor, F., Connallon, T., Kwok, T. et McDonald, M. J. (2020). Horizontal gene transfer potentiates adaptation by reducing selective constraints on the spread of genetic variation. *Proceedings of the National Academy of Sciences, 117*(43), 26868-26875. https://doi.org/10.1073/pnas.2005331117

Wu, W., Du, K., Kang, X. et Wei, H. (2021). The diverse roles of cytokinins in regulating leaf development. *Horticulture Research, 8*(1), 1-13. https://doi.org/10.1038/s41438-021-00558-3

Xin, X.-F., Kvitko, B. et He, S. Y. (2018). Pseudomonas syringae: what it takes to be a pathogen. *Nature Reviews Microbiology, 16*(5), 316-328. https://doi.org/10.1038/nrmicro.2018.17

Xiong, C. et Lu, Y. (2022). Microbiomes in agroecosystem: Diversity, function and assembly mechanisms. *Environmental Microbiology Reports, 14*(6), 833-849. https://doi.org/10.1111/1758-2229.13126

Yadav, R. K. P., Karamanoli, K. et Vokou, D. (2005). Bacterial colonization of the phyllosphere of mediterranean perennial species as influenced by leaf structural and chemical features. *Microbial Ecology, 50*(2), 185-196. https://doi.org/10.1007/s00248-004-0171-y

Yashiro, E., Spear, R. N. et McManus, P. S. (2011). Culture-dependent and culture-independent assessment of bacteria in the apple phyllosphere: Apple phyllosphere bacteria. *Journal of Applied Microbiology, 110*(5), 1284-1296. https://doi.org/10.1111/j.1365-2672.2011.04975.x

Yim, W. J., Kim, K. Y., Lee, Y. W., Sundaram, S. P., Lee, Y. et Sa, T. M. (2014). Real time expression of ACC oxidase and PR-protein genes mediated by Methylobacterium spp. in tomato plants challenged with Xanthomonas campestris pv. vesicatoria. *Journal of Plant Physiology, 171*(12), 1064-1075. https://doi.org/10.1016/j.jplph.2014.03.009

Youseif, S. H., Abd El-Megeed, F. H., Humm, E. A., Maymon, M., Mohamed, A. H., Saleh, S. A. et Hirsch, A. M. (2021). Comparative Analysis of the Cultured and Total Bacterial Community in the Wheat Rhizosphere Microbiome Using Culture-Dependent and Culture-Independent Approaches. *Microbiology Spectrum, 9*(2), e00678-21. https://doi.org/10.1128/Spectrum.00678-21

Yuan, Z., Ye, J., Lin, F., Wang, X., Yang, T., Bi, B., Mao, Z., Fang, S., Wang, X., Hao, Z. et Ali, A. (2023). Relationships between Phyllosphere Bacterial Communities and Leaf Functional Traits in a Temperate Forest. *Plants, 12*(22), 3854. https://doi.org/10.3390/plants12223854

Yurimoto, H., Hirai, R., Matsuno, N., Yasueda, H., Kato, N. et Sakai, Y. (2005). HxlR, a member of the DUF24 protein family, is a DNA-binding protein that acts as a positive regulator of the formaldehyde-inducible hxlAB operon in Bacillus subtilis. *Molecular Microbiology, 57*(2), 511-519. https://doi.org/10.1111/j.1365-2958.2005.04702.x

Yurimoto, H. et Sakai, Y. (2023). Interaction between C1-microorganisms and plants: contribution to the global carbon cycle and microbial survival strategies in the phyllosphere. *Bioscience, Biotechnology, and Biochemistry, 87*(1), 1-6. https://doi.org/10.1093/bbb/zbac176

Yurkov, V. et Hughes, E. (2017). Aerobic Anoxygenic Phototrophs: Four Decades of Mystery. Dans P. C. Hallenbeck (dir.), *Modern Topics in the Phototrophic Prokaryotes* (p. 193-214). Springer International Publishing. https://doi.org/10.1007/978-3-319-46261-5_6

Zarraonaindia, I., Owens, S. M., Weisenhorn, P., West, K., Hampton-Marcell, J., Lax, S., Bokulich, N. A., Mills, D. A., Martin, G., Taghavi, S., van der Lelie, D. et Gilbert, J. A. (2015). The Soil Microbiome Influences Grapevine-Associated Microbiota. *mBio, 6*(2), e02527-14. https://doi.org/10.1128/mbio.02527-14

Zervas, A., Zeng, Y., Madsen, A. M. et Hansen, L. H. (2019). Genomics of Aerobic Photoheterotrophs in Wheat Phyllosphere Reveals Divergent Evolutionary Patterns of Photosynthetic Genes in Methylobacterium spp. *Genome Biology and Evolution, 11*(10), 2895-2908. https://doi.org/10.1093/gbe/evz204