UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ECOLOGICAL AND EVOLUTIONARY PERSPECTIVES ON MICROBIAL COMMUNITY ASSEMBLY

THESIS

SUBMITTED

IN PARTIAL FULFILLMENT

FOR THE Ph.D. IN BIOLOGY

BY

GENEVIÈVE LAJOIE

JULY 2020

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

PERSPECTIVES ÉCOLOGIQUES ET ÉVOLUTIVES SUR L'ASSEMBLAGE DES COMMUNAUTÉS MICROBIENNES

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN BIOLOGIE

PAR

GENEVIÈVE LAJOIE

JUILLET 2020

ACKNOWLEDGEMENTS

I would first and foremost like to thank Steve Kembel for making it possible for me to explore a new domain of ecology during this PhD and for giving me the freedom to develop new skills and knowledge in ways that were meaningful for me. I would next like to thank Tanya Handa for having the generosity to engage with me like a colleague throughout these years and being such an inspiring mentor. Thanks also to Dan Kneeshaw for his warm welcome into his lab community at UQAM, which made me feel valued and supported when I needed it most.

Several more people have shared their enthusiasm and passion for science with me these past few years and in doing so contributed to making me a better scientist. I would like to thank in particular Tonia de Bellis, with her insatiable curiosity and sharp questioning of assumptions that always pushed me forward, and Geneviève Bourret, who helped me through the last steps of lab work and managed to make it so interesting and enriching. I am also grateful to Lochin Brouillard, forever mindful of the subjectivity involved in any scientific research and also one of the most perceptive readers I know. Many thanks also to Christophe Malaterre and Antoine C. Dussault for stimulating and nourishing my interest in epistemology, thence enhancing my approach to functional ecology.

My next thanks are for the wonderful assistants I've had the pleasure to interact with in collecting data for my thesis: Ariane Lafrenière and Dominique Tardif. Not only were they self-motivated and dedicated in moving field and lab work forward, they were also delightful co-workers and friends. I also thank all members of the Kembel lab, the Kneeshaw lab and the QCBS student network who helped make my everyday experience of this PhD more purposeful and fun.

Finally, I would like to thank my partner Raphaël for his indefectible support and encouragements every step of the way. He contributed invaluable coaching in computer science that helped me become a more rounded scientist. His insatiable curiosity and bubbling creativity also fueled many vitalizing discussions through these past years, making me a better researcher and communicator, but also a better person. Sincere thanks to Robin and Julien, cinephiles and readers extraordinaire, who continue to amaze me year after year with their endless passions and the breadth of their knowledge. They have been extraordinary sources of motivation and inspiration in both work and life. My last thanks go to my family, my grand-mother and my dearest cousins for believing in me and encouraging me to make a path of my own in pursuing my interests and passions throughout my life. I've had the privilege to be surrounded by inquisitive, self-driven and hard-working people from an early age and I believe it has greatly contributed in making me the researcher I am today. I will always be grateful for that.

et ce visage est tellement venu longtemps et souvent
et à peine pourtant regardé mais toujours désiré pareillement
comme tous ces gens assis sur les banquettes de cuir
qui attendent un éblouissement définitif
quelque espace s'ouvrant
vers quelque paysage parfaitement accessible
et parfaitement beau

Marie Uguay
Autoportraits

TABLE OF CONTENTS

LIST OF FIGURES

Figure 4.3 Principal components analyses (PCoA) of bacterial community composition of the phyllosphere of 33 tree host species. Two PCoA were calculated using Bray-Curtis distances among samples, one using samples from all species (a-c) and the other using samples from the sugar maple host only (d-f). Each point represents a bacterial community sample from a host species at a given plot and site. Samples are color-coded by the taxonomic family of the host. Ellipses representing the standard error of the average of scores are drawn for host species for which he had at least 5 samples (panel a) and for sites (panels c,f). We fitted host traits and relative abundances, as well as mean annual temperature (MAT) onto each PCoA in a way that maximises the correlation between these variables and the configuration of the PCoA and plotted them on the graphs. We performed this fitting respectively for individual- (a,d), plot-level (b,e) and site-level (c,f) average host traits (black lines), as well as plot-level and site-level host relative abundances (lines colored by host taxonomic family). Only the host species occurring in more than 3 sites were used for plotted in site-level analyses. Species codes are the following: *Abies balsamea* (L.) [ABIBAL]; *Acer pensylvanicum* (L.) [ACEPEN]; *Acer rubrum* (L.) [ACERUB]; *Acer saccharum* (Marshall) [ACESAC], *Acer spicatum* (Lam.) [ACESPI], *Betula alleghaniensis* (Britt.) [BETALL], *Betula papyrifera* (Marshall) [BETPAP], *Corylus cornuta* (Marshall) [CORCOR], *Fagus grandifolia* (Ehrh.) [FAGGRA], *Fraxinus pennsylvanica* (Marshall) [FRAPEN], *Ostrya virginiana* (Mill.) K. Koch [OSTVIR], *Picea rubens* (Sarg.) [PICRUB], *Prunus pensylvanica* (L.f.) [PRUPEN], *Quercus rubra* (L.) [QUERUB], *Tilia americana* (L.) [TILAME]. Host trait codes are the following: specific leaf area ($mm^2 \cdot mg^{-1}$) [SLA]; wood density ($g \cdot cm^{-3}$) [Wood.dens]; leaf calcium concentration ($mg \cdot g^{-1}$) [Ca]; leaf phosphorus concentration ($mg \cdot g^{-1}$) [P]. Site codes are presented in Table 4.1. 104

Figure 4.4 Variation in specialization (DSI*) of bacterial phyllosphere communities among the most abundant tree host species (a) and among sites (b). The DSI* metric is a standardized score characterizing the phylogenetic range of hosts an ASV interacts with varies between -1 (fully generalist) and +1 (fully specialist ASVs). In panel a, data points are mean specialization scores of ASVs per host species per site, weighted by the ASV abundance on the host. We only show host species that were present in at least 3 sites. Different letters among taxonomic

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

α_{div}	Alpha diversity
ANOVA	Analysis of variance
ASV	Amplicon sequence variant
β_{div}	Beta diversity
BLAST	Basic Local Alignment Search Tool
Ca	Calcium
CSR	Competitive - Stress-tolerant - Ruderal
CTAB	Cetrimonium bromide
df	Degrees of freedom
DNA	Deoxyribonucleic acid
DSI	Distance-based specialization index
DSI*	Standardized distance-based specialization index
EDTA	Ethylenediaminetetraacetic acid
Fdis	Functional dispersion
K	Potassium
KEGG	Kyoto Encyclopedia of Genes and Genomes
LHS	Leaf - Height - Seed
MAT	Mean annual temperature
Mg	Magnesium
MPD	Mean phylogenetic distance
MSE	Mean squared error
P	Phosphorus
PCA	Principal components analysis
PCoA	Principal coordinates analysis
PCR	Polymerase chain reaction
PERMANOVA	Permutational multivariate analysis of variance
RNA	Ribonucleic acid
ROS	Reactive oxygen species
SLA	Specific leaf area

SWATH-MS	Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra
TCA	Tricarboxylic acid cycle
tRNA	Transfer RNA
Tukey's HSD test	Tukey's honest significant difference test
UTM	Universal Transverse Mercator
UV	Ultraviolet
WD	Wood density
γ_{div}	Gamma diversity

RÉSUMÉ

Comprendre la diversité de la vie microbienne est crucial pour aborder plusieurs questions fondamentales en écologie et en biologie évolutive. Par l'étude des microbes, nous commençons à évaluer l'universalité des processus écologiques responsables de la distribution des espèces via un beaucoup plus grand spectre de formes et de fonctions biologiques que considéré traditionnellement. Nous avons également une meilleure appréciation du rôle des interactions biotiques dans la coexistence et l'évolution d'espèces cooccurrentes de différents niveaux trophiques. Forte de plusieurs études reliant la composition des communautés microbiennes des humains et végétaux à la santé et la productivité de leur hôte, la recherche appliquée en écologie microbienne a également des impacts importants sur notre conception des phénotypes élargis des plantes et des animaux.

Notre compréhension des processus façonnant la vie microbienne a jusqu'à maintenant dépendu principalement de descriptions taxonomiques de populations et de communautés microbiennes. En écologie des macro-organismes, les approches de recherche basées sur les traits des organismes ont renseigné l'étude de la distribution des espèces et de leur évolution en donnant des explications mécanistiques pour les patrons observés. Ces approches demeurent toutefois sous-utilisées en écologie microbienne. Parmi les défis à surmonter, la très grande diversité des phénotypes microbiens rend notamment le choix des traits à l'étude difficile, tout en complexifiant leur interprétation fonctionnelle. Une autre frontière majeure dans notre compréhension de l'écologie microbienne concerne le rôle des hôtes dans l'assemblage de leur communautés microbiennes. Alors que de la variation dans la composition taxonomique des microbes a été documentée entre génotypes et espèces d'hôtes, nous comprenons encore peu les mécanismes sélectifs responsables de tels patrons. On ignore également la façon dont les associations hôte-microbes varient en fonction du contexte écologique dans lequel elles prennent place.

Ma recherche doctorale aborde deux objectifs principaux : 1) améliorer notre compréhension de la structure de la diversité microbienne entre les écosystèmes à l'aide d'approches basées sur les traits, 2) déterminer les processus écologiques et évolutifs responsables des associations entre bactéries et hôtes dans des écosystèmes forestiers.

À l'aide d'une revue critique de la littérature, mon premier objectif spécifique était d'évaluer les différentes façons dont les approches basées sur les traits pourraient contribuer à la science de l'écologie microbienne. Il visait également à souligner les défis et opportunités pour continuer de faire avancer l'utilisation et l'adoption de ces approches. Mon second objectif spécifique était de déterminer les grands axes de variation dans les traits bactériens au niveau mondial. Nous avons utilisé une méta-analyse de jeux de données métagénomiques et génomiques bactériens pour identifier des groupes de gènes fonctionnels expliquant le plus de variation à la fois entre les écosystèmes et entre les clades bactériens. Mon troisième objectif spécifique était d'évaluer les bases adaptatives du renouvellement taxonomique des bactéries entre espèces d'hôtes dans la forêt tropicale. Nous avons caractérisé la diversité fonctionnelle des communautés bactériennes des surfaces foliaires (i.e. phyllosphère) de 17 espèces d'arbres à l'aide de séquençage métagénomique et évalué la présence d'appariement dans les traits des hôtes et des bactéries. Nous avons également déterminé le rôle des hôtes comme filtres environnementaux bactériens en fonction de leurs traits et de leur phylogénie. Mon dernier objectif spécifique était d'évaluer comment l'assemblage bactérien de la phyllosphère sur leur hôte était influencé par la variation environnementale à large échelle, l'identité de l'hôte et la composition du voisinage de l'arbre hôte. Nous avons abordé cet objectif en collectant et analysant des échantillons microbiens de la phyllosphère de plus de 30 espèces d'arbres à travers un gradient latitudinal de 5 °C s'étendant le long d'une transition de la forêt de feuillus à la forêt boréale.

Nous montrons que l'utilisation des approches basées sur les traits représente un pas vers l'avant dans notre investigation des mécanismes d'adaptation microbienne aux gradients environnementaux, mais également dans la mise en place d'une pratique plus intégrative de l'écologie microbienne. Nous révélons ensuite la présence de stratégies écologiques majeures expliquant la variation dans les traits bactériens entre clades et écosystèmes, suggérant la présence de moteurs écologiques et évolutifs universels de la variation fonctionnelle bactérienne dans le monde. Nous apportons également des preuves empiriques d'appariement adaptatif entre les bactéries et leurs hôtes dans la phyllosphère d'espèces d'arbres tropicales, révélant un rôle significatif pour le phénotype des hôtes dans l'assemblage de leurs communautés bactériennes. Notre étude en forêt tempérée révèle de surcroît un effet de la composition de la communauté d'arbres avoisinante dans l'appariement entre les communautés bactériennes et leurs hôtes, menant à une homogénéisation de la composition bactérienne d'espèces d'hôtes cooccurrentes.

Cette thèse pose de robustes fondations pour un meilleur usage des traits bactériens et de l'hôte pour comprendre l'origine et le maintien de la diversité bactérienne à l'intérieur des écosystèmes et entre eux. Ce faisant, nous identifions un chemin pragmatique pour rendre l'écologie microbienne plus généralisable. En identifiant des

mécanismes par lesquels les hôtes organisent l'assemblage des communautés bactériennes sur leurs feuilles et celles d'autres espèces, nous supportons également l'hypothèse que la vie avec les hôtes représente un axe important de variation écologique chez les bactéries. Somme toute, par l'usage de diverses méthodologies de recherche, incluant une revue conceptuelle, une méta-analyse, et des études de terrain, cette recherche contribue des vues nouvelles et approfondies sur les moteurs écologiques et évolutifs de la diversité microbienne.

Mots clés : écologie microbienne, associations hôtes-symbiontes, diversité fonctionnelle, phyllosphère, métagénomique

ABSTRACT

Understanding the diversity of microbial life is crucial to addressing several fundamental questions in ecology and evolutionary biology. Through the study of microbes, we are starting to evaluate the universality of ecological processes driving species distributions across a much greater spectrum of biological forms and functions than traditionally considered. It has also improved our appreciation of the role for biotic interactions in driving coexistence and the evolution of co-occurring species across trophic levels. With numerous studies linking the composition of human and plant microbial communities to their host health or productivity, applied research in microbial ecology is further having tremendous impacts on our conception of plant and animal extended phenotypes.

Our understanding of processes shaping microbial life has so far relied mostly on the taxonomic description of microbial populations and communities. Research approaches based on the traits of organisms have informed the study of species distribution and evolution in macro-organismal ecology by providing mechanistic explanations for observed patterns. Still, these approaches remain underused in microbial ecology. Challenges to overcome include the extremely large diversity of microbial phenotypes, making the choice of study traits difficult and complexifying their functional interpretation. Another major frontier in our understanding of microbial ecology regards the role of hosts in shaping the assembly of their microbial communities. While variation in the taxonomic composition of microbes have been reported among host genotypes and species, there is still little understanding of the selective mechanisms driving such patterns. We also have little appreciation of the way host-microbe associations vary as a function of the ecological context in which they take place.

My thesis research addresses two main objectives: 1) to improve our understanding of the structure of microbial diversity across ecosystems using trait-based approaches, 2) to determine the ecological and evolutionary processes driving associations between bacteria and their hosts in forested ecosystems. Through a critical survey of the literature, my first specific objective was to evaluate the different ways in which trait-based approaches could contribute to the science of microbial ecology. It also aimed to outline challenges and opportunities in moving these approaches forward. My second specific objective was to determine the main axes of variation in bacterial traits

worldwide. We used a meta-analysis of metagenomic and genomic bacterial datasets to identify groups of functional genes that explained the most variation both among ecosystems and among bacterial clades. My third specific objective addressed the adaptive bases of bacterial taxonomic turnover among host species in a tropical forest. We characterized functional diversity of bacterial communities from the leaf surfaces (i.e. phyllosphere) of 17 tree species using metagenomic sequencing and evaluated the presence of trait matching between hosts and bacteria. We also determined the role of hosts as environmental filters on the traits of bacteria as a function of their traits and phylogeny. My last specific objective was to evaluate how phyllosphere bacterial assembly on their hosts was influenced by large-scale environmental variation, host identity and the composition of the host tree neighbourhood across the landscape. We addressed this objective by collecting and analyzing microbial phyllosphere samples from more than 30 tree species across a 5 °C latitudinal gradient spanning a transition from deciduous to boreal forest.

We show that the use of trait-based approaches represents a promising way forward in investigating mechanisms of microbial adaptation to environmental gradients, but also in attempting to build a more integrative practice of microbial ecology. We next uncover major bacterial ecological strategies across bacterial clades and ecosystems, suggesting the presence of universal ecological and evolutionary drivers of bacterial trait variation worldwide. We also provide empirical evidence for adaptive matching between bacteria and their hosts in the phyllosphere of tropical tree species, revealing a significant role of host phenotypes in structuring bacterial community assembly. Our study in the temperate forest further reveals an effect of the neighbouring host community in determining the match between bacterial communities and their hosts, observed as an homogenization of bacterial composition between co-occurring host species.

This thesis lays a strong foundation for an improved use of bacterial and host traits in understanding the origin and maintenance of bacterial diversity within and among ecosystems. By doing so, we point a much-needed way forward to making microbial ecology more generalizable. By identifying mechanisms by which hosts structure bacterial community assembly on their leaves and those of other tree species, we also provide evidence supporting life with hosts as an important axis of ecological variation in bacteria. Overall, through the use of diverse research methodologies, including a conceptual review, a meta-analysis, and field-based studies, this research contributes novel and comprehensive insights into the ecological and evolutionary drivers of microbial diversity.

Keywords : microbial ecology, host-symbiont associations, functional diversity, phyllosphere, metagenomics

INTRODUCTION

0.1 Making sense of microbial diversity

Microbes account for about 20 percent of the world's biomass and colonize practically all habitats on Earth (Bar-On et al. 2018). Their diversification over the course of billions of years of evolution is now reflected in the very large taxonomic and phylogenetic diversity of microbes observed both within and among ecosystems (Haggerty and Dinsdale 2017, Delgado-Baquerizo et al. 2018). Despite a recent accumulation of genetic data characterizing microbial life worldwide, we still have little understanding of the structure of that functional diversity and more generally how to use this functional data to address ecological questions. Particularly, much remains to be understood about the drivers of functional variation among microbial clades and taxa, and how this diversity contributes to microbial sorting across habitat types. These questions are especially prominent for host-associated microbial communities. While microbial taxa have been found in association with hosts and appear to interact with them in a way that affects either the symbiont or the host fitness or both (Provorov and Vorobyov 2009), the study of host-associated microbes is still largely restricted to single host species (e.g. humans, crops, livestock) and microbial species (e.g. pathogens) of economic importance. As a result, we are still mostly unable to evaluate to what extent hosts are contributing to the generation and maintenance of microbial diversity both within host communities and across the landscape.

In this thesis, I will be addressing two main objectives: 1) to improve our understanding of the structure of microbial diversity across ecosystems using trait-

based approaches, and 2) to determine the ecological and evolutionary processes driving associations between bacteria and their hosts in forested ecosystems. I specifically address the following questions. First, how can we use the study of microbial functions to better understand the structure of microbial diversity worldwide? Second, what are the adaptive mechanisms underlying microbial taxonomic variation among microbial clades and ecosystems? Third, what is the role of hosts in structuring functional diversity of bacterial communities in the tree phyllosphere? Fourth, how do host-symbiont associations vary across spatial scales and how are they determined by the abiotic and biotic features of their habitat? I will concentrate on the bacterial portion of microbial diversity in addressing these questions. With important roles in global nutrient cycling, bacteria represent especially valuable study organisms. The relative ease through which to describe their diversity using universal sequencing primers and the growing databases of bacterial gene sequences facilitating their identification also makes them a particularly tractable microbial group to study.

0.2 On the use of trait-based approaches in microbial ecology

As tokens of past and current selective constraints on the survival of organisms across spatially and temporally variable environments, phenotypes are instrumental to our understanding of species ecology (and evolution). The study of functional traits - the morphological, physiological or phenological characteristics of individuals that have consequences on their fitness (*sensu* Violle et al. 2007) – has been important for decades in other ecological fields such as plant ecology to inform our understanding of species distributions and interactions (McGill et al. 2006). Functional perspectives on the study of ecology have supplemented traditional approaches focusing on species identity as the primary unit of observation by improving the mechanistic understanding of species-environmental associations (Keddy 1992), refining predictions of species responses to environmental change across different ecosystems (e.g. Frenette-Dussault et al. 2013), and by facilitating comparison and generalization of ecological dynamics

across study species and systems (Westoby 1998, Cornwell et al. 2008, Handa et al. 2014). While phenotype-oriented studies of organismal distributions and interactions are common now in fields such as plant ecology, trait-based approaches to ecological questions are still largely underused in microbial ecology.

0.2.1 Challenges related to defining microbial phenotypes

Delays in the adoption of trait-based approaches for microbes may stem from unresolved questions regarding the choice of microbial traits that should be studied and the way these should be measured. Challenges that have impeded progress on these questions include the lack of a working definition of microbial functional traits, the diverging methodological traditions of microbial physiologists and geneticists, and the large diversity of forms and life habits of microbes. The definition and identification of ecologically and evolutionarily relevant microbial phenotypes brings about conceptual issues that did not pose as much of an issue in macro-organismal ecology. For example, in a biological domain where lateral gene transfer may occur between co-occurring species, and where complex aggregates of organisms may acquire emergent properties (e.g. biofilms), should phenotypes be considered properties of genes, individual cells, or communities? At which level should they be measured? The definition and quantification of fitness in microbes remains a challenge, particularly regarding the metrics that should be used and the biological scale at which it should be measured.

The diversity of microbial lifestyles is another facet of the challenge to the standardized use of functional trait approaches. It encompasses three distinct issues. First, a large diversity of traits may be of potential adaptive significance. Second, many potentially important microbial traits may not have been documented yet (Sberro et al. 2019). This problem is especially evident when trying to attribute functions to gene sequences for which associated proteins are rarely experimentally characterized, and in many cases do not sufficiently match any characterized protein to allow their

function to be predicted by homology (Danchin and Fang 2016, Price et al. 2018). Third, this diversity of traits has led to a diversity of ways by which researchers select traits to study (Ramírez-Flandes et al. 2019), leading to the study of potentially distinct gene families depending on the specific interests and expertise of a research group; examples include methane oxidation (Krause et al. 2014), glucose utilization (Morrissey et al. 2016), and nitrogen cycling (Nelson et al. 2016). In turn, we are witnessing a diversity of ways in which microbial traits are studied by different microbial scientists, from culture-based to gene-based approaches, among which direct correspondences have not often been established.

0.2.2 Next steps for trait-based approaches in microbial ecology: functional classifications

Ecological classification systems have been developed over several decades in ecology as a way to explain the evolution of life histories and think about fundamental constraints on the distribution of organisms. Among the first, the r-K selection spectrum distinguished between species that reproduce quickly and invest less in each offspring (r-selected), and species reproducing more slowly but with more investment in the success of each reproductive unit (K-selected) (Dobzhansky 1950). This framework for defining major trade-offs in life-history strategies was later used for building models of population dynamics which contributed to the development of the theory of island biogeography (MacArthur and Wilson 1967), and in understanding the role of density-dependent regulation in driving the evolution of resource-use and life-history traits (Reznick et al. 2002). Expanding on the r-K spectrum, plant ecologists later proposed the competitor – stress-tolerant – ruderal (CSR) classification system to explain species response to environmental variation (Grime 1977). The classification is based on the response to two types of environmental gradients, stress and disturbance, with (i) competitive species growing best in low-stress and low-disturbance habitats, (ii) stress-tolerant species thriving better under high stress and low disturbance, and (iii) ruderals being best adapted to low stress and high-disturbance habitats. While useful

conceptually, this framework has faced criticism for the difficulty with which plants and animals could be assigned to each of these categories outside of study in culture, making it difficult to use it routinely for testing ecological questions at large scales and across study systems (but see Li and Shipley 2017, Pierce et al. 2017). Therefore, the identification of measurable traits representing major ecological trade-offs across species and across space emerged as an important next step to improve the utility of ecological classification systems.

In plant ecology, the leaf-height-seed (LHS) scheme was consequently developed in answer to those considerations (Westoby 1998, Westoby et al. 2002). Three important axes of adaptive strategies that explained ecological trade-offs and variation among plant species at local and global scales were first identified. Three measurable phenotypic traits were respectively proposed to describe variation along each of these axes, namely: (i) specific leaf area (the area of the leaf divided by its dry mass) as a proxy of resource conservation strategies, (ii) height as a proxy for energy acquisition and response to disturbance strategies, and (iii) seed weight as a proxy for dispersal and colonization strategies. Upon identification of other axes important in explaining ecological variation among plant species, such as below-ground resource investment strategies (Li et al. 2017), this framework has been expanded through years (Díaz et al. 2016). It is now used widely as a basis for standardized measurements of plant functional traits across the globe (e.g. Pérez-Harguindeguy et al. 2013). These traits are also being increasingly used in building predictive models of ecosystem function (Cadotte 2017).

Similar functional classification schemes for microbes have recently been proposed in microbial ecology (Fierer et al. 2007, Ho et al. 2013), though with mixed success. While some have tried to apply functional classification schemes from other domains of study such as plant and animal ecology (e.g. Chagnon et al. 2013, Ho et al. 2013 with the CSR classification scheme), others have aimed at devising new

functional classifications based on phenotypic traits common to the microbial world, such as the complexity of the carbon chains that they can process (Martiny et al. 2015). Most experimental characterizations of microbial strategies have focused on a few well-studied microbial clades that could be grown on a variety of carbon substrates, mostly in aquatic and soil habitats (e.g. Lennon et al. 2012, Evans and Wallenstein 2014, Shapiro and Polz 2014); as a result, the extent to which they represent the major ecologically and evolutionarily informative axes of phenotypic variation in microbes remains unknown.

0.3 Role of hosts in driving microbial diversity across the landscape

Matches between hosts and symbionts are supported by several lines of evidence, including variation in community composition of symbionts among genotypes of a given host species or species of a given host genus (Clark et al. 2001, Bailey et al. 2005, Whitham et al. 2012, Pita et al. 2013). Trait-based approaches are beginning to help us understand the ecology and evolution of host-associated microbes, by identifying host traits that play a selective role in driving microbial distributions among hosts (Kembel et al. 2014), or more rarely in identifying drivers of co-diversification between hosts and symbionts (Zangerl and Berenbaum 2003, Anderson et al. 2010). Building on these approaches, in this thesis I am particularly interested in examining the role for plant hosts and their phenotypes in structuring microbial diversity on the aboveground parts of plants including leaves, a habitat commonly referred to as the phyllosphere.

0.3.1 The phyllosphere as a study system

Interactions between plants and their symbionts have mostly been studied in plant-herbivore and plant-pathogen systems, due to their historically negative impacts in agriculture and their simpler nature, often involving only one plant species and one

main symbiont species. As pairwise interactions become better understood and with the availability of more efficient computational tools, more complex relationships can be modelled and multiple symbionts studied at the same time, allowing a more realistic investigation of the structuring of interactions in nature. Still, the dynamics of plant-microbiome associations remain poorly understood, particularly with regards to mutualisms and commensalisms and the specific mechanisms by which these associations may be maintained through time and space.

The best-studied plant-symbiont systems are from agricultural soils, mostly focusing on organisms such as rhizobacteria and mycorrhizal fungi which have been found to have positive impacts on plant fitness (reviewed in Pineda et al. 2013) and have thus represented attractive opportunities for applied research. The extent to which such model organisms may be used for above-ground plant-microbial interactions is however debatable. Notably, the phyllosphere may be exposed to more variable and less stable environments (Hirano and Upper 2000, Lindow and Brandl 2003, Bringel and Couée 2015). Phyllosphere microbes are also likely to be subjected to very different environmental filters (e.g. exposure to UV rays or high temperatures from which soil microbes may be relatively sheltered) (Jacobs et al. 2005) and may therefore greatly differ in their ecology and evolution than soil microbes for example.

The phyllosphere is an immense microbial habitat worldwide, as large as 100,000,000 km^2 and harbouring more than 10^{26} bacterial cells (Lindow and Brandl 2003, Vorholt 2012), making it a particularly relevant host-associated habitat to study. Phyllosphere microbial communities include bacteria (Yang et al. 2001, Lindow and Brandl 2003), fungi (Inácio et al. 2002), viruses and archaea (Taffner et al. 2018), with bacterial members being the most abundant and diverse. Bacterial communities are mostly composed of members of the phyla Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria, spanning a large breadth of bacterial evolutionary history (Bulgarelli et al. 2013).

Phyllosphere bacteria and their hosts can influence each other's health through several mechanisms. Host plants are typically a source of carbon and nitrogen for the microbes on their leaves, as well as shelters from water stress (Lindow and Brandl 2003). Microbes may reciprocally influence plant health through nitrogen fixation (Moyes et al. 2016), the production of plant growth enhancers (e.g. auxin) (Cox et al. 2018), production of antimicrobial compounds and the induction of systemic resistance to improve plant defense against pathogens and herbivores (Carvalho and Castillo 2018). Phyllosphere microbes are also drawing interest for their potential applications for improving human health (e.g. through the production of antimicrobial components; Pérez et al. 2016), and for crop health (Kerdraon et al. 2019). Despite these interesting features, as for many host-associated microbes, we still understand little of the factors driving their distribution both across ecosystems and among hosts. In the second part of this thesis, I will thus investigate how hosts drive the ecology and distribution of the yet largely unknown phyllosphere, focusing on bacteria living on the outside of plant tissues (i.e. epiphytes).

0.3.2 Drivers of phyllosphere microbial community assembly in association with hosts

Relationships between the relative abundance of different microbial strains and the abiotic features of their environment have been reported for some time in microbiology and have been the guiding principle for the use of selective growth media in strain isolation and culture (Lagier et al. 2015). Such a trait-matching process is collectively referred to in ecology as species sorting. For example, culture media determine differential growth among bacterial strains, a striking example of which is heavy-metal contaminated soil matrices in which only a small set of microbial species may actually grow (Roane and Pepper 1999). Field-based studies of microbial biogeography are more recent (Bryant et al. 2008) and have been linked with the emergence of culture-free sequencing technologies to characterize microbial diversity without requiring culturing. Based on the use of sequencing approaches, climatic

features of the abiotic environment such as temperature and precipitation have been found to play a role in determining the dominant species in field microbial community across regional to continental spatial scales (Drenovsky et al. 2010, Garcia-Pichel et al. 2013, Chen and Shapiro 2015, Coleman-Derr et al. 2016).

The biotic environment of a microbial community, that is to say the organisms with which the microbial community interacts directly or indirectly, also plays a role in driving its composition. Thinking of plants as landscapes for micro-organisms, we can recognize them as forming potentially important and geographically variable selective gradients for microbes, with plant physical and physiological characteristics such as tissue nutrient concentration or growth form determining the resources available for symbionts to establish, grow and reproduce. Differences in leaf microbial community composition and diversity have been reported among plant species with very different types of leaves (broadleaf vs. coniferous) in temperate forests (e.g. Redford et al. 2010, Laforest-Lapointe et al. 2016). While a greater abundance of Proteobacteria and Firmicutes and an overall lower diversity of bacteria was detected on the thinner, less waxy angiosperm leaves, a greater abundance of Actinobacteria and Acidobacteria as well as a greater overall diversity of bacteria was recorded on the thicker and longer-lived leaves of gymnosperms (Laforest-Lapointe et al. 2016). Similar differentiation of leaf microbial communities among plant taxa have also been found in tropical forests (Kembel et al. 2014).

Various properties of leaves may influence the species composition of phyllosphere bacteria, and these may depend on the specific ecosystem in which the plants are found. Finkel and colleagues (2016) reported adaptations to the salt-secretion and high UV radiation at the leaf surface in bacterial members of desertic plants, while leaf water content was the primary environmental driver structuring epiphytic bacterial abundance in dry Mediterranean environments (Vokou et al. 2012). Leaf traits indicative of resource-use acquisition strategies in plants, such as leaf mass per area,

and the wood density versus growth-mortality trade-off explained the most variation in bacterial community composition among host species in the phyllosphere of tropical trees (Kembel et al. 2014).

While we know of bacterial adaptations to life in the phyllosphere and of host traits driving bacterial taxonomic turnover among hosts, few studies to date have combined information on both bacterial and host traits to understand the emergence of associations between bacterial and plant taxa. Identifying matches between traits of both partners would help understand selective processes through which hosts may influence bacterial community assembly.

0.4 Variation in host-symbiont matching across the landscape

As the abiotic and the biotic environment change through space, selective pressures on microbial communities and the match between the characteristics of the environment or the plant host should vary as well, with consequences for the evolution of microbes (Thompson 2005). Still, whether variation in microbial community assembly on their plant host among sites results from variation in environmental features of the landscape or variation in characteristics of individual hosts or the host community remains an open question.

0.4.1 Influence of sampling site and the abiotic environment

The importance of sampling site in explaining variation in microbial community composition across space has been evaluated in several habitats. On the one hand, there is evidence for large-scale circulation of microbes around the globe; cross-continental winds have been shown to move microbes directionally from Africa to North America (DeLeon-Rodriguez et al. 2013), and between the Sahara and the Pyrenees (Barberan et al. 2014). Evidence for global circulation of microbes in the troposphere has also been uncovered, though being limited to a number of resistant

microbial species (DeLeon-Rodriguez et al. 2013). On the other hand, decreases in compositional similarity among microbial communities with increases in their geographic separation (i.e. distance-decay patterns) have been detected in some systems, indicating dispersal limitation could play a role in structuring microbial community assembly (e.g. Barberán et al. 2015) among sites.

Environmental variation can also explain microbial variation among sites, though the strength of environmental variables relative to geographical location depends on the study. In vineyard soil microbial communities, Morrison-Whittle & Goddard (2015) found that selection from the local environment explained as much as four times more variation in the composition of fungal communities than geographic location, though both factors were significant drivers of assembly. Sampling site however explained more variation among endophytic bacterial communities of the pine phyllosphere than climatic variation (Firrincieli et al. 2020). Such differences among studies could be due to the type of organisms considered and their dispersal capacities, but also on the spatial structure of the microbial habitat studied, providing microbes a spatially continuous or discontinuous matrix for dispersal.

0.4.2 The spatially structured plant landscape

Despite growing interest in understanding mechanisms of bacterial community assembly across space, it is still unclear what role there is for variation in biotic features of the landscape in driving bacterial turnover among sites relative to abiotic variables such as temperature or humidity. Studies to date have mostly touched on these questions by comparing the relative influence of host identity versus site in explaining microbial community similarity between samples (Knief et al. 2010, Redford et al. 2010, Coleman-Derr et al. 2016, Laforest-Lapointe et al. 2016) and have shown variable importance of species and species-site interactions in driving microbial community assembly across large sampling areas. Yet, whether dynamics of filtering and matching between bacteria and their hosts may vary across the landscape as a

function of the characteristics of the hosts and the host communities remains to be investigated.

The spatial structure of plants in the landscape determine patterns of dispersal and gene flow between populations of symbionts and their plant hosts. For example, the relative abundance of a particular plant host in the landscape should influence the likelihood of an associated microbe encountering that plant, leading to a positive relationships between landscape abundance of a host and the strength of its interactions with that symbiont (Kuussaari et al. 2001, Agrawal 2006). In laboratory experiments, Kuussaari and colleagues (2001) found a link between the heterogeneity of host plant species in the landscape and the degree of specialization or generalization of one of their shared parasites. The diversity of host tree species plantations has similarly been found to be linked with the diversity of their associated arthropod communities (Setiawan et al. 2016).

The connectivity of host populations in the plant landscape may also determine the extent of symbiont homogenization among co-occurring host species. In a mesocosm experiment on microarthropod meta-community dynamics in moss habitats, spatial connectivity among moss patches affected diversity observed both among and across arthropod communities (Chisholm et al. 2011). When patches were more tightly connected, more frequent exchanges between communities tended to homogenize species composition among sites, while when they were further apart, individual patches tended to diverge more between one another. Similarly, high levels of immigration from the surrounding terrestrial landscape have been documented to affect community composition of freshwater networks within the boreal forest (Ruiz-González et al. 2015). Acting as environmental filters or spatially structured microbial pools, plant hosts as part of a plant landscape therefore have multiple means of impacting the ecology and evolution of their associated organisms.

The structure of the plant landscape is also likely to affect the occurrence and evolution of symbiont specialization on their hosts. When selection is strong and opportunities to interact with alternative partners are low, given enough time local adaptation should occur and favor specialization, even in species rich communities (Thompson 2005). As selection from host individuals decreases and/or gene flow increases in the landscape and thus increases the probability of symbionts encountering functionally diverse hosts, we expect host-symbiont relationships to loosen and symbionts to become more generalist. Symbiont specialization may also vary along abiotic selection gradients (Wolinska and King 2009). First, along a productivity gradient, specialization is expected to be stronger when resources are scarce, such that benefits from the mutualisms may be higher (Callaway et al. 2002, Pineda et al. 2013). In more productive environments, resources may be plentiful enough that even more inefficient species may perform well enough to survive on their own (Thrall et al. 2007). Evidence for stronger mutualistic interactions under water stress have been documented in plant-mycorrhizal systems (O'Brien et al. 2018), though the breakdown of such mutualisms have also been reported under extreme nutrient limitations (Treseder and Allen 2002). Second, strong spatially variable selection in a heterogeneous landscape is expected to lead to greater specialization because of the fitness costs imposed to the evolution of generalization in the presence of limits to plasticity (Rainey and Travisano 1998, Markussen and Marvig 2014). With much of the theory on symbiont specialization having been developed in plant-pollinator systems or host-parasite systems, we still have little understanding of how these processes could influence the match between diverse communities of microbes and their hosts.

0.5 Presentation of this thesis

In this thesis, I address four main questions. First, how can we use the study of microbial functions to better understand the structure of microbial diversity worldwide?

Second, what are the adaptive mechanisms underlying microbial taxonomic variation among microbial clades and ecosystems? Third, what is the role of hosts in structuring functional diversity of bacterial communities in the phyllosphere of tree species? Fourth, how does host-symbiont associations vary across spatial scales and how is it determined by the abiotic and biotic features of their habitat? I addressed these questions using a diversity of research approaches.

0.5.1 Chapter 1 − Trait-based approaches in microbial ecology

The first chapter is an extensive conceptual investigation of the way trait-based approaches have been used in microbial ecology and how they could potentially be improved for ameliorating our ecological knowledge of microbes. We first present the different roles that trait-based approaches are occupying in improving ecological knowledge using a framework derived from epistemology, a branch of philosophy that investigates the nature of knowledge, as well as whether and how it can be obtained (Williams 2001). Using comparisons from the macro-organismal literature, we then outline the main challenges that the adoption and use of such approaches have been facing in microbial ecology and conclude with four main recommendations in answering these challenges and making the most of the study of traits in microbial ecology. This chapter was published in 2019 as an Opinion piece in *Trends in Microbiology*.

0.5.2 Chapter 2 − Fundamental bacterial strategies across clades and ecosystems

Building on conclusions from Chapter 1, the second chapter is a meta-analysis of worldwide bacterial functional gene data aimed at identifying the principal axes of variation in bacterial traits worldwide. Finding such universal axes has two general motivations regarding the development of ecological knowledge, namely to reduce dimensionality in the potentially large diversity of traits exhibited by bacteria and to facilitate investigations and comparisons of microbial ecological dynamics and

distributions across scales and study systems. My rationale here is that these major trait syndromes will be uncovered by assessing axes of correlated trait variation that 1) contribute the most to variation among microbial communities across ecosystem types and 2) contribute the most to variation among microbial clades. These axes should reflect respectively recent (ecological sorting) and longer-term selection pressures. This chapter is currently (July 2020) under review at *Nature Ecology and Evolution*.

0.5.3 Chapter 3 − Adaptive matching between phyllosphere bacteria and their tree host

I empirically investigate the role of hosts in driving the assembly of their microbial community in Chapters 3 and 4. Chapter 3 first addresses whether taxonomic associations between phyllosphere bacteria and their tree hosts are driven by adaptive matching between the partners at a local scale. Using metagenomic sequencing, we first characterize the functional diversity of epiphytic phyllosphere bacterial communities from 17 tree species in a neotropical forest. We then test three main predictions. First, bacterial functions should vary among host plant species and be correlated with the functional traits of the hosts. Second, cophylogenetic associations between trees and bacteria should lead to phylogenetic signal in bacterial functions present on different plant hosts. Third, bacterial functions present on leaves should be filtered by the host, since conditions on the leaves of different host plants create a selection pressure on the functions of bacteria able to persist on those leaves. This chapter was published in 2020 in *Microbiome*.

0.5.4 Chapter 4 − Bacterial community assembly and specialization in a multi-host landscape

This final chapter addresses the drivers of phyllosphere bacterial composition and specialization among multi-species tree communities at large spatial scales. Using microbial phyllosphere samples collected from 33 tree host species across a large climatic gradient, we test whether community assembly of epiphytic phyllosphere

bacteria on a focal host species may be influenced by the co-occurrence of functionally different host species across the landscape. We predict that the microbiota of host species co-occurring with different host species be more similar to that of the other host species as opposed to host species that are occurring in essentially monospecific stands. We also hypothesize that the temperature stress gradient, host filtering strength and host functional diversity will affect the prevalence and strength of specialization of bacteria on their host across this heterogeneous landscape. This chapter is currently (July 2020) under review at *Ecological Monographs*.

CHAPTER 1

MAKING THE MOST OF TRAIT-BASED APPROACHES FOR MICROBIAL ECOLOGY

(Article published in 2019 in *Trends in Microbiology*, 27: 814-823)

Geneviève Lajoie and Steven W. Kembel

Département des Sciences Biologiques, Université du Québec à Montréal
141 Avenue du Président-Kennedy, Montréal, Canada, H2X 1Y4

1.1 Abstract

There is an increasing interest in applying trait-based approaches to microbial ecology, but the question of how and why to do it is still lagging behind. By anchoring our discussion of these questions in a framework derived from epistemology, we broaden the scope of trait-based approaches to microbial ecology from one oriented mostly around explanation towards one inclusive of the predictive and integrative potential of these approaches. We use case studies from macro-organismal ecology to concretely show how these goals for knowledge development can be fulfilled and propose clear directions, adapted to the biological reality of microbes, to make the most of recent advancements in the measurement of microbial phenotypes and traits.

1.2 Shifting paradigms : moving to trait-based ecology

Counts of individual organisms and species across space and time have provided valuable insights into the processes governing species distributions since ecology's early days (MacArthur and Wilson 1967, Jackson 2009), but in recent decades these approaches have been criticized for providing only a partial understanding of the adaptive mechanisms driving ecology and evolution. By focusing on the study of phenotypic characteristics that influence organismal fitness across environmental gradients regardless of species identity, trait-based ecology aims to provide mechanistic explanations to ecological patterns and more robust predictions of ecological dynamics and ecosystem function. Grounded in the long-lasting tradition of studying relationships between traits and fitness in evolutionary and population ecology it has in the last few decades been fueled by conceptual developments in the fields of plant and animal ecology (Calow 1987, Keddy 1992, Violle et al. 2007).

Thanks to the increasing availability of data on the diversity of microbial populations and communities, trait-based approaches to microbial ecology are gaining in popularity (Shafquat et al. 2014, Martiny et al. 2015, Rojo et al. 2017, Bahram et al. 2018, Heintz-Buschart and Wilmes 2018) (Box 1). Direct observations of microbial traits and indirect inferences based on genetic data are increasingly used for investigating fundamental ecological questions and have already contributed to the development of knowledge in microbial ecology. We examine these contributions below.

1.3 Trait-based approaches have expanded our understanding of microbial ecological processes

One of the most recognized roles of trait-oriented approaches to microbial ecology has been to provide mechanistic **explanations** (see Glossary) of ecological

patterns. Bacterial traits have served in identifying adaptive mechanisms important for survival across different types of environments (e.g. plant roots − Hartmann et al. 2009, Rowe et al. 2018; human organs − Huttenhower et al. 2012; sponge tissues − Kamke et al. 2013; soil − Leff et al. 2015, Malik et al. 2017). By analyzing the genomes of single cells of *Poribacteria*, Kamke and colleagues (2013) discovered metabolic pathways indicative of the ability to degrade chains of proteoglycans – important components of their sponge host tissues – thereby providing a mechanism by which these bacteria could survive in their host. A study of the functional genes of soil bacterial communities across a soil pH gradient revealed that adaptation to high-pH soils was characterized by a greater abundance of multiple transporters (e.g. ABC transporters) allowing a direct uptake of substrates and cofactors (Malik et al. 2017). Attention to microbial traits has also led to important advancements in understanding the consequences of organismal adaptations and interactions for ecosystem functioning and productivity (Lindström et al. 2010, Raes et al. 2011, Wallenstein and Hall 2012, Hall et al. 2018). Variation in the diversity of microbial traits based on functional genes found in metagenomic samples of ocean water explained shifts in the primary productivity of these communities across the globe, providing insight into the role of ocean microbes in sustaining global productivity (Raes et al. 2011).

Developing functional explanations for observed ecological patterns also has the benefit of providing mechanistic bases for the development of **corroboratory predictions** (*sensu* Maris et al. 2018), aimed at testing the validity of ecological hypotheses, models or theories. Traits have been used to develop predictions on the importance of different ecological and evolutionary drivers of community assembly through time and space (Severin et al. 2013, Staley et al. 2014). To distinguish the relative importance of selection and neutral processes in driving the assembly of microbial communities, researchers have compared the trait similarity of microbes living in the same community to communities composed from microbes whose traits were drawn randomly from across all samples. A trait similarity higher than expected

by chance in observed communities suggests selection on the traits of microbes in several systems (Burke et al. 2011, Shafquat et al. 2014, Staley et al. 2014).

Functional ecology also holds the further promise of integrating ecological data, methodologies and explanatory schemes across research groups and disciplines (see O'Malley 2013) – the operationalization of which also constitutes its greatest challenge. **Data integration** involves the creation and use of tools and standards for assembling and comparing data collected within and among taxa (Leonelli 2013), the analysis and interpretation of which helps improve understanding. Nowadays, it typically requires online infrastructure for standardizing and storing data to facilitate their use and interpretation by researchers of different backgrounds. Data integration has been one of the strengths of microbial ecology, having relied on the development of databases for storing, organizing and sharing large amounts of genetic data (Zhulin 2015, Nilsson et al. 2019). Benefitting from those infrastructures, phenotypic data and functional annotations of full genomes and metagenomes are now being added to existing or new databases, such that trait information is more readily retrievable and comparable (e.g. Nguyen et al. 2016, Chen et al. 2017, Basenko et al. 2018, Cornwell et al. 2018, Reimer et al. 2019). The growth of protein description databases has also helped develop more precise and accurate functional predictions (The UniProt Consortium 2017). Data integration in microbial functional ecology is lastly being fostered by the development of elaborate methodologies (e.g. Moretti et al. 2016), refined ontologies (e.g. Chibucos et al. 2014, The Gene Ontology Consortium 2019) and standardized pipelines (e.g. Keegan et al. 2016) for collecting and processing massive standardized trait data sets (see also Box 1). Such methodologies are further making the collection of data more uniform and comparable among research groups, facilitating generalization.

Methodological integration concerns the development and use of a range of methods for the study of a given ecological pattern or process. It is aimed at developing a multi-faceted understanding of the results that improves on using each method

individually (Leonelli 2013). The concurrent use of phenotypic microarrays and next-generation sequencing have for example been used to characterize the real-time functional capabilities of specific microbial taxa to understand adaptive mechanisms underlying their endophytic lifestyle (Blumenstein et al. 2015). The parallel sequencing of a microbial community's genomes and transcriptomes has similarly helped characterize differences between the **fundamental** and **realized niches** of these communities (Ofek-Lalzar et al. 2014, Rojo et al. 2017).

Finally, **explanatory integration** involves the use of combination of hypotheses or theories developed in other disciplines in a new area of research, which may or may not lead to theoretical unification (Leonelli 2013). While a call for explanatory integration in microbial ecology to foster ecological understanding was made more than a decade ago (Prosser et al. 2007), such types of integration are now just emerging. For example, Werner and colleagues (Werner et al. 2014) proposed a reapplication of market theory adapted from economics to provide explanations of cooperative behaviors in microbes by characterizing resource investment strategies (a key concept in functional ecology) across varying conditions. In order to partition the relative contributions of different processes carried on by microbial communities to dinitrogen production in a marine habitat (here anammox and denitrification), Reed and colleagues (Reed et al. 2014) adapted models of chemical dynamics developed in biogeochemistry to functional gene abundance data from environmental genomic studies. Comparing their model with experimental data, they were able to confirm a larger role for denitrification in N_2 production. This type of integration however remains rare.

When achieved via functional traits, explanation, prediction and integration may finally serve a further goal for the development of knowledge in ecology. They provide a foundation for the **generalization** of research results irrespective of taxonomic identity across the globe, facilitating the search for general laws, theory development

and the elaboration of large-scale predictive models. A world-wide comparison of the relative abundance of nitrogen-cycling pathways in soil microbial communities has for example revealed that while the abundance of nitrogen pathways tended to vary biogeographically as a function of C and N concentrations, their relative proportions tended to correlate across soil samples (Nelson et al. 2016). This observation supported the hypothesis that habitats in which microbes can successfully exploit one pathway will also support higher number of cells that can exploit other N pathways, possibly leading to faster nutrient cycling rates.

1.4 Opportunities and challenges in the study of functional microbial ecology

The various types of studies mentioned above provide examples of the opportunities for using traits in microbial ecology with the objective of improving ecological understanding. Specific opportunities provided by microbial study systems include their large variety of physiologies and resource-use strategies, providing a playground for the study of adaptive mechanisms and the eco-evolutionary generation of biological diversity. For example, the incorporation of organismal optimum temperatures and light intensities for growth, as well as their capacity for assimilating nitrate and metabolizing silica all contribute to improving models of community structure and predictions of ecosystem function and biogeography in marine phytoplankton (Follows et al. 2007). From integrative and pragmatic standpoints, microbial ecologists can also benefit from existing infrastructure developed for the sharing of trait data, as well as several free online platforms for standardizing the treatment and analysis of functional trait data (Dudhagara et al. 2015, Shi et al. 2019). This potential has however not yet been fully realized (see Fig. 1.1). We next examine current challenges in the implementation of microbial functional ecology and their consequences for the different aspects of knowledge development.

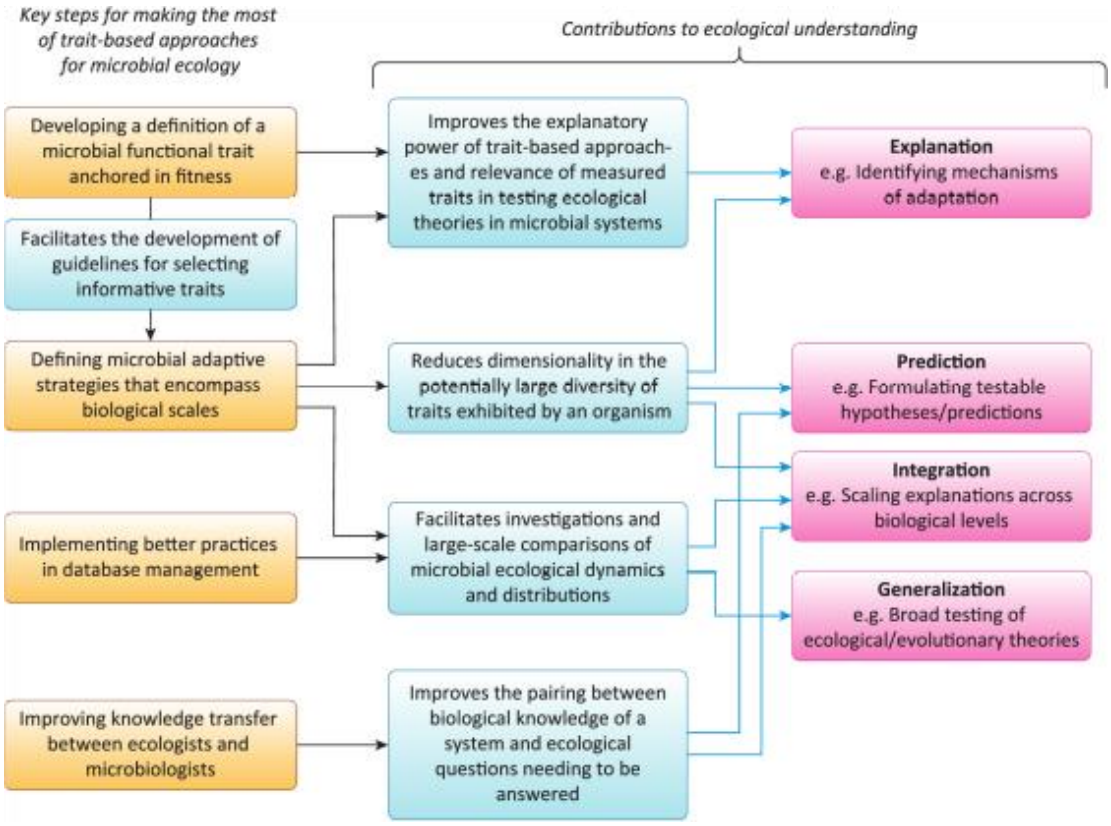

Figure 1.1 Key steps for trait-based approaches in improving understanding of microbial ecology. Each step can contribute to ecological understanding via different mechanisms, described in the blue boxes.

1.4.1 Lack of a working definition of a microbial functional trait

As much as scientific progress has been made by the use of traits in microbial ecology, individual studies have rarely defined the functional trait concept for microbes or explicitly linked traits to components of fitness as has been done for macro-organisms (but see Johnson et al. 2006, Lennon and Lehmkuhl 2016). This has limited the capacity of traits to identify adaptive mechanisms and the potential for explanatory

power. The lack of a standardized definition of microbial traits has further limited our possibility to compare results across trait-based studies, impacting the potential for integration. This issue may stem in part from the difficulty of applying existing concepts of functional traits developed in plant and animal ecology to the reality of microbial life (Box 2).

1.4.2 Large diversity of microbial lifestyles

The diversity of microbial lifestyles is another facet of the challenge to the standardized use of functional trait approaches. It encompasses three distinct issues. First, a large diversity of traits may be of potential adaptive significance. Second, many potentially important microbial traits may not have been documented yet (Sberro et al. 2019). This problem is especially evident when trying to attribute functions to gene sequences for which associated proteins are rarely experimentally characterized, and in many cases do not sufficiently match any characterized protein to allow their function to be predicted by homology (Danchin and Fang 2016, Price et al. 2018). Third, this diversity of traits has led to a diversity of ways by which researchers select traits to study (see Ramírez-Flandes et al. 2019), leading to the study of potentially distinct gene families depending on the specific interests and expertise of a research group; examples include methane oxidation (Krause et al. 2014), glucose utilisation (Morrissey et al. 2016), and nitrogen cycling (Nelson et al. 2016). In turn, we are witnessing a diversity of ways in which microbial traits are studied (see Box 1).

These situations have led to consistent context-dependence in interpreting ecological patterns and dynamics of microbial traits among systems, making integration and generalization more challenging among research groups and study systems. In part to address such concerns and also to account for the fact that the functions of many individual genes are still poorly understood, microbial traits are frequently classified and analyzed at broad levels, for example ‘metabolism’ or ‘cellular processes’ (see Box 1). This approach seems necessary given the inability to

characterize the precise function of many genes, but it limits inference of adaptive mechanisms driving differences in relative abundances of microbes across habitats. Finding the right level at which to aggregate microbial trait data to maximise explanatory power will require more investigation and improvement in our ability to annotate and classify gene functions (Österlund et al. 2017, Ramírez-Flandes et al. 2019).

The complexity of microbial trait measurement methods likewise constitutes a barrier to integration and generalization, promoting compartmentalization of research groups around technical specialties (e.g. metabolomics, metagenomics, proteomics – see Box 1). This compartmentalization is further encouraged by the prohibitive costs of some technologies, with costs for analysis running to hundreds or thousands of dollars per sample (e.g. SWATH-MS for assessing protein identity − Schilling et al. 2017), leading to technical limitations and cost determining methodology for many research groups. Finally, the analysis of high-dimensional microbial trait space data poses a technical challenge to traditional statistical analyses, both in terms of bioinformatics infrastructure and expertise requirements, and due to the expense associated with replication of samples (Johnstone and Titterington 2009).

1.4.3 Incomplete and biased databases

Annotation of microbial functional traits requires the use of reference databases, but current databases of microbial traits are phylogenetically incomplete and biased. These databases have far better representation of microbial taxa associated with certain ecosystems, in particular those microbes of importance for human health and those that can be grown in culture (Overmann et al. 2017). While these data are valuable for certain microbial systems such as the human microbiome, the extent to which the ecology of human-associated microbes matches that of their relatives in other environments limits the quality of the inference that can be drawn from using their genomes for predicting traits of microbes from environmental samples (Choi et al.

2016). Biases in the phylogenetic origin of microbes in trait databases further limits the investigation of general (i.e. context-independent) drivers of ecological and evolutionary dynamics in microbial communities.

Certain data management practices have also been limiting progress in integration. The general lack of metadata associated with sequence data complicates meta-analyses on fundamental ecological questions among studies and ecosystems. For example, different studies may define or measure different environmental variables when quantifying microbial functions. Data curation practices prior to integration in a database such as sample preservation conditions are often undescribed, despite the fact that these conditions can have significant impacts on the measured diversity and relative abundance of key taxa in the samples (Vandeputte et al. 2017).

1.4.4 Lack of clear ecological hypotheses in many trait-based studies

Initial studies on microbial traits have mostly focused on describing phenotypes and physiological mechanisms for understanding the biology of microorganisms and for identification of different microbial taxa in culture (Janda and Abbott 2002). While some have argued that microbial ecology is still in a 'discovery' phase where collection of data without specific tests of ecological hypothesis is normal or desirable (Tripathi et al. 2018), there have been ongoing calls by microbial ecologists to develop a more explicitly hypothesis-based science of microbial ecology in order to move forward (Jessup et al. 2004, Prosser et al. 2007). Some go further in arguing that microbial systems could actually represent ideal systems to test and expand on existing theory (Meyer and Leveau 2012, Bruns 2018). While researchers have been using trait-based approaches to identify linkages between microbial community structure and ecosystem processes, direct empirical tests of these predictions are still largely lacking (Bier et al. 2015, but see Glassman et al. 2018).

1.5 Next steps for trait-based approaches in microbial ecology

1.5.1 Developing a definition of a microbial functional trait

A first major step to be taken in improving the contribution of microbial functional trait approaches would be to develop a definition of a functional trait in the microbial realm. In particular, this definition would include identification of the units of selection that are being discussed and worked on. A clear definition of a microbial functional trait would facilitate the development of guidelines for selecting informative traits, which should consequently improve the explanatory power of trait-based approaches. We suggest adopting the definition of a functional trait in current use for plants and other macro-organisms – that a functional trait is any attribute of an organism that can be linked to its fitness (Violle et al. 2007). This adoption will require a shift from purely sequence-based approaches to quantifying microbial functions, towards incorporation of data from direct measurement of microbial population and cellular growth, performance and survival (e.g. monitoring of cell densities, biomass incorporation or respiration rates) (Bai et al. 2015, Widder et al. 2016).

1.5.2 Defining microbial adaptive strategies that encompass biological scales

A next step forward in microbial ecology research would be the search for major axes of adaptive variation within and among taxa, as has been performed in fields such as plant ecology (Grime 1977, Westoby 1998, Díaz et al. 2016) (Box 3). In other words, questions such as "Are there universal adaptive strategies across microbes?" or "What types of functional traits have driven the generation of their evolutionary diversity?" should be addressed. The investigation of overarching adaptive strategies among organisms also serves the pragmatic purpose of reducing dimensionality in the potentially large diversity of traits exhibited by an organism. While some researchers have attempted to apply some of the categorizations developed in macro-organismal ecology to the microbial realm (Fierer et al. 2007, Evans and Wallenstein 2014), their

use has so far been limited as these categories have not been easily applicable and measurable across taxa. Reducing thousands of potential traits to a smaller number of measurable functions that are consistently correlated with each other and with the ecological strategies of microbes will help researchers focus investigations and in turn address common issues such as time and budget limitations. For the foreseeable future, collecting microbial trait data will remain fairly expensive and time-consuming, making the use of a limited set of functional traits to study even more important.

1.5.3 Implementing better practices in database management

Much of the research in microbial ecology makes use of online databases of information on microbial phenotypes, functional genes, and proteins. An ongoing challenge to make the most of these data is the improvement of metadata collection and reporting to facilitate investigations and large-scale comparisons of microbial ecological dynamics and distributions (see also Nayfach and Pollard 2016, Tripathi et al. 2018). Such metadata include the methods used in preliminary analysis of the data and environmental conditions in which the data were collected, including relevant features of the host or habitat from which microbes were collected. While minimal information standards have been developed (Yilmaz and Al. 2011, McQuilton et al. 2016), their adoption and appropriation by practicing scientists remains difficult without proper large-scale consultations of all experts concerned. Built-in support for third-party annotations of records by expert users may also represent a way forward in improving the completeness of data in curated databases (Nilsson et al. 2019). Enforcement of these standards by database managers is nevertheless challenging and will require better funding for the operation and curation of these databases.

1.5.4 Improving knowledge transfer between ecologists and microbiologists

The limited use of microbial traits in testing ecological theories might be due to a divide in the current practice of microbial ecology by microbiologists versus non-

microbial ecologists, separated by a history of distinct scientific traditions and publishing journals (Koskella et al. 2017). The pairing between ecological questions needing to be answered and the biological knowledge of a system that could best serve these questions is made difficult for each side in the absence of active knowledge transfer. While there is no panacea in bridging that divide, it will be extremely important to work towards developing common definitions of concepts and providing their explicit definitions and explanations in publications and conferences that cross disciplinary boundaries (e.g. Tipton et al. 2019). Fostering pre- and post-publication peer review by researchers with different backgrounds will also help make the most of each discipline's experience.

1.6 Concluding remarks

Fostered by the development of high-throughput sequencing technologies, trait-based approaches to microbial ecology have accelerated the development of ecological knowledge of a highly diversified branch of the tree of life. These approaches have yet been successful in documenting the mechanisms by which microbes adapt to their environment, providing explanations to the variation in microbial life observed across several systems as well as insight into the generation of biological diversity. A strong bioinformatics capital in the discipline has further simplified online data sharing and thus increased opportunities for integration of results worldwide.

Knowledge development in microbial ecology could however be improved by a more consistent use of functional traits in generating predictions for testing ecological theories, and in better data and theory sharing between all practitioners of microbial ecology to facilitate integration and generalization of research results. We have argued here that the first step in reaching these goals should be to reach an agreement on what constitutes a valuable microbial trait to study (Klassen 2018). Building on definitions of functional traits developed in plant ecology, we suggest that adopting a concept of

functional traits anchored in adaptation and fitness would increase both the explanatory power of trait-based approaches and their relevance in testing ecological theories in microbial systems. The identification and use of major microbial adaptive strategies, combining numerous covarying traits, could further facilitate the methodological integration of trait-based results among research teams studying trait variation at different biological levels and with different methodologies and simplify the high-dimensionality of microbial trait data, which remain challenging to analyze and interpret. All in all, adopting a plan of action that seeks to firmly link microbial functions with fitness offers the promise to greatly accelerate knowledge development in microbial ecology.

1.7 Acknowledgements

We are grateful to Tonia de Bellis, Christophe Malaterre and Antoine C. Dussault for constructive conversations and input in the writing of this manuscript. This research was funded by a graduate fellowship from the Natural Sciences and Engineering Research Council of Canada and the Canada Research Chairs program.

1.8 Boxes

Box 1. Measuring microbial traits.

While the use of microbial functional traits in the framework of functional ecology – generally conceptualized as characteristics of microbes that might have an importance for their survival in an environment – is relatively recent, there is a long history in microbiology of measuring phenotypic traits of microorganisms. For example, while recent work in microbiology has moved to the use of sequencing-based approaches to identify microbial taxa, a compendium of phenotypic attributes or traits of bacterial taxa (Bergey 1923) was widely used for bacterial species identification and

diagnostic purposes for most of the 20th century. We here describe the most common approaches in use today, by classifying them into direct and indirect approaches.

Direct approaches refer to any trait measurement method that characterize traits of microbes through direct observation of phenotypes. They comprise traditional techniques of microscopy and cultivation for studying morphological characteristics of microbes (e.g. shape, cell wall structure) (Bergey 1923, Nielsen and Nielsen 2005). They also include phenotypic arrays, quantifying the physiological response of microbes (e.g. respiration) to a large range of substrates or stressors (Bochner 1989). Resource-use traits of microbes can then for example be described as the ability to metabolize different carbon compounds such as fructose, or to survive at different salt concentrations. Direct approaches may also involve the monitoring of metabolites (e.g. glucose, fumarate) produced by microbes of interest in culture or in the field providing a snapshot of their physiological state (Zhong et al. 2018). This approach, commonly performed through nuclear magnetic resonance or mass-spectrometry analysis, is referred to as metabolomics (Tang 2011). Lastly, metaproteomics refers to the analysis of proteins produced by a given sample of microbes, with each of the proteins with known roles for the organism being considered a trait (Maron et al. 2007). It is usually performed through mass spectrometry of isolated proteins.

Indirect approaches quantify microbial traits using the sequencing and analysis of genes via genomics, metagenomics including targeted sequencing of marker genes as well as shotgun sequencing of environmental DNA (National Research Council (US) Committee on Metagenomics: Challenges and Functional Applications. 2007), or sequencing of messenger RNA (via transcriptomics or metatranscriptomics) (Bashiardes et al. 2016). These approaches rely on the comparison of gene sequences to databases of described genes or proteins to infer their function and potential use to the microbes. The emergence of high-throughput sequencing have improved the quality of ecological inferences possible through such approaches by increasing the breadth

and depth at which diverse microbial communities can be described. Since interpreting the ecological function of single genes is not straightforward, microbial ecologists have commonly used gene hierarchy schemes to describe microbial traits, classifying genes by their contribution to higher-level traits such as metabolic pathways, or environmental sensing pathways (Tatusov et al. 1997, Ogata et al. 1999).

Box 2. Challenges in translating the concept of functional trait from macro-organismal to microbial ecology

The concept of functional trait used in macro-organismal ecology is essentially a selective one: ecologists aim to use traits that explain differences in fitness among individuals (Violle et al. 2007). For plants and animals, the notion of fitness and heritability of traits is relatively straightforward, as they are generally multi-cellular organisms that reproduce and transmit their traits vertically to their descendants. As a result, both the trait and the measure of fitness (e.g. the number of descendants) can be traced back to a single individual. Microbial ecologists trying to adapt this concept of the trait to the reality of microbial biology have been confronted with several challenges.

First, it has been difficult to apply an individual-based definition of trait to microbes provided their potential for horizontal gene transfer with other cells irrespective of phylogenetic identity (Abby et al. 2012). Other phenomena, such as the aggregation of microbes into biofilms that can be selected upon as a group (Ereshefsky and Pedroso 2013), or oppositely the presence of genetically diverse nuclei observed in single fungal mycelia (Ma et al. 2016a) make it even harder to target individual microbes as units of selection.

The difficulty in measuring trait and fitness at the level of individual cells is another facet of the problem. Commonly used environmental sequencing approaches do not allow the attribution of sequences to individuals, but are rather grouped by

sequence variants or taxonomic units (Callahan et al. 2016). The lowest level at which they are defined is thus generally the population, requiring the use of fitness measures such as the population growth rate, or population size. Technical solutions such as flow cytometry and single-cell genomics (Gawad et al. 2016) offer the possibility of characterization of functional traits of individual cells, but these approaches are still in their early stages of development. Whether these approaches will provide a standard and easily applicable way to perform trait-based ecology remains unknown. Whether a definition of functional traits focusing on individual cells or organisms can or should be adopted in microbial ecology is yet another open question (see main text and Inkpen et al. 2017).

Box 3. History of research on functional strategies in macro-organismal ecology

The classification of living organisms has a long history and has allowed comparisons of taxa to each other via measurable differences in their morphology, physiology or behaviors (Mayr 1982). One of the first major contemporary ecological classification systems was the r-K selection spectrum which aimed to explain life-history evolution (Dobzhansky 1950). This classification recognizes a trade-off between r-selected species, reproducing at fast rates with less investment in each offspring, and K-selected ones, reproducing at slower rates and investing more in the success of each reproductive unit. Popularized by MacArthur and colleagues in the 1960s and 1970s, it was most famously used in building models of population dynamics as part of the theory of island biogeography (MacArthur and Wilson 1967). Since then, it has more generally been used to conceptualize how density-dependent regulation and resource availability may be shaping the evolution of life-history (Reznick et al. 2002).

Building on the r-K spectrum, the CSR (competitor – stress-tolerant – ruderal) classification system was later developed by plant ecologists (Grime 1977) to explain species variation to environmental variation, based on two types of gradients: stress

and disturbance. Competitive species (C) grow best in low stress and low disturbance habitats, while the stress-tolerant (S) thrives under high stress and low disturbance, and the ruderal (R) is most adapted to low stress and high disturbance environments. Until recently (Li and Shipley 2017, Pierce et al. 2017), the difficulty of classifying animals and plants using these systems without prior study in cultivation or captivity made it difficult to apply them at large scales to facilitate integration and generalization of research results across biological scales and biomes (see main text).

The identification of main axes of measurable trait variation across species and the investigation of their global distribution was the next step in addressing this issue. Thus plant ecologists developed the leaf-height-seed (LHS) scheme (Westoby 1998, Westoby et al. 2002), representing three important axes of adaptive strategies that could easily be measured across many species and which explained more precisely the ecological trade-offs and variation among plant species at local and global scales. The LHS scheme proposed three phenotypic traits of plants that could be used as surrogate measures for plant ecological strategies: specific leaf area (the area of the leaf divided by its dry mass) as a proxy of resource conservation strategies, height as a proxy for energy acquisition and response to disturbance strategies, and seed weight as a proxy for dispersal and colonization strategies. Subsequent development of plant functional trait strategy schemes have included the incorporation of below-ground resource investment strategies (Li et al. 2017), and as world-wide observations of plant functional traits are collected, this framework has been expanded (Díaz et al. 2016, Laliberté 2017). Plant functional traits are increasingly reported in open-access databases (Kattge et al. 2011), and the availability of these data have helped explain the distribution of plant species among habitats and coexistence within habitats (Laughlin et al. 2010, Kunstler et al. 2016). They have also been used in building predictive models of ecosystem function (Cadotte 2017).

1.9 Glossary

Functional trait: Morphological, physiological, phenological or behavioural trait that impacts fitness by its effects on growth, reproduction or survival (Violle et al. 2007, Pey et al. 2014).

Explanation: Identification and description of the mechanisms underlying invariant causal relationships (Paslaru 2009).

Corroboratory prediction: Expectation that can be compared to scientific observations to test hypotheses, models and theories and support or not to the understanding of a phenomenon (Maris et al. 2018).

Integration: Formation of an account of a phenomenon that is built from a variety of ideas possibly coming from different levels of organization or disciplines (Brigandt 2013).

Data integration: Design and implementation of tools and standards for assembling and comparing data (Leonelli 2013).

Methodological integration: Creation and use of various methods for developing a more multi-faceted understanding of an ecological phenomenon or process than what could be obtained by using these methods individually (Leonelli 2013).

Explanatory integration: Use or combination in a new field of research, of hypotheses, models or theories developed in other disciplines (Leonelli 2013).

Generalization: Postulation of the occurrence of a pattern or process on a whole system from observation on a part. Generalization through abstraction can help

reduce the complexity of a system to facilitate its interpretation (Vepsäläinen and Spence 2000).

Fundamental niche: The range of environmental conditions individuals of a species may thrive under.

Realized niche: The portion of the range of conditions individuals of a species are actually found to inhabit, due to constraints on the occupancy of their fundamental niche.

CHAPTER 2

FUNDAMENTAL ECOLOGICAL STRATEGIES STRUCTURE GLOBAL BACTERIAL DIVERSITY

(Article under review at *Nature Ecology and Evolution*)

Geneviève Lajoie and Steven W. Kembel

Département des Sciences Biologiques, Université du Québec à Montréal
141 Avenue du Président-Kennedy, Montréal, Canada, H2X 1Y4

2.1 Abstract

The microbial realm is characterized by an exceptional diversity of forms and functions across Earth's ecosystems (Huttenhower et al. 2012, Sunagawa et al. 2015, Bahram et al. 2018). The discovery of major axes of correlated functional variation among species and habitats has revealed the fundamental trade-offs structuring both functional and taxonomic diversity in eukaryotes (Díaz et al. 2016, Pigot et al. 2020). However, we still do not know whether similar axes of functional variation shape global bacterial diversity, nor whether they could explain bacterial taxonomic turnover among ecosystems. Here we reveal the existence of universal bacterial ecological strategies using global genomic and metagenomic datasets. We identify three primary axes of correlated functional variation explaining both evolutionary differentiation within the Bacteria and their ecological sorting along major environmental gradients. Functional variation along these three axes is characterized by traits related to 1) DNA metabolism, 2) metabolism of secondary compounds, and 3) signalling and attachment to hosts, with most variation in these functional axes occurring among different bacterial families. Our results support clade-based sorting of bacteria across ecosystems as a major driver of global bacterial functional diversity. By reducing the high functional diversity of bacteria to fewer fundamental axes of variation, our study offers a way forward in generalizing our understanding of the drivers of biological diversity

2.2 Introduction

The identification of major axes of variation in life-history strategies among organisms and habitats has led to advances of both theory and practice in ecology and evolutionary biology (Grime 1977, Westoby et al. 2002, Lajoie and Kembel 2019). For example, in plants, large-scale screening of plant functional traits (Grime 1977, Westoby 1998, Díaz et al. 2016) led to the discovery that a few major axes of covarying traits related to leaf economics, seed mass, and plant height explain most of the global variation in plant functional and life-history strategies, and that each of these strategy axes could be characterized by the measurement of one or a few key functional traits (Westoby 1998, Wright et al. 2004). The results of this discovery have been twofold; first, it improved our understanding of the major environmental drivers and evolutionary origins of functional variation among organisms (Ackerly and Reich 1999, Wright et al. 2004). Second, the identification of a subset of measurable functional traits that can be used to characterize ecological strategies helped coordinate measurement efforts around a reduced set of interpretable functions, thus facilitating attempts at generalization among study systems worldwide (Kattge et al. 2011, Díaz et al. 2016, Madani et al. 2018).

While recent conceptual and technological developments have led to a proliferation of microbial trait measurements, functional categorization schemes proposed to date in microbial ecology have mostly consisted of direct translations of functional strategy schemes developed for macro-organisms, such as the r-K selection spectrum or Grime's CSR (competitive, stress-tolerant, ruderal) scheme (Fierer et al. 2007, Bissett et al. 2010, Wallenstein and Hall 2012, Evans and Wallenstein 2014, Santillan et al. 2019). While the application of these approaches has provided insights into the diversity of microbes in certain habitats (Malik et al. 2020), it has proven challenging to classify microbes into these categories, especially outside of those microorganisms that can be grown in culture for which phenotypic measurements are

available. The lack of a coherent functional strategy scheme for microbes and more generally a lack of data about the key sets of covarying microbial traits and strategies has prevented the widespread adoption of functional strategy schemes in microbes (Lajoie and Kembel 2019).

Here, we develop an *a posteriori* trait screening approach defining bacterial functional traits based on the relative abundance of gene families in genomic and metagenomic data sets (Louca et al. 2016) in order to identify ecological strategies structuring bacterial diversity across the tree of life and across ecosystems (Calow 1987). We go beyond existing large-scale assessments of metagenomic variation along environmental gradients (Dinsdale et al. 2008, Fierer et al. 2012, Thompson et al. 2017, Ramírez-Flandes et al. 2019) by identifying bacterial ecological strategies from the study of trait correlations across both genomes and environmental metagenomes. To evaluate the contribution of evolutionary processes to ecologically important trait variation across metagenomes, we tested whether the axes of trait covariation explaining the most variation among metagenomes from different environments would also be phylogenetically structured among genomes by comparing trait correlation structures between bacterial communities (metagenomes) and bacterial clades (genomes). We lastly assessed the taxonomic level at which traits explained most variation in both datasets in order to test hypotheses about the phylogenetic origins of ecologically important traits (David and Alm 2011, Martiny et al. 2013, Dolan et al. 2017). Such information is also paramount to improving predictions of ecosystem functioning from bacterial taxonomic composition data (Goberna and Verdú 2016).

2.3 Methods

2.3.1 Metagenomic dataset collection and processing

We searched the IMG/M system (Chen et al. 2019) for shotgun metagenomic datasets from environmental and host-associated sources. To limit biases due to sequencing technology (Clooney et al. 2016) and processing, only datasets that had been sequenced with Illumina Hi-Seq at the DOE Joint Genome Institute (JGI) were included. Using the Ecosystem search tool, we selected up to four samples per habitat from the main types of ecosystems listed under the category "Ecosystem Type" in this database, leading to a final set of 69 metagenomic sample records (Table S2.1). For each sample, we downloaded the KEGG functional annotations of sequences that had been performed by the JGI's Microbial Genome Annotation Pipeline on assembled and unassembled reads. We also downloaded each of their taxonomic annotations, consisting in the best BLAST hits of protein-coding genes (>30% identity) in each sample. In order to facilitate comparisons with the genomic dataset, metagenomic sequences were filtered to the bacterial kingdom only, which represented the vast majority of annotated sequences (~94% of all taxonomic annotations).

Counts of protein-coding (i.e. functional) genes retrieved for each sample were first rarefied to the smallest number of functional genes retrieved for any one sample (~74,000) to control for differences in sequencing depth among samples. We then calculated the average copy number of all functional genes per functional pathway and sample. These pathways were defined following the KEGG BRITE gene hierarchy (Kanehisa et al. 2014) which classifies genes by shared functionality at three different hierarchical levels, the lowest of which we considered here. We counted a pathway as present if at least one of its constitutive functional genes were present. We lastly generated a table of relative abundances of each of these functional pathways per sample. The total number of functional pathways was 312. Taxonomic annotations

were rarefied to 41,500 hits per sample. Tables of the relative abundances of each taxon in each sample were generated at the phylum, class, order and family levels.

2.3.2 Genomic dataset collection and processing

The genomic dataset, consisting of the functional annotations, taxonomic identification and phylogenetic relationships of more than 27,000 fully sequenced bacterial genomes, was retrieved from the AnnoTree server (Mendler et al. 2019). We simplified the dataset by randomly retaining a single genome at each tip of the phylogeny, leading to a final dataset of 15,973 genomes (Annex A: Fig. S2.1). Counts of protein-coding genes for each genome were transformed to relative abundances of KEGG pathways as with the metagenomic dataset. The total number of pathways was 312.

2.3.3 Identification of main functional axes in metagenomic dataset

We used a hierarchical clustering approach in order to identify groups of correlated functions that could constitute ecological strategies. We first generated a principal coordinate analysis of the metagenomic functional table based on the binary distance between samples. We then fitted the functional pathways onto the ordination to obtain vectors maximizing the correlation of these pathways with each ordination axis. These vectors thus explain the most variation among the metagenomic samples (Oksanen et al. 2013). To identify correlated axes of functional variation, we next calculated distances among the functional pathways using the coordinates of these vectors. We then performed a hierarchical clustering on these distances using Ward's minimum variance criterion to minimize within-cluster variance and generated a dendrogram of these relationships. We assessed support in the clustering through multiscale bootstrap resampling (n=10,000), as implemented in the R package pvclust (Suzuki and Shimodaira 2006). We lastly calculated each functional pathway's contribution to the variance in the principal coordinate analysis as the scores of each of

their vectors obtained from the fitting in order to assess the importance of the different groups in driving functional differentiation among metagenomes.

2.3.4 Phylogenetic structure of functional variation

We assessed the phylogenetic covariation of functional traits by performing a phylogenetic PCA (Jombart and Dray 2010) on the genomic functional dataset. This approach finds axes of functional variation that maximize the product of variance of the scores and their phylogenetic autocorrelation phylogenetic variation, so to reveal axes of correlated trait variation that are phylogenetically structured (Jombart et al. 2010). This analysis was performed on a genomic dataset rather than the metagenomic dataset, because it contained more complete annotations permitting a more accurate phylogenetic placement of sequences (Darling et al. 2014). Using the loadings obtained through this PCA, we performed a hierarchical clustering analysis on Euclidean distances among samples using Ward's minimum variance criterion.

We compared the structure of functional covariation in the metagenomic and genomic-based clusterings with a Procrustes analysis performed on their cophenetic distance matrices. The two dendrograms were pruned for this purpose to include the same functions. The significance of the Procrustes statistic, describing the similarity between the two datasets, was assessed by comparing the observed statistic to a distribution of 999 statistics generated through permutations of the original data (Oksanen et al. 2013). We generated a tanglegram of the two clusterings for visual comparison of the functional groupings resulting from the two datasets. The trees were untangled prior to plotting using R package dendextend (Galili 2015).

2.3.5 Phylogenetic depth of functions

To tell whether functional strategies important in driving differences among ecosystems tended to vary at a recent or ancient scale in Bacteria, we tested the mean

depth at which each of the functional pathways was conserved across clades within the genomic dataset using the consenTRAIT approach from Martiny et al. 2013, implemented in the R package castor (Louca and Doebeli 2018). Because this approach only accommodates binary data, we transformed our relative abundance data into presence-absence. As such, this test represents a conservative estimate of the depth at which traits relative abundances may actually vary. To be considered as possessing the trait, a clade had to have at least 70% of its constituting genomes possessing the trait. To evaluate whether the mean depth of each trait was different than expected from a random distribution of trait values, we compared the observed depths to those estimated based on a null model in which tips were randomly assigned the presence or absence of the traits (n=99). For comparison, we calculated mean phylogenetic depths at which taxonomic identities varied for each taxonomic level from phylum to genus using the same approach.

2.3.6 Contributions of environmental variables and bacterial taxnomy in driving metagenomic functional variation

We calculated the relative contributions of environmental variation among ecosystems and of the taxonomic composition of communities to metagenomic functional variation using a variation partitioning approach. We first performed PCAs on each of the taxonomic relative abundance table in order to extract major axes of taxonomic variation for each taxonomic level as well as to reduce dimensionality in the datasets. We then performed variation partitioning of the functional dataset using the first ten PCA axes of the phylum, order, family and genus data tables in order to assess what portion of the variation could be explained by each factor alone and by combinations of taxonomic factors.

2.4 Results and discussion

2.4.1 Defining main strategies driving bacterial functional turnover across ecosystems

To identify the principal strategies that drive variation in bacterial functional turnover across ecosystems, we searched for axes of correlated functional trait variation among microbial metagenomes from different ecosystems using a dataset of 69 metagenomes from diverse habitats (Annex B: Table S2.1). Functional annotations of sequences for all metagenomes were obtained from JGI's Microbial Genome Annotation Pipeline and filtered to bacterial sequences only. They were rarefied to ~74,000 sequences per sample and then aggregated by Tier 3 functional pathways as defined by the KEGG BRITE Hierarchy. These functional gene annotations were used as functional traits for further analyses. We used a principal coordinates analysis (PCoA) to identify the main axes of functional variation across metagenomes from different ecosystems. We next used multiscale bootstrap resampling to identify clusters of highly correlated trait variation that explained the most trait variation across samples.

We identify three strongly supported clusters of independently covarying bacterial traits (Fig. 2.1, detailed in Annex A: Fig. S2.2). These strategies – namely DNA metabolism, secondary compounds metabolism and signalling and attachment to host – explained variation across the transition from soil to aquatic microbiomes (Fig. 2.2 – Axis 1) and the transition from host-associated to environmental microbiomes (Fig. 2.2 – Axis 2). A fourth cluster of traits associated with oxidative stress response was also found to be strongly associated with the latter transition (Fig. 2.1, Fig. 2.2). Together, these first axes encompassed 34.1% of total trait variation. Each of these strategies were also observed to explain differences among microbiomes along the third and fourth axes of variation through sister clusters with similar trait composition (Fig. 2.1). Along these axes, most variation was aligned along a transition from strongly

symbiotic bacterial communities to epiphytic and free-living ones (Annex A: Fig. S2.3 – Axis 3) and from terrestrial to aquatic ecosystems (Annex A: Fig. S2.3 – Axis 4).

Communities associated with soils were characterized by groups of correlated functions linked with the degradation and biosynthesis of xenobiotics and toxins (Fig. 2.1 – C9 & 10). The fact that this strategy occurs in soil bacterial communities is not surprising and supports existing evidence of the prevalence of competitive dynamics among microbes in these diverse and densely populated habitats (Charlop-Powers et al. 2014, Tyc et al. 2017). Our results suggest that this axis of microbial trait variation is similar to the competitive axis of Grime's CSR strategy scheme (1977), though based on different premises; while Grime's CSR scheme defines a competitive plant species as one having faster growth rates and allocation to resource-acquisition structures, the competitive axis found in environmental metagenomes is related to the ability of bacteria to respond to and participate in chemical warfare.

Bacteria associated with aquatic microbiomes possessed suites of traits linked with DNA metabolism and repair (Fig. 2.1 – C2 & C1) and with oxidative stress response mechanisms (Fig. 2.1 – C8 & C7), both indicative of stress-tolerance strategies analogous to Grime's. Ecologically important stresses associated with aquatic communities include low resource availability, temperature fluctuation, and UV radiation (Matallana-Surget et al. 2012), that can be responsible for the production of reactive oxygen species (ROS) and the degradation of DNA. The use of the TCA cycle in controlling intracellular concentrations of ROS (Mailloux et al. 2007) and of antioxidants such as riboflavin (Ashoori and Saedisomeolia 2014) and alpha lipoic acid (Packer et al. 1995) to scavenge them are known to play a role in resistance to oxidative stress. The importance of DNA metabolism in aquatic habitats is also coherent with the presumed important role of dissolved DNA as a source of energy and nutrients for bacteria in such habitats (Lennon 2007).

The trait clusters that were most strongly associated with a host-associated ecological strategy were characterized by functional pathways involved in signalling and attachment to host cells, particularly membrane receptors (Högbom and Ihalin 2017) (Fig. 2.1 – C3 & C4). This category of traits does not appear to have any equivalent in previously defined functional strategies of micro- or macro-organisms. In the case of bacterial communities taking part in symbiotic relationships with hosts, several traits associated with the metabolism of terpenoids and polyketides were also part of the host-associated strategy. Terpenoids and polyketides being major plant secondary metabolites. Bacterial traits associated with the metabolism of major secondary metabolites such as terpenoids and polyketides likely represent adaptations specific to life in endophytic associations with plants or the plant-based diets of symbiotic host

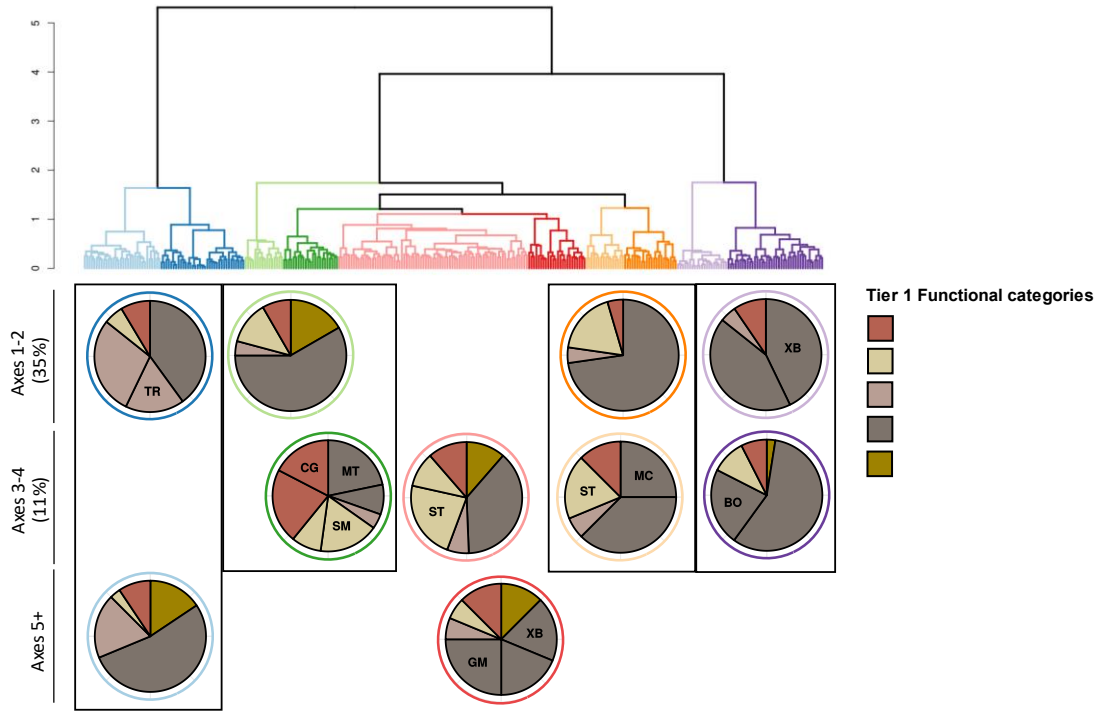

Figure 2.1 Dendrogram indicating correlations among bacterial functional pathways across 69 metagenomic samples. Main clusters of correlated traits are color-coded and labelled (C1-C10) in the dendrogram. The trait composition of each cluster is indicated in a piechart and color-coded using the Tier 1 functions of the KEGG functional hierarchy. Tier 2 functions that represented at least 15% of the total number of traits in each cluster are indicated with two letter codes. BO: Biosynthesis of other secondary metabolites; CG: Cell growth and death; GM: Glycan biosynthesis and metabolism; MC: Metabolism of cofactors and vitamins; MT: Metabolism of terpenoids and polyketides; SM: Signaling molecules and interaction; ST: Signal transduction; TR: Translation; XB: Xenobiotics biodegradation and metabolism. The axes to which they contribute most in the principal coordinates analysis of metagenomic functions (see Fig. 2.2) are indicated on the left. The major strategies explaining variation among metagenomes are indicated with black boxes.

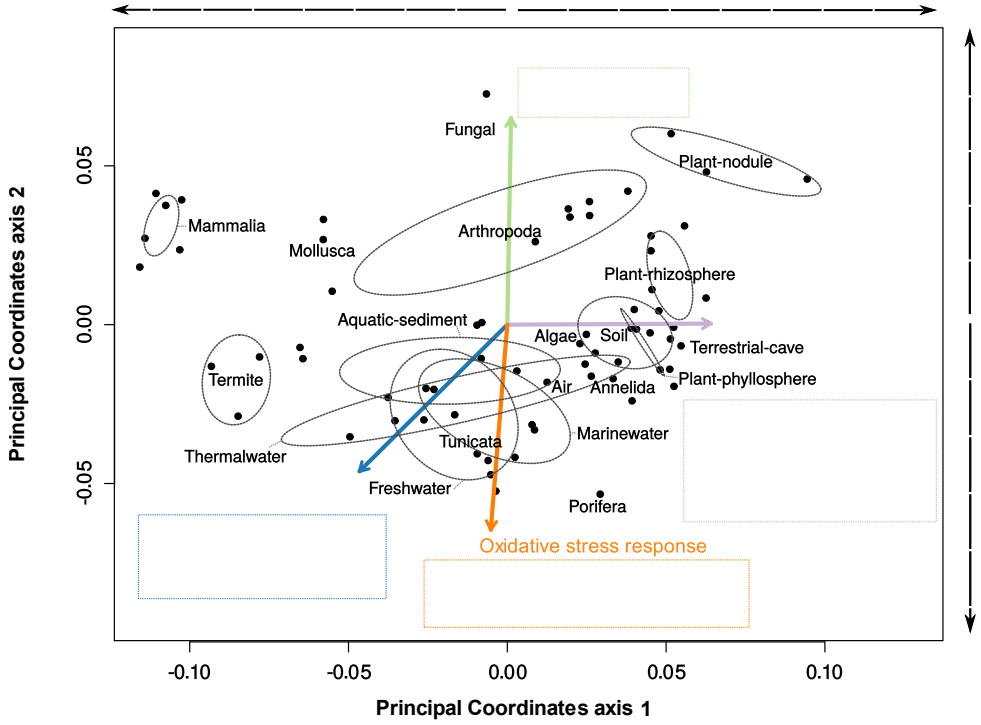

Figure 2.2 Principal coordinates analyses of the metagenomic functional dataset. Axes 1 and 2 are shown. Colored arrows represent the mean position of the traits contributing most to variance across these dimensions, by functional cluster (as depicted in Fig. 2.1). These most important traits are indicated in colored boxes next to the corresponding arrow. Bacterial phyla that correlate the most with the axes are indicated on the outer portion of the graph, along with the direction of the correlation. Ellipses define the average position of the points in each environmental group.

2.4.2 Evolutionary structure of correlated trait variation across metagenomes

Correlated axes of trait variation across ecosystems may be shaped by two major processes: the filtering of groups of traits differentially selected along environmental gradients, as well as evolutionary selection or constraints on the variation of these traits through time (Futuyma 2010, Muir 2015). To test the contribution of phylogenetic processes to the presence of correlated trait variation across metagenomes we compared the clustering of functional traits in metagenomes with a clustering analysis of an extensive genomic dataset in which we account for phylogenetic correlations among traits. Our main prediction was that the strategies of ecological importance to bacterial communities would be phylogenetically structured among bacterial genomes, such that functional pathways would cluster similarly in both datasets (Annex B: Table S2.2). To identify clusters of phylogenetically correlated traits across the bacterial tree of life, we first performed a phylogenetic principal component analysis ordination on functional annotations of 15,973 bacterial genomes (Mendler et al. 2019) (Annex A: Fig. S2.1) to reveal axes of trait variation taking into account the phylogenetic non-independence of evolving genomes (Jombart et al. 2010). We then identified clusters of correlated traits on these phylogenetically corrected ordination axes. Major clusters of bacterial functions are similar in both metagenomic and genomic datasets (Procrustes test: $m_{12} = 0.5051$, $r = 0.700$, $p = 0.001$), with the broad retention of major groups of covarying functions in the phylogenetic clustering analyses (Fig. 2.3, detailed in Annex A: Fig. S2.4).

Of the most ecologically important functional groups of traits presented above, clusters linked with DNA metabolism (Fig. 2.1 - C2 & C1) and secondary compounds metabolism (Fig. 2.1 - C9 & C10) were consistently clustered in both genomes and metagenomes, along with the host signalling and attachment cluster associated with symbiotic microbiomes (Fig. 2.1 - C4). These results suggest that selection or constraint (e.g. genetic correlations (Futuyma 2010) or functional barriers to

recombination (Hendrickson et al. 2018)) have been important in shaping the evolution of clusters of correlated traits in bacteria in a consistent fashion across both clades and habitats, leading to predictable variation in the ecological strategies of bacteria associated with particular clades (Fig. 2.2, Annex A: Fig. S2.3). The fact that we found such clusters despite the likely presence of many transient organisms in environmental metagenomes samples investigated provides additional confidence in these results.

Despite identifying three main axes of functional covariation among bacterial traits, other drivers of functional trait correlations appeared to play a role in structuring trait variation among ecosystems (Annex B: Table S2.2). First, metagenomically-derived trait clusters such as those associated with oxidative stress response in aquatic systems (Fig. 2.1 - C7 & C8), or those generally correlated with host-associated life (Fig. 2.1 - C3) were not observed in the genomic dataset (Fig. 2.3). These clusters might thus result from environmental filtering on pre-existing trait variation or lateral gene transfer among host-associated clades rather than a fundamental constraint or trade-off in genome architecture and evolution. Phylogenetic biases in the genomic dataset could also prevent us from detecting some functional strategies in genomic analyses. Second, most of the remaining traits were very loosely clustered in both metagenomic and genomic clusters (Fig. 2.3). The fact that these traits explained little of the functional variation among communities in the metagenomic dataset suggests they do not represent part of a functional strategy that has arisen in response to the major axes of environmental variation encompassed by this dataset. Their lack of correlations in the genomic dataset suggests that they are likely to have been evolving idiosyncratically or to be subject to more extensive horizontal gene transfer. It is possible that other types of functional strategies than those identified above are also important for driving within-community niche partitioning among bacteria, but at the broad scales of the bacterial tree of life and among metagenomes from distinct habitat types these types of trade-offs are not important enough to be consistently identified as clusters of covarying traits.

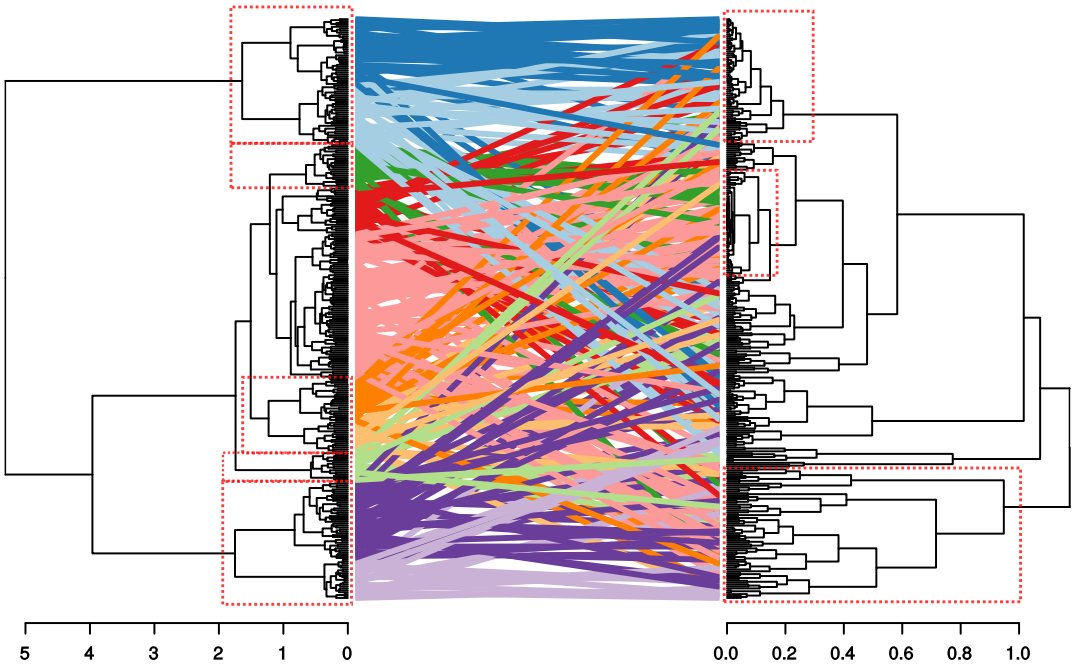

Figure 2.3 Comparison of functional clusters for bacteria based on annotation of metagenomic (left) and genomic (right) datasets. Dendrograms indicate correlations among bacterial functional pathways; functional pathways are color coded by metagenomic clusters (see details in Fig. 2.1). Lines connect the same functional pathways in each data set. Red boxes indicate the main functional strategies identified for bacterial genomes (see details in Fig. 2.1, Fig. 2.2).

2.4.3 Phylogenetic depth of ecologically important trait variation

One of the interests of understanding the phylogenetic structure of trait variation in microbes has been to improve predictions of functional traits and ecosystem functioning of microbial communities from taxonomic community composition data (Goberna and Verdú 2016, Dolan et al. 2017). Studies assessing the phylogenetic depth of trait conservatism in Bacteria have suggested a relationship between the complexity of a trait and the phylogenetic depth at which it arose, with complex traits like methanogenesis that involve multiple functional pathways being conserved more deeply in the phylogeny than relatively simple traits such as organic substrate utilisation (Martiny et al. 2013). Still, the phylogenetic depth at which the traits that are most important for driving ecological gradients in bacterial community composition are conserved remains untested. Understanding these patterns is important for predictive purposes as well as for linking the evolution of traits in Bacteria to their biogeographical distributions.

To determine the depth at which functional pathways explained most variation in the bacterial phylogeny, we used the consenTRAIT approach which evaluates the phylogenetic similarity of groups of organisms sharing discrete traits (Martiny et al. 2013), here the presence or absence of functional pathways. We compared these depths to those obtained for the different taxonomic levels encompassed by our dataset to test at which level taxonomic variation best explained the functional composition of metagenomes. For this analysis, we used variation partitioning of metagenomic functions using the taxonomic composition of communities (phylum, class, order and family) as explanatory variables.

The mean depth at which functional pathways varied among bacterial clades did not vary considerably among major functional clusters (Annex A: Fig. S2.5). Most trait variation was observed at the level of taxonomic families or slightly lower. The importance of ecological variation among bacterial families in driving variation along

the main functional axes of variation in the metagenomic dataset was also evident based on the variation partitioning analysis in which families could explain the majority of functional variation among metagenomes (Annex A: Fig. S2.6). These results support a role for complex adaptations in mediating adaptation across large environmental gradients, as they reflect the phylogenetic depths at which other complex microbial traits such as methylotrophy or survival in saline environments vary among microbial clades (Martiny et al. 2015, Goberna and Verdú 2016). They also support an important role for clade-based bacterial sorting across environments in maintaining the match between organismal traits and their environment.

We have identified large-scale functional strategies across the bacterial domain, and a stimulating research venue will be to test the scale-dependence of these trait relationships across different biological scales (Anderegg et al. 2018). We might expect that the numerous traits that explain only a small proportion of global variation among metagenomes could play important roles in niche differentiation within certain habitats, such that different functional axes could be important in driving adaptive variation and the sorting of Bacteria at smaller environmental and phylogenetic scales (Martiny et al. 2009). Overall, the integration of evidence from different biological and phylogenetic scale represents a fruitful venue for expanding our understanding of the generation and maintenance of bacterial diversity.

2.5 Conclusion

In conclusion, we used a data-driven functional trait screening approach to identify the major axes of functional trait covariation in bacterial metagenomes and genomes based on gene functional annotations. By comparing functional trait clusters in both datasets, we show that the major strategies driving functional differentiation between bacterial genomes can be scaled up to the level of metagenomes of co-occurring bacteria in ecological communities. There is limited overlap between the

bacterial ecological strategy axes described here and the trait classification schemes developed for plants and animals. While we find evidence for bacterial traits related to metabolism and resource utilisation driving the functional differentiation among bacterial communities and clades, we have shown that many major clusters of functional traits that we identify differentiate bacterial strategies based on biotic interactions. Taken together, our results offer a first step in quantifying global microbial functional trait strategies based on both environmental and genomic data. By reducing the high-dimensionality of trait variation observed among microorganisms around a small number of fundamental axes of trait covariation, we make an important step towards generalization of microbial ecology and of the drivers of biological diversity across study systems (Lajoie and Kembel 2019).

2.6 Acknowledgements

We are grateful to Jesse Shapiro and Mario Muscarella for comments on a version of this manuscript. GL was funded by a PhD fellowship from the Natural Sciences and Engineering Council of Canada and SWK by the Natural Sciences and Engineering Council of Canada and the Canada Research Chairs program.

2.7 Data and code availability statement

Accession of metagenomic datasets can be achieved through the IMG/M server (https://img.jgi.doe.gov/cgi-bin/m/main.cgi) by entering the IMG/M taxon object IDs provided in Annex B (Table S2.1) in the Quick Genome Search tool. Genomic functional and phylogenetic datasets can be accessed through the Downloads tool of the AnnotTree server (http://annotree.uwaterloo.ca/app/downloads.html). The full database can be downloaded at https://bitbucket.org/doxeylab/annotree-database/src/master/. The code used for processing data and performing the analyses described in this manuscript will be deposited on GitHub upon acceptance.

CHAPTER 3

ADAPTIVE MATCHING BETWEEN PHYLLOSPHERE BACTERIA AND THEIR TREE HOSTS IN A NEOTROPICAL FOREST

(Article published in 2020 in *Microbiome*, 8: 70)

Geneviève Lajoie, Rémi Maglione and Steven W. Kembel

Département des Sciences Biologiques, Université du Québec à Montréal
141 Avenue du Président-Kennedy, Montréal, Canada, H2X 1Y4

3.1 Abstract

The phyllosphere is an important microbial habitat but our understanding of how plant hosts drive the composition of their associated leaf microbial communities and whether taxonomic associations between plants and phyllosphere microbes represent adaptive matching remains limited. In this study we quantify bacterial functional diversity in the phyllosphere of 17 tree species in a diverse neotropical forest using metagenomic shotgun sequencing. We ask how hosts drive the functional composition of phyllosphere communities and their turnover across tree species, using host functional traits and phylogeny.

Neotropical tree phyllosphere communities are dominated by functions related to the metabolism of carbohydrates, amino acids and energy acquisition, along with environmental signalling pathways involved in membrane transport. While most functional variation was observed within communities, there is non-random assembly of microbial functions across host species possessing different leaf traits. Metabolic functions related to biosynthesis and degradation of secondary compounds, along with signal transduction and cell-cell adhesion were particularly important in driving the match between microbial functions and host traits. These microbial functions were also evolutionarily conserved across the host phylogeny.

Functional profiling based on metagenomic shotgun sequencing offers evidence for the presence of a core functional microbiota across phyllosphere communities of neotropical trees. While functional turnover across phyllosphere communities is relatively small, the association between microbial functions and leaf trait gradients among host species supports a significant role for plant hosts as selective filters on phyllosphere community assembly. This interpretation is supported by the presence of phylogenetic signal for the microbial traits driving inter-community

variation across the host phylogeny. Taken together, our results suggest that there is adaptive matching between phyllosphere microbes and their plant hosts.

3.2 Background

The phyllosphere – the aerial surfaces of plants including leaves – is a widespread microbial habitat that hosts a diversity of microorganisms that play key roles in plant ecology and evolution (Vorholt 2012). Phyllosphere microbes play key roles in plant health (Saleem et al. 2017, Wagi and Ahmed 2017) and human health (Berg et al. 2014), and can influence ecosystem function (Laforest-Lapointe et al. 2017). At a broad taxonomic scale, phyllosphere bacterial communities are consistently dominated by taxa including Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria (Bulgarelli et al. 2013), indicating that plants also influence the composition of their microbial partners. A key goal of phyllosphere microbial ecology research has been to identify the adaptive basis of such relationships between plants and associated microbes.

Comparative studies of the taxonomic composition of phyllosphere microbial communities across plant hosts have demonstrated the importance of host identity as a key driver of variation in phyllosphere microbial taxonomic diversity. At fine taxonomic scales, the composition of these communities varies predictably across host plant species (Kim et al. 2012, Lambais et al. 2014, Laforest-Lapointe et al. 2016) and across genotypes within host plant species (Bailey et al. 2005, Schweitzer et al. 2008). Plants and associated bacteria also show cophylogenetic associations, with clades of plants and bacteria consistently occurring together (Redford et al. 2010, Kim et al. 2012, Kembel et al. 2014), suggesting close adaptive associations between plants and their phyllosphere microbes.

Determining whether plant-microbe associations in the phyllosphere have an adaptive basis will require establishing how both plant and microbial functions are related across a range of host species. Plant functional traits – measures of morphology and physiology that capture key axes of variation in plant life history and ecology

(Violle et al. 2007) – have been targeted as a potential proxy for explaining microbial community turnover among plant species. These traits determine the potential for nutrient, metabolite and secondary compound leaching from the plant, which should largely determine the quality of a leaf as a habitat for phyllosphere microbes (van der Wal and Leveau 2011). In support of this hypothesis, plant functional traits such as leaf mass per area, leaf elemental composition, and growth rate are correlated with phyllosphere microbial community turnover both within (Hunter et al. 2010) and among plant species (Whipps et al. 2008, Yadav et al. 2008, Barott et al. 2011, Bodenhausen et al. 2014, Kembel et al. 2014).

Several studies have reciprocally identified the broad-scale microbial functional categories and adaptations that epiphytic microbes possess for living on plants (e.g. Krohn-Molt et al. 2013, Akinsanya et al. 2015, Finkel et al. 2016, Sambles et al. 2017). Functions including the biosynthesis of osmoprotectants such as trehalose and betaine and the production of extracellular polysaccharides are enriched in the phyllosphere and are thought to provide key adaptations to life on leaf surfaces by allowing microbes to attach to the leaf surface and by providing resistance to environmental stresses and plant defenses (Hartmann et al. 2009, Rastogi et al. 2013). The enrichment of rhodopsin genes in leaf bacterial communities exposed to high light also points to a role for those pigments in improving microbial fitness through higher energy acquisition on sun leaves (Atamna-Ismaeel et al. 2012). However, studies of microbial functions in the phyllosphere have largely been based on comparison of one or a few host plant species. How microbial functions map onto variation in host plant functions in diverse natural communities thus remains largely unknown. As a result, it is not clear whether plant microbiota exhibit the pattern of taxonomic turnover but functional homogeneity across hosts that has been observed in some animal microbiota (Human and Project 2012) or if a turnover in microbial functions can also be observed across functionally different tree species.

In this study, we quantified the functional repertoire of microbial communities on leaves of multiple tree species in a neotropical forest on Barro Colorado Island (Panama) using metagenomic shotgun sequencing. Sampling was performed in a 50-ha long-term plot of old-growth tropical forest within which ~300 tree species have been recorded, most of them evergreen (Condit et al. 1999). We asked which microbial functions are abundant in the phyllosphere, and how these functions are linked to the taxonomy and functional traits of plant hosts. Our central hypothesis was that the plant-microbe taxonomic associations previously observed in this forest (Kembel and Mueller 2014, Kembel et al. 2014) should be driven by adaptive matches between microbes and host plants, leading to several key predictions. First, we predicted that microbial functions should vary among host plant species and be correlated with the functional traits of the hosts. Second, we predicted that cophylogenetic associations between trees and microbes should lead to phylogenetic signal in microbial functions present on different plant hosts. Third, we predicted that microbial functions present on leaves should be filtered by the host, since conditions on the leaves of different host plants create a selection pressure on the functions of microbes able to persist on those leaves.

3.3 Methods

3.3.1 Microbial DNA collection, extraction and sequencing

Microbial communities were collected from the leaves of 24 individual trees from 17 tree species (1-2 samples per species) in the tropical lowland rainforest of Barro Colorado Island, Panama, in December 2010. These samples were selected from a larger pool of samples (Kembel and Mueller 2014, Kembel et al. 2014) for which we had sufficient quantities of high-quality DNA, selecting host species to maximize the phylogenetic and functional diversity of hosts. Methodological details of sample collection are described by Kembel et al. (2014). Briefly, 50-100g of fresh leaves were

collected from the subcanopy of one tree of each species. Microbial cells were then washed from each leaf sample using phosphate buffer [1 M Tris•HCl, 0.5 M Na EDTA, and 1.2% CTAB] and collected by centrifuging at 4,000 × g for 20 min. DNA was extracted using MoBio PowerSoil DNA extraction kits and samples stored at -80°C for future analyses. We quantified DNA concentrations and sequenced both extraction negative controls and PCR negative controls for these samples as part of previously published analyses of bacterial 16S and fungal 28S amplicon sequencing of these samples (Kembel and Mueller 2014, Kembel et al. 2014); none of the negative control samples contained measurable concentrations of DNA and upon sequencing they contained fewer DNA sequences than the minimum cut-off for inclusion in analyses As a result, they were all excluded from subsequent analyses in previously published studies and the present study. To quantify the metagenomic structure of each microbial community, we constructed a paired-end metagenomic shotgun library including a random sample of the whole community DNA composition using an Illumina Nextera XT® kit (Illumina reference FC-131-1024). These libraries were then sequenced using Illumina MiSeq paired-end 2 x 250 base pair sequencing (V2 kit, Illumina reference MS-102-2003). Analyses were performed on these 24 samples unless stated otherwise. Results were not influenced by including replicates of the same species (see tests below).

3.3.2 Microbial taxonomy and functional trait annotation

Metagenomic shotgun sequencing yielded 14,642,408 reads in total. We trimmed sequences to remove Illumina adapters and truncate end-bases with a quality score less than 20, and removed sequences shorter than 25bp, leaving 14,634,072 trimmed and quality-controlled reads. Taxonomic annotation of all sequences in each microbial community was performed to restrict functional analyses only to bacterial sequences. We annotated metagenomic reads using Kaiju, which annotates taxonomic identity of reads by comparing sequenced reads to the microbial subset of the NCBI

BLAST non-redundant protein database (Menzel et al. 2016). Out of the 7,317,036 sequences, we were able to annotate taxonomy to at least the taxonomic level of domain/kingdom for 2,138,885 sequences, of which 2,100,491 sequences were from Bacteria, representing 29% of total sequences. All subsequent taxonomic and functional analyses were based on this subset of sequences identified as belonging to the Bacteria. Of these Bacterial sequences, 1,902,749 were annotated to at least the phylum level, representing 26% of total sequences. Analysis of taxonomic composition was carried out on this subset of sequences annotated to at least the bacterial Phylum level. We rarefied all samples to 20,100 randomly chosen sequences per sample for taxonomic composition analyses, resulting in a total of 482,400 sequences for taxonomic analyses (relative abundances of major taxa are shown in Annex C: Fig. S3.1).

Functional annotation of microbial sequences was performed via protein homology searches using the KEGG annotation framework (Ogata et al. 1999, Kanehisa et al. 2014) via the software COGNIZER (Bose et al. 2015). Analyses resulted in the identification of functional genes and categories for 873,082 sequences representing 12% of sequences. In total, of the 7,317,036 bacterial sequences that were obtained from the metagenomic sequencing of all samples, 722,936 sequences were taxonomically annotated as bacteria and had a functional annotation. Only these sequences that were both bacterial and functionally annotated were used for the functional analyses. We lastly classified each of these sequences into functional categories, defined by the BRITE functional hierarchy manually curated for the KEGG annotation system based on published literature (Kanehisa et al. 2014). This hierarchy contains four different levels, which were designed as Tier 1, Tier 2, Tier 3 and functional genes, ranging from the more general to the more specific functional assignment (see Staley et al. 2014). Most analyses were performed at the Tier 3 level, in the intent of reaching a balance between the complexity of the data and its interpretability. In a few instances, Tier 3 categories were perfectly correlated across

samples so we removed the duplicates from the dataset in order to reduce its dimensionality (Annex D: Table S3.1).

3.3.3 Plant functional traits and phylogeny

We obtained measurements of plant functional traits for all plant species from a dataset collected previously on Barro Colorado Island (Wright et al. 2010). This trait database initially included 21 whole-plant and leaf traits, but we reduced these traits to a subset of 16 traits with limited overlap in functional significance (Pérez-Harguindeguy et al. 2013) (Annex C: Fig. S3.2). This reduced set of traits included height at maturity, sapling growth rate and sapling mortality rate as whole-plant resource-use traits, leaf area and leaf dry matter content as leaf structural traits, and a suite of leaf elemental chemistry traits including concentration of aluminum, calcium, copper, magnesium, phosphorus, zinc and nitrogen content. A phylogenetic hypothesis for host plant species was obtained by grafting tree species onto a dated megatree of angiosperms provided by Zanne et al. (2014) using Phylomatic v.3 (Webb and Donoghue 2005).

3.3.4 Variation in phyllosphere functions among versus within samples

We determined the contributions of within- and among-sample variation in function of total functional variation among metagenomic samples using additive diversity partitioning, where $\gamma_{div} = \alpha_{div} + \beta_{div}$ (Veech et al. 2002). The percentage of alpha diversity was calculated as the amount of alpha entropy divided by the amount of total entropy across all communities. The percentage of beta diversity was calculated as 1 minus the percentage of alpha diversity. These metrics were calculated using the R package entropart (Marcon and Hérault 2015). Analyses were performed at three levels of functional aggregation (Tier 1 to Tier 3). We tested whether the presence of two samples rather than one for some of the sampled species would affect this diversity partitioning by subsampling the dataset to include all possible combinations of samples

totally a single sample per species (n=128) and rerunning the analyses. This subsampling did not affect our results (Annex C: Fig. S3.3), such that we kept the 24 samples in the subsequent analyses. We then compared sources of turnover for functions and taxonomy between samples by performing the same analysis from the taxonomically annotated metagenomic sequences, defined at levels from phylum to species.

3.3.5 Associations between microbial and plant traits

We performed a principal component analysis (PCA) of functional trait matrices and identified the functions contributing most to variation along the first axes of variation using R package FactoMineR (Lê et al. 2008). We fitted the plant traits onto this ordination to identify correlations between bacterial traits driving the PCA and the plant traits. We evaluated the influence of tree species replicates in our samples on these results and did not uncover important differences in the main drivers of functional differences among samples when excluding these duplicates. We also performed a Procrustes analysis (Oksanen et al. 2013) on the duplicated samples to evaluate whether their functional composition were correlated and obtained a very high correlation coefficient (r=0.989). As such, we can assume that individuals from the same species are structuring their leaf microbial communities in a similar way and do not drive important functional differences among samples. All 24 samples were thus kept in this analysis.

We quantified the phylogenetic signal in associations between microbial functions and host plant phylogeny using function *multiphylosignal* from R package Picante (Kembel et al. 2010) to calculate Blomberg's K. This statistic quantifies whether a microbial trait exhibits stronger phylogenetic signal than expected by chance under a Brownian motion model of trait evolution. The higher the K statistic, the more phylogenetic signal in the trait. We identified microbial functions with strong phylogenetic signal by comparing the variance of independence contrasts observed for

each microbial function to those obtained through a null model where taxa labels have been shuffled across the tips of the phylogeny (n=9,999 randomizations). We considered a microbial function to exhibit strong phylogenetic signal if it fell in the top 5% of the distribution of signal based on the randomization test ($P < 0.05$ according to randomization test). We selected a single random sample per host species for those host species with more than one sample prior to calculating phylogenetic signal. We repeated this for different random subsamples and it did not qualitatively change the results so we report phylogenetic signal for a representative random subsample.

3.3.6 Host filtering of microbial functions and taxa

The degree of host filtering on microbial communities was assessed by comparing the occurrence of traits in observed communities to those obtained from 9,999 randomizations of community trait matrices. Host filtering was detected as an over- or under-representation of the given trait in individual communities. Randomizations were generated by permutations of the trait matrix preserving row and column totals. For each site and bacterial trait combination, we compare the observed frequency of the trait to the random values to assess whether it was lower or higher than expected by chance. To compare the strength of functional vs. taxonomic filtering, we applied the same procedure to the taxonomic datasets defined at each of six taxonomic levels, from the phylum to the species.

3.4 Results

3.4.1 Metagenomics shotgun sequencing characterization of phyllosphere microbial functions

Overall, we detected 4587 different functional genes across all samples based on annotation of metagenomic shotgun sequencing of tropical tree phyllosphere communities. Functions related to metabolism were the most abundant overall in our

dataset, making up 45% of all functionally annotated sequences (Fig. 3.1). The principal metabolic functions in the phyllosphere were related to metabolism of amino acids (e.g. amino acid related enzymes), nucleotides (e.g. purine and pyrimidine metabolism), carbohydrates (e.g. pyruvate, glyoxylate and dicarboxylate metabolism), and energy (e.g. oxidative phosphorylation & TCA cycle) (Fig. 3.1). Groups of functional genes related to environmental and genetic information processing also had a high relative abundance, mainly membrane transport (e.g. transporters), translation (e.g. aminoacyl-tRNA biosynthesis), and signal transduction (e.g. two-component systems).

3.4.2 Variation in phyllosphere functions and taxa among versus within samples

The bacterial functions present on tree leaves were remarkably consistent among different samples. The vast majority of functional variation occurred within samples (>97%), with a very small contribution of functional turnover among samples (<3%) to total functional diversity, regardless of the functional level under study. Most taxonomic diversity was also observed within samples, with a contribution of beta-diversity increasing from 1 to 4.4% of total diversity with a refinement of the taxonomic scale utilized (Table 3.1). The principal component analysis of bacterial community functional composition indicated that metabolic functions related to biosynthesis and degradation of secondary compounds and antibiotics, as well as functions related to signal transduction and cell-cell adhesion were the most strongly varying among hosts (Fig. 3.2; Annex D: Table S3.2). We detected 16 Tier 3 functions that exhibited strong phylogenetic signal with respect to the host phylogeny (functions with phylogenetic signal in top 5% of values compared to null distribution compared to K statistic randomization test; $P<0.05$; Fig. 3.3). These functions were mostly involved in the metabolism of terpenoids and polyketides, signal transduction and cellular processes.

3.4.3 Associations between microbial and plant traits and host filtering

Many of the plant traits displayed some level of correlation with the principal axes of microbial functional community composition. Among these, morphological leaf traits (e.g. leaf area, leaf mass per area) were most strongly associated with the first two axes of microbial functional variation. Leaf elemental concentrations of copper, aluminum and manganese were also strongly correlated with these first dimensions. The plant trait gradients explained altogether ~17% of variation in functional composition among microbial communities. The vast majority of the microbial Tier 3 functions were more abundant or less abundant than 95% of the values obtained from null model keeping both the total abundance of a trait and the number of traits in a community constant (Table 3.2). The filtering signal was slightly stronger for the microbial taxa than for the microbial functions (Table 3.2).

Table 3.1 Functional and taxonomic additive diversity partitioning of bacterial communities across 24 tree phyllosphere samples. The percentage of alpha diversity was calculated as the amount of alpha entropy divided by the amount of total entropy across all communities. The percentage of beta diversity was calculated as 1 minus the percentage of alpha diversity.

	Functional			Taxonomic					
	Tier 2	Tier 3	Functional gene	Phylum	Class	Order	Family	Genus	Species
Alpha diversity (%)	100.0	99.8	97.2	99.0	99.0	99.0	98.8	98.2	95.6
Beta diversity (%)	0.0	0.2	2.8	1.0	1.0	1.0	1.2	1.8	4.4

Table 3.2 Occurrences of Tier 3 functions and taxa across 24 tree phyllosphere samples. Occurrences of Tier 3 bacterial functions and taxa that are respectively more or less abundant than 95% of the values obtained from a null model randomizing abundances of functions and taxa across hosts (n=9,999).

	Number of combinations in the top 5% of the null model values	% of total	Number of combinations in the bottom 5% of the null model values	% of total	Total number of combinations
Functions					
Tier 3 functions	4360	70	930	15	6192
Taxa					
Phylum	1397	69	279	14	2016
Class	1073	63	405	24	1704
Order	2426	62	988	25	3888
Family	5597	64	2014	23	8808
Genus	26723	66	6690	16	40608
Species	183337	76	23479	10	240288

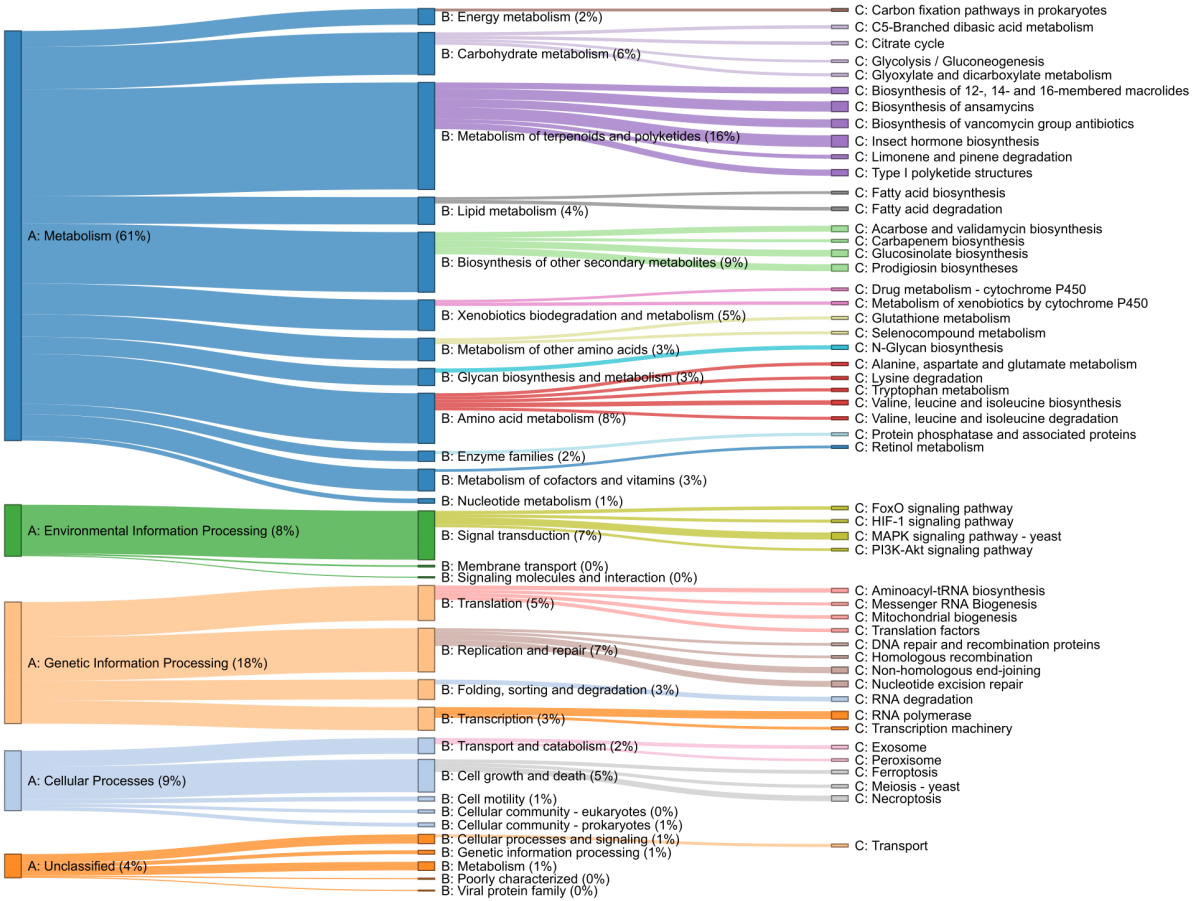

Figure 3.1 Relative abundance of the most abundant functional pathways detected across 24 tree phyllosphere samples in a neotropical forest in Panama. Functional pathways are classified using the KEGG functional hierarchy (Kanehisa et al. 2014).

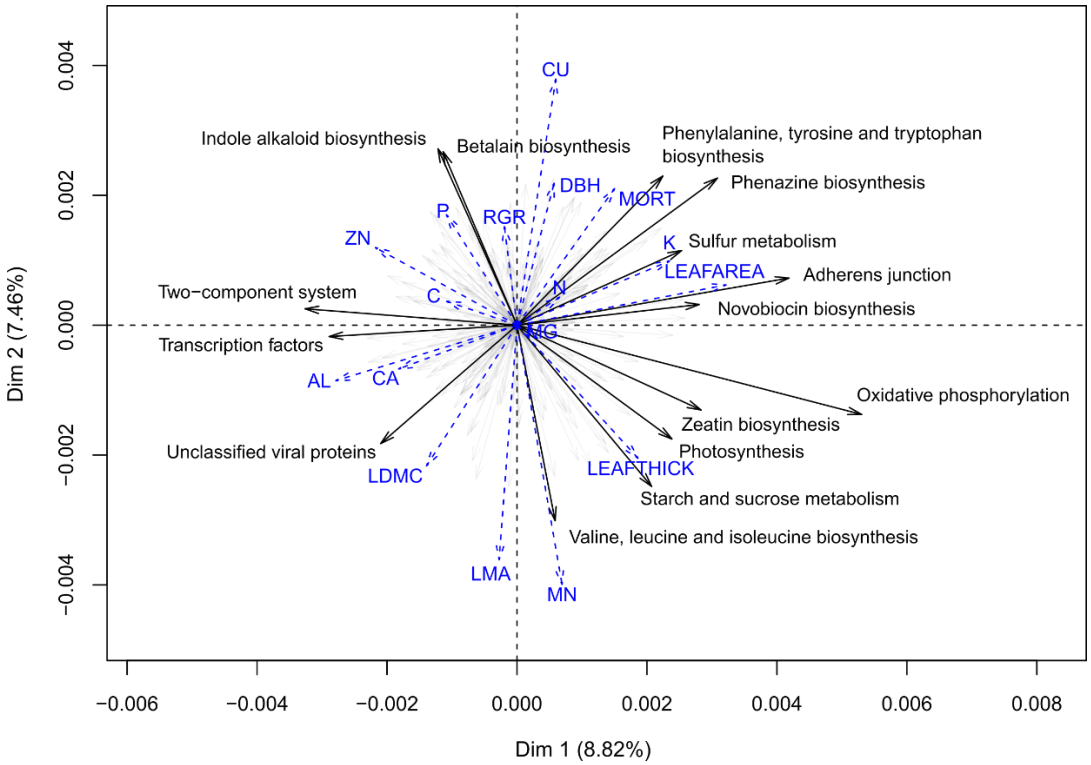

Figure 3.2 Principal components analysis (PCA) of microbial functional composition from the phyllosphere of neotropical trees. The 20 Tier 3 functions contributing the most to variation among samples are indicated as black arrow. Plant traits were fitted onto the PCA in a configuration that would maximize correlation with the PCA axes and are represented as blue dashed lines. Plant trait abbreviations are the following: Aluminum (AL), Calcium (CA), Carbon (C), Copper (CU), Diameter at breast height (DBH), Leaf area (LEAFAREA), Leaf dry matter content (LDMC), Leaf mass per area (LMA), Leaf thickness (LEAFTHICK), Manganese (MN), Mortality (MORT), Nitrogen (N), Phosphorus (P), Potassium (K), Relative growth rate (RGR), Zinc (ZN).

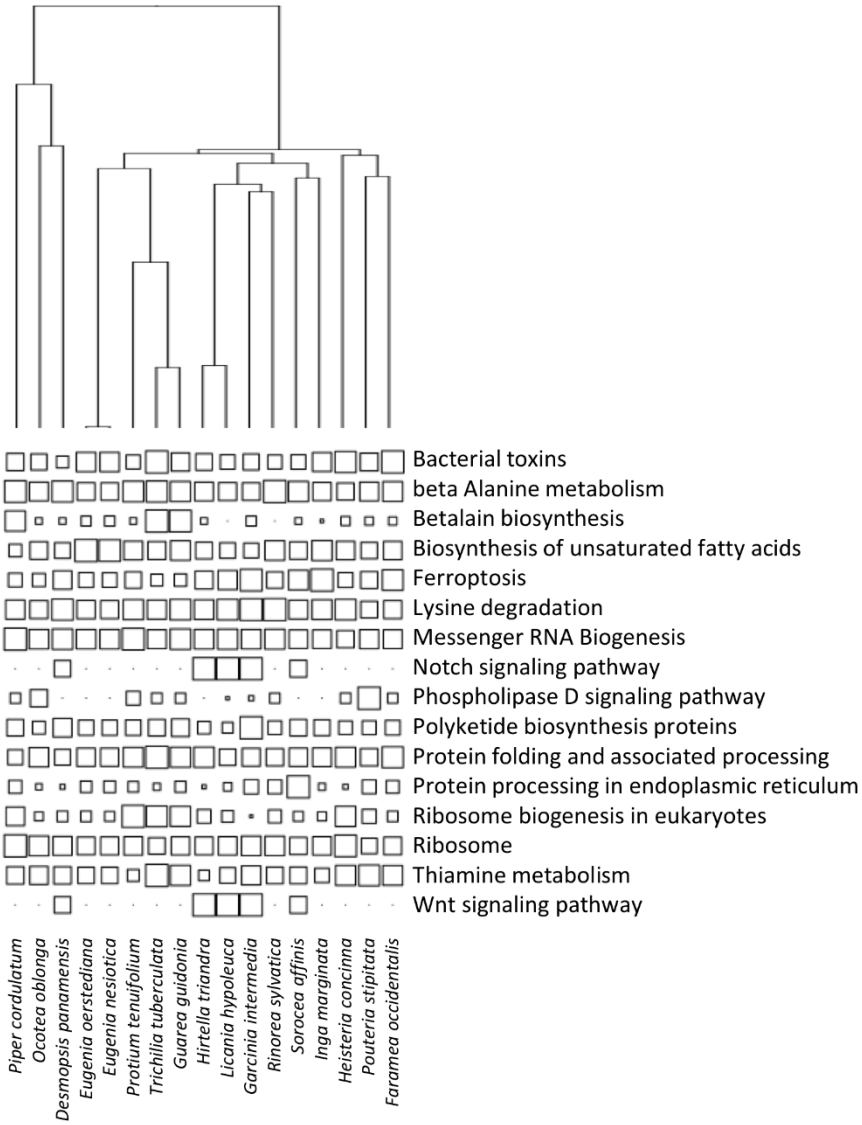

Figure 3.3 Distribution of microbial functions with respect to plant phylogeny. Distributions are shown for the subset of Tier 3 microbial functions with phylogenetic signal (K statistic) in the top 5% of values compared to the expected distribution of phylogenetic signal according to the K statistic randomization test ($P < 0.05$). Symbol size indicates the scaled relative abundance of microbial functions for each host species.

3.5 Discussion

The functional composition of tree phyllosphere microbial communities in a tropical forest in Panama is largely consistent with those reported in the literature, regardless of the type of plant studied, suggesting the presence of a core functional microbiota in phyllosphere microbial systems. Core functional microbiota in host-associated systems have also been reported for other hosts. Our study supports findings of an important role for the metabolism of carbohydrates and amino acids in bacterial survival in the phyllosphere (Yadav et al. 2008, Ryffel et al. 2015, Müller et al. 2016) that is consistent with the abundance of these compounds in leaf leachates and photosynthates. The main mechanism of energy acquisition from these compounds appeared to be the TCA (citric acid) cycle, as reported in experimental studies of bacterial colonization of the phyllosphere (Müller et al. 2016). Membrane transporters were also reported to be an important component of the epiphytic microbe functional repertoire, maximising the ability to monopolize otherwise limiting resources (Delmotte et al. 2009). The abundance of signal transduction functional pathways, involved in the rapid sensing and response to environmental change, would lastly be coherent with the high variability in conditions of humidity, light and temperature in that microbial habitat (Rastogi et al. 2013).

The low functional variability in microbiota observed among tree species represents a further line of evidence supporting the presence of a core phyllosphere functional microbiota. This low variability, observed even at fine functional levels, could be the consequence of essentially similar constraints imposed by the generally harsh leaf environment on its microbial communities, regardless of the specific physiological traits of the host plant species. This low functional turnover among communities was also associated with a low taxonomic turnover, contrasting with reports from phyllosphere-associated temperate systems where species identity was a

strong driver of taxonomic composition of the microbial communities (Laforest-Lapointe et al. 2016). These results could be explained by a finer-scale partitioning of taxa among neotropical than temperate tree species, or a greater overlap in species functional types limiting strong associations between microbial taxa and their hosts. Such differences should be further investigated.

Despite the high levels of convergence in microbial functions among the phyllospheres of different tree species, several lines of evidence support a role for plant species taxonomic and functional identity in driving microbial community assembly. Tree traits explained a notable portion of the functional turnover among microbial communities. Traits correlated with microbial functional turnover (e.g. leaf area, leaf mass per area) are mostly part of the leaf economics spectrum (Wright et al. 2004), a functional strategy scheme describing photosynthetic resource-use efficiency in plants, which is coherent with what we know of phyllosphere microbial physiology. The ability of a tree to be conservative of its resources and generate thicker and better protected leaves (i.e. high leaf mass per area) is likely to limit the leaching of nutrients from the leaf to the phyllosphere, in turn constituting a filter on resource-use strategies in microbes. The high correlation of leaf mass per area with turnover in microbial communities is coherent with a previously described role for cuticle characteristics in determining functional turnover among leaf microbial communities (Hunter et al. 2010, Bodenhausen et al. 2014). The high correlation of aluminum and copper concentrations in leaves with microbial functional variation may be explained by their role as antibiotics. The predominance of two-component systems associated with high aluminum and copper concentrations suggests that the ability to sense and quickly respond to fluxes in these elements at the cell surface might constitute an efficient stress-response to deal with these conditions (Kaczmarczyk et al. 2014). This type of plant trait gradient is analogous to the leaf chemical gradient described by Yadav and colleagues (Yadav et al. 2005), who reported variation in leaf colonization by phyllosphere microbes on different tree species as a function of their total leaf

phenolics content. Taken together, these interpretations are concordant with the importance of energy metabolism, secondary metabolites and antibiotics production as well as environmental sensing in driving functional turnover of microbes among tree species.

Other lines of evidence support the idea that the plant host plays a selective role on microbial community assembly, such as the detection of bacterial traits that exhibit a strong phylogenetic signal with respect to the host plant phylogeny. While this pattern might arise from the filtering of microbes on phylogenetically structured selective plant traits or from co-evolution of the two partners, it is regardless indicative of an influence of the host on the functional make-up of bacterial phyllosphere communities. Interestingly, the set of pathways that are important in driving functional turnover among communities belong to the same functional categories as the ones that are phylogenetically structured among plant hosts, supporting the proposed match between these bacterial functions and their host's functional and taxonomic identity. The fact that the relative abundance of a large set of functions was different within communities than that expected by chance given their relative abundance across samples, also supports a role for individual tree species in structuring the functional composition of their phyllosphere bacterial communities. The higher filtering of most microbial taxa relative to microbial functions suggests a role for unmeasured trait variation in driving functional turnover among communities.

The relatively small but significant contribution of functional turnover among microbial communities to the total functional diversity observed across samples suggests that the functions that are of importance in driving the distribution of bacteria across different host trees are actually relatively few compared to those enabling the bacteria to pass the overall "phyllosphere filter" that is needed to survive in the phyllosphere habitat. It remains unknown whether the majority of functional pathways that do not vary among trees are actually important for the ecology of the microbes, or

if that trait variation is adaptively neutral within communities. It is also possible that some pathways important for microbial adaptations to leaf physiological gradients are not yet functionally described and are part of the large number of sequences that could not be functionally annotated. Ongoing efforts to better characterize gene functions will help improve the precision of ecological inferences in environmental metagenomes.

3.6 Conclusions

In conclusion, we have identified a core functional microbiota in the phyllosphere of neotropical trees. While most functional variation was observed within individual microbial communities, we reveal a functional matching between the traits of microbes and the traits of plants across 17 tree species, emphasizing the role for energy metabolism, secondary metabolites and antibiotic production as well as environmental sensing in mediating bacterial adaptation to leaf trait gradients in the canopy. Our identification of the adaptive drivers of phyllosphere microbial community composition in this neotropical ecosystem represents a good starting point for identifying the types of microbial traits that could be routinely studied by phyllosphere microbial ecologists to address global questions on the ecological and evolutionary dynamics of phyllosphere microbes. Empirical testing of the fitness consequences of variation in those traits will represent an important next step in understanding adaptive processes in the phyllosphere.

3.7 Acknowledgements

We thank Travis Dawson and Luis Tovar for assistance with sample processing and sequencing.

3.8 Data and code availability statement

The datasets generated and/or analysed during the current study are available in a MG-RAST repository: https://www.mg-rast.org/linkin.cgi?project=mgp91848. The scripts used to perform analyses for the current study are available in a GitHub repository: https://github.com/glajoie1/panama_metagenome.

CHAPTER 4

HOST NEIGHBOURHOOD SHAPES BACTERIAL COMMUNITY ASSEMBLY AND SPECIALIZATION ON TREE SPECIES ACROSS A LATITUDINAL GRADIENT

(Article under review at *Ecological Monographs*)

Geneviève Lajoie and Steven W. Kembel

Département des Sciences Biologiques, Université du Québec à Montréal
141 Avenue du Président-Kennedy, Montréal, Canada, H2X 1Y4

4.1 Abstract

Hosts shape microbial diversity in natural ecosystems by acting as selective filters on environmental microbes. Selection gradients are also observed among host species and genotypes, with turnover in the composition of microbial communities and variation in the strength of symbiont specialization taking place as a function of host taxonomy and phenotypes. With many studies focusing on pairwise interactions between hosts and symbionts, little is understood about the influence of the host community as a whole in shaping host-microbe interactions and how these may change according to the environmental context in which these interactions take place.

Here we investigate the role of host community composition at the local and regional scale in modulating turnover in phyllosphere bacteria among 33 tree host species and within sugar maple (*Acer saccharum*) across a large-scale transition from deciduous to boreal forest. We hypothesize that mass effects from functionally different host species across the landscape will limit the strength of the match between a focal host species and its microbiota. We further hypothesize that temperature stress gradient, host filtering strength and host functional diversity will affect the prevalence and strength of specialization of bacteria on their hosts across this heterogeneous landscape.

We found that the presence of alternate dominant host species at the local and regional scales influenced the composition of symbiont communities on individual host populations and species. Host traits were important drivers of microbial community composition across the landscape, such that we found the strongest differences in the composition of sugar maple microbiota when co-occurring with functionally contrasting species such as conifers. The relative abundance of the focal host species was similarly one of the most important predictors of the specialization of its microbiota, along with the temperature gradient. These results suggest that the

transmission of phyllosphere microbes from the dominant tree community members may be constraining the match between a species and its symbionts, particularly at its range limits. This study provides important insights for predicting host-symbiont mismatches with variation in the distribution of species as a result of climate change.

4.2 Introduction

Host-microbe associations are widespread across plants and animals (Paster et al. 2001, Huttenhower et al. 2012). Hosts influence microbial diversity structure within ecosystems by selecting specific sets of microbial taxa from the environment, as evident by differences in the identity of microbes living inside or outside host tissues and the environmental matrix surrounding the host (Smalla et al. 2001, Hentschel et al. 2002, Björk et al. 2013). Though some members of host microbial communities are invariable among host genotypes and species (the *core* microbiota) (Schmitt et al. 2012, Ainsworth et al. 2015, Yeoh et al. 2017), host taxonomy and traits also drive microbial species sorting among host genotypes and species (the *peripheral* microbiota) (Ley et al. 2008, Hunter et al. 2010, Kembel et al. 2014). Evidence for adaptive matching between host and microbial traits in maintaining these associations is also emerging (Lajoie et al. 2020).

Despite these advances in understanding drivers of symbiont species sorting across hosts, little is still understood of the relative importance of source-sink dynamics from the host neighbourhood in affecting host-microbe associations. At a local scale, a few studies have provided evidence suggesting that the sharing of habitats among different animal species could make their microbiota more similar through direct or indirect contact (e.g. van Veelen et al. 2017, Perofsky et al. 2019, but see Ivens et al. 2018). In a forest ecosystem, transfer of gut bacteria among mammal species depended on the type of niche occupied, with the more connected terrestrial mammals sharing more microbes than arboreal species displaying less contact with each other (Perofsky et al. 2019).

The relative influence of species sorting versus source-sink dynamics is expected to vary with variation in the identity and relative abundances of co-occurring host species of a focal host across space. Some studies have shown variation in the

composition of the microbial community of focal hosts across large spatial gradients (Finkel et al. 2011, Laforest-Lapointe et al. 2016), but the extent to which such variation can be explained by variation in abiotic factors (e.g. climate), variation in traits of the host, or variation in the biotic context experienced by the symbionts (e.g. host community structure) remains largely unknown. Comparing the prevalence of these processes among, as well as within, host species can inform us about the potential importance of these processes in affecting adaptation and evolution in a host species.

With several microbial strains associating with a limited number of related host genotypes or species (Konno et al. 2011, De Mares et al. 2017, Eck et al. 2019), hosts also appear to structure microbial diversity by shaping opportunities for microbial specialization. Specialization is here defined as the level of phylogenetic similarity among hosts that a given symbiont associates with (Jorge et al. 2017). In some cases, variation in host genotypes and traits is also linked with patterns of diversification in their microbes (Lei and Olival 2014, Miyake et al. 2016). Studies of these questions have mostly focused on a small number of pathogenic or strongly symbiotic microbial partners. As a result, we have little understanding of the drivers of microbial specialization on their hosts and the extent to which this process is important in structuring microbial diversity in natural systems.

As for microbial community composition, the ecological contexts in which host-symbiont interactions take place is expected to affect the quality of the match between microbes and their hosts, leading to variation in the strength of microbial specialization across a spatially structured landscape (Thompson 2005, Mihaljevic 2012, Chamberlain et al. 2014). First, the Stress Gradient Hypothesis predicts that with an increase in stress in a habitat, we should observe an increased frequency of mutualistic interactions between partners (Bertness and Callaway 1994). In an extension of that predictive framework, O'Brien et al. (2015) suggested that an increase in mutualistic interactions should be conducive to co-adaptation, which should in turn

favor greater specialization of symbionts on their host and vice-versa. Stressful conditions can be defined as conditions where organisms are under strong selective pressure, resulting in a reduction in fitness of populations (Hoffman and Hercus 2000). Using this concept, the level of stress can be evaluated as the level of phylogenetic clustering in the microbial community, representative of the strength of filtering experienced by the community (Webb et al. 2002, Horner-Devine and Bohannan 2006).

In host-associated microbes, stress can be qualified in two ways. We can first think of abiotic stress gradients defined as variation in environmental conditions that impact microbial fitness. For example, growth at 0°C was linked with longer generation times and higher nutrition requirements than at 10-35°C in aquatic bacteria (Wiebe et al. 1992). We can also think of stress gradients experienced via host phenotypes, for example through variation in host traits affecting carbon or nutrient availability for microbial communities such as cuticle thickness in plants (Lindow and Brandl 2003), or gut pH in animals (Ilhan et al. 2017).

Second, context-dependence in the extent of symbiont specialization can also take place through variation in source-sink dynamics across the landscape, since variation in the relative abundance of alternative hosts among sites can affect opportunities for specialization of symbionts to a focal host (Agrawal 2004). For example, butterfly larvae specialize by feeding on a given plant species when it is more abundant regionally (Kuussaari et al. 2001). This effect was explained by an increase in direct encounter rates, but also by a greater adaptation of populations to regionally more abundant plants. The spatial scale at which such effects are likely to be observed however seems to vary among study systems (Tack et al. 2014).

Lastly, in conjunction with these density-dependent effects, the functional composition of the local or regional host community can also constrain the evolution of specialization. A greater availability of functionally similar hosts (lower host

functional diversity) could increase opportunities for host species shifts by the symbionts, decreasing opportunities for specialization (Agrawal 2006). Alternatively, in examining the evolution of multi-host parasites, the ability to exploit functionally different hosts is likely to be constrained by trade-offs between traits involved in exploiting the different hosts, such that specialization should be more likely (Gandon 2004).

4.2.1 Hypotheses and predictions

Here, we examine the role for the local and regional host community composition on the assembly of their bacterial microbiota through an extensive study of associations between temperate tree species and their leaf bacterial communities in eastern North America. We first evaluate the role of host identity, host community composition and climatic variation in determining leaf bacterial community composition on more than 30 tree species across a latitudinal gradient spanning 5°C in mean annual temperature. We then test different hypotheses about variation in the prevalence and strength of bacterial specialization on their hosts across the landscape among and within host species (Fig. 4.1).

We expect that host community composition (tree community structure) will be important in determining bacterial community composition of individual host tree species. We predict that bacterial composition of focal host species will become more similar to that of neighbouring host species as the focal host becomes less abundant. This would happen as a result of an increased exposition of the focal host to bacterial communities from neighbouring species when the focal host is less abundant. Epiphytic plant bacteria being dispersible as aerosols by wind and clouds (Lindemann and Upper 1985), we expect source-sink dynamics to take place at local, but also at regional scales.

We next hypothesize that bacterial specialization will vary both as a function of host taxonomy and traits and as a function of the ecological context in which the

host-symbiont associations take place (Thompson 2005, Chamberlain et al. 2014). First, following the Stress Gradient hypothesis, we predict an increase in specialization with an increase with abiotic and biotic stress for the bacterial communities. More specifically, we predict that specialization should increase with a decrease in mean annual temperature. It should also vary as a function of the leaf physiology of the host, namely be higher on hosts that have more conservative resource-use strategies and that exhibit higher filtering of their bacterial communities.

We lastly expect that as for bacterial community composition, the level of specialization of bacteria on a given tree host species will vary as a function of the relative abundance and functional diversity of its neighbouring tree species (Agrawal 2004, 2006). We predict that specialization will be higher when the focal host is more abundant locally and regionally. This would result from the opportunity to have more consistent interactions with the focal hosts, facilitating specialization. By extension, we also predict that bacterial specialization will decrease as a function of the functional diversity of the host community, which should relax selection on specialization to the focal host phenotypes.

.

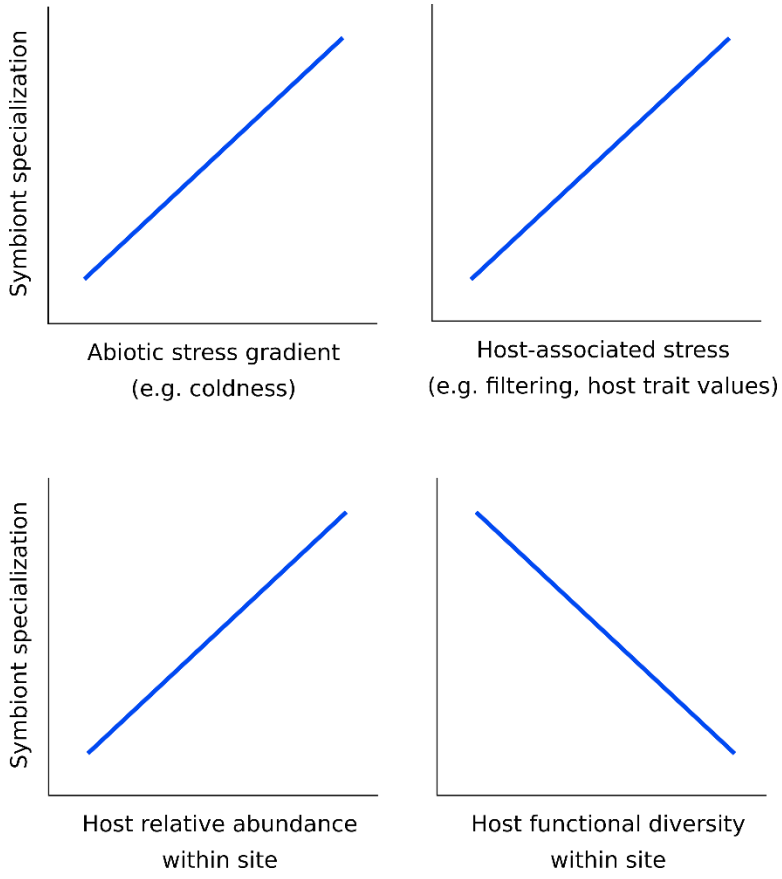

Figure 4.1 Predictions on the context-dependence of symbiont specialization on their host.

4.3 Methods

4.3.1 Data collection

4.3.1.1 Sampling design

During summer 2017, we sampled epiphytic microbial communities from the leaf surface (phyllosphere) of sugar maple (*Acer saccharum*) and its surrounding tree communities across the north-eastern portion of its range. Nine sites were laid out in a grid across Quebec, eastern Ontario and north-eastern United States, each one separated by ~150 km from its closest neighbour (Annex E: Fig. S4.1a). Mean annual temperature over the 1981-2010 period was determined for the station nearest to each site using publicly-available regional climatic data from Environment Canada (Gouvernement du Canada 2017) for Canadian sites, and the United States National Center for Environmental Information (National Atmospheric and Oceanic Administration 2017) for the Vermont site (Table 4.1). At each site, we established three 20 x 20 m plots in stands where sugar maple was dominant (>70% cover), and three in stands where it was rare (<30% cover). In Québec, we located such stands using ecotype forest maps from the Ministère des Forêts, Faune et Parcs of Québec (Direction des inventaires forestiers du Ministère des Forêts Faune et Parcs 2015). Within each site, all plots were separated by at least 800 m.

We estimated regional abundance of each tree species in our species pool in a 25 km radius around median geographic coordinates of plots at each site (Annex E: Fig. S4.1b). For Quebec sites, we estimated tree species relative regional abundance at each scale using ecotype forest maps, describing tree species composition at the stand level across the province from aerial photographs and field surveys (Direction des inventaires forestiers du Ministère des Forêts Faune et Parcs 2015). We obtained similar information on tree species composition around our Ontarian sampling site from the Ontario Forest Resources Inventory (Ontario Ministry of Natural Resources

and Forestry 2007). For the Vermont site, we evaluated regional tree species relative abundances through the Forest Inventory and Analysis Database of the United States Forest Service, consisting in a network of plots in which individual tree species are identified and measured (US Department of Agriculture - Forest Service - Northern Research Station 2019). All of these georeferenced data were analyzed with ArcGIS v.10.5 (ESRI 2011).

Table 4.1 Description of sampling sites. Latitudes and longitudes represent the centroid of latitude and longitude for all plots at that site. Mean annual temperature was recorded for the closest governmental meteorological station.

Site name	Abbreviation	Latitude (UTM)	Longitude (UTM)	Mean annual temperature (1981-2010) (°C)
Parc des Monts-Valins (Québec, Can.)	VAL	48.60088	-70.90967	2.8
Parc de la Vérendrye (Québec, Can.)	VER	47.08255	-76.52711	3.1
Centre écologique La Huardière, Saint-Michel-des-Saints (Québec, Can.)	SMS	46.67675	-74.14121	3.1
Parc de la Jacques-Cartier (Québec, Can.)	JAC	47.17828	-71.38214	3.4
Parc du Mont-Mégantic (Québec, Can.)	MEG	45.44989	-71.11256	4.0
Parc de la Gatineau (Québec, Can.)	GAT	45.60540	-76.04605	5.6
Gault Nature Reserve, Mont-Saint-Hilaire (Québec, Can.)	MSH	45.54956	-73.13212	6.1
Proctor Maple Research Center (Vermont, USA)	PRO	44.52747	-72.86562	7.4
Frontenac Provincial Park (Ontario, Can.)	FRO	44.52042	-76.53622	7.8

4.3.1.2 Microbial community sampling and DNA analysis

Within each plot, we determined tree composition by visually estimating the relative abundance (% cover) of each woody species for which we observed at least one individual >1 m height. We then sampled three individuals per species present for both leaf microbial communities and tree traits. We first used sterile gloves and a pole pruner to collect leaves from each of the three individuals of each species at a height of ~5 m in the canopy to sample microbial communities. All leaf samples from the same species in each plot were mixed together in a sterile bag and kept on ice until brought back to the lab and refrigerated in the dark at 4°C.

Within two days of field collection, epiphytic microbes were washed from leaf surfaces using 100 mL of a 1:50 dilution of Redford buffer (1M Tris, 500 mM EDTA, 1.2% CTAB). The suspension was centrifuged at 3300 g for 25 minutes after which the pellet was collected and added to an extraction tube before being frozen at -20°C until DNA extraction. Microbial DNA was extracted using QIAGEN Powersoil kits, following the instructions of the manufacturer. Using a 1-step PCR protocol adapted from Fadrosh et al. (2014), we amplified the bacterial V5-V6 region of the bacterial 16S rRNA gene to determine community composition using primers 799F and 1115R (Kembel et al. 2014). These primers exclude plant chloroplasts and cyanobacterial sequences and thus avoid PCR contamination by host plant DNA amplification (Rastogi et al. 2010, Redford et al. 2010). Their use is justified since cyanobacteria are rare members of the tree phyllosphere (Delmotte et al. 2009, Vorholt 2012). Twenty-five microliter PCR solutions consisted of 5 µL 5X Phusion HF Buffer (Thermo Fisher Scientific), 0.5 µL dNTPS (10µM each), 0.75 µL DMSO, 0.5 µL each primer (10 µM), 0.25 µL Phusion Hot Start II polymerase (2U/µL) (Thermo Fisher Scientific), 1 uL genomic DNA template and 16.5 uL molecular-grade H_2O. Reactions were performed for each sample using the following conditions: 30 s initial denaturation at 98 °C, followed by 35 cycles of 15 s at 98 °C, 30 s at 64 °C, and 30 s at 72 °C, with a final 10-

min elongation at 72 °C. PCR products were normalized using a SequalPrep Normalization kit (Thermo Fisher Scientific), pooled and then purified using AMPure (Beckman Coulter Life Sciences) to remove contaminants. The DNA library was prepared by mixing equimolar concentrations of DNA from each sample, and then sequenced using Illumina MiSeq 250-bp paired-end sequencing. In total, 53 samples of sugar maple, and 270 samples of other coexisting species were obtained across sampling sites, for a total of 323 samples of epiphytic bacterial communities spanning 34 tree and tall shrub species.

4.3.1.3 Tree trait measurement

We sampled a range of tree traits representative of different plant adaptive axes to characterize the host community at each plot. In each plot, we measured wood density on three individuals per species as a trait indicative of the stability, defence, carbon gain and growth potential of trees (Sungpalee et al. 2009). For example, fast growers will typically have low wood density. While this trait value has a lower cost and higher hydraulic capacity, it is also associated with lesser biomechanical strength and resistance against pathogens. For measuring this trait, we extracted a wood core from the same three individuals per plot for sugar maple, and from a total of 5 individuals per site for each other species using a wood corer. In the lab, we rehydrated the wood cores for 30 min. We then measured their volume by water displacement in a graduated cylinder (Chave 2005). Samples were then dried at 100°C for 72h before being weighed. Wood density was calculated as the ratio between the dry mass and the fresh volume of each core.

We also measured leaf traits across our range of plots and tree species. From the same branch where microbial communities were sampled for each tree, we collected the three largest leaves and pooled them by species and site for further measurements. Upon collection, we first measured leaf area by scanning each leaf and determining its area using ImageJ. We then dried the samples at 70°C for 72h and

measured their dry mass in order to calculate specific leaf area (SLA), the ratio between the fresh area of a leaf and its dry mass. While leaf area is related to a plant's light capture strategies, SLA is a proxy for photosynthetic efficiency and resource-use vs. conservation tradeoffs (Reich et al. 1994). Using these samples, we also measured leaf nutrient concentrations (Ca, Mg, K, P) for all sugar maple samples, and for one sample per other tree species per site, totalling 155 samples. To measure leaf nutrient concentrations, we first ground the leaf material from each of these samples into fine powder and digested with sulfuric acid and hydrogen peroxide (Parkinson and Allen 1975). Ca, Mg and K were measured on 1/20 dilutions of digested samples with flame atomic absorption spectrometry (Varian SpectrAA 220FS, Varian Australia). P was measured on digested samples through the colorimetric automated method with ammonium molybdate and ascorbic acid (Lachat Instruments, Hach USA).

4.3.2 Data preparation

4.3.2.1 Bacterial community composition

We obtained a total of 10.1M sequences across our 323 bacterial samples. Sequence processing, amplicon sequence variants (ASV) identification, and taxonomic annotation were carried out using the R package ‘DADA2’ version 1.10 (Callahan et al. 2016) using default parameters except where noted. In order to calculate a community composition matrix, we first filtered the sequences based on read quality and trimmed the forward reads at 210 bp and the reverse reads at 170 bp to get rid of low-quality tails. We then identified amplicon sequence variants (ASV) in each sample using pseudo-pooling, in which information is shared among samples and then each sample processed independently. Pseudo-pooling was used to increase the sensitivity of sample inference to sequence variants that may be present in several samples but at very low frequencies. We next merged paired reads and removed non-target sequences that were shorter than 293 bp or longer than 322 bp along with chimeras. These operations resulted in a final set of 9.2M sequences from 18,155 ASVs across all

samples, with a median of 29,340 sequences and 497 ASVs per sample. Taxonomy was assigned to ASVs by comparison with the SILVA SSU r132 database, confirming that all sequences were of bacterial origin and did not contain chloroplasts. For subsequent analysis, we rarefied the bacterial community composition matrix to 7,000 sequences per sample using the R package 'vegan' (Oksanen et al. 2013), leading to the removal of 27 samples that contained fewer than 7,000 sequences and reducing the host species pool to 33 species (Annex F: Table S4.1). After rarefaction, 2,072,000 sequences and 14,777 ASVs remained and were used for all subsequent analyses.

4.3.2.2 Bacterial phylogeny

We built a 16S bacterial phylogeny using the sequences of all ASVs in our dataset. We first aligned the sequences to a GreenGenes core set alignment, then filtered them to eliminate positions that were gaps in 80% of the sequences using QIIME (Caporaso et al. 2010). We then inferred approximately-maximum-likelihood phylogenies with the GTR+CAT model using the FastTree2 tool and picked the phylogeny with highest likelihood (Price et al. 2010). We removed 167 ASVs with extremely long subtending branches in the phylogeny that resulted from poor alignment of the associated ASV sequences.

4.3.2.3 Tree trait community weighted means and functional diversity

In order to limit collinearity in our tree trait dataset, we evaluated correlations between all traits measured across all samples and kept only a subset of traits for which the correlation coefficient was less than 0.5 between each other. As a result, our final tree trait dataset only includes wood density, specific leaf area, leaf calcium concentration and leaf phosphorus concentrations. For each of these traits, we then calculated average plot-level population values by averaging trait values across all individuals of each species per plot. We also calculated site-level species mean traits by averaging a species' measured trait values across all plots where it was present for

each site. Using these species average values, we then calculated weighted mean trait values for each plot by weighing the trait values of each species in the plot by their plot relative abundance and summing these values across species. We calculated site weighted mean trait values in the same way, using the species regional relative abundance as weights.

Lastly, we calculated host functional diversity using the functional dispersion metric (FDis) for each plot and site (Laliberté and Legendre 2010). Functional dispersion represents the average distance of each species to the species centroid in a PCoA space calculated using trait dissimilarities between species. Functional dispersion metrics were calculated using a trait dissimilarity matrix generated from trait values of all five traits for each plot and site. These calculations were performed using package 'FD' in R (Laliberté and Legendre 2010).

4.3.2.4 Tree host phylogeny

A phylogeny for the sampled tree species was constructed by concatenating the dated angiosperm and gymnosperm phylogenies of North American trees published by Ma and colleagues (2016). Concatenation of the two trees was performed by grafting the angiosperm tree onto the gymnosperm tree with a divergence time set at 377 Mya (Ran et al. 2018) using R package 'ape' (Paradis and Schliep 2018).

4.3.3 Analyses

4.3.3.1 Drivers of bacterial community composition across the landscape

We first evaluated whether there were differences in bacterial community composition as a function of the plot, plot type (high- or low- sugar maple abundance) and site. We calculated a Bray-Curtis distance matrix from the rarefied bacterial community composition of each sample on which we performed a PERMANOVA analysis using plot nested within type and site, type nested within site and site as

explanatory factors. We also evaluated the role of host taxonomy in driving differences in bacterial communities among samples using a PERMANOVA with host species, genus and family as nested explanatory factors.

We then tested whether variation observed respectively among plots and among sites could be linked with variation in host species identity and host community composition, while controlling for environmental differences between them. We thus conducted a variation partitioning analysis of the bacterial community composition data, as a function of host species identity, plot-level tree community composition, site-level tree community composition and environmental variation among samples (mean annual temperature). We conducted individual models per spatial scale, incorporating host species identity or abundance, tree trait measures and mean annual temperature, to tell which tree traits were most important in explaining these patterns. All analyses were performed using R package 'vegan' (Oksanen et al. 2013).

4.3.3.2 Bacterial specialization on their tree hosts

We evaluated bacterial specialization on their hosts for each ASV at each site using the rescaled version of the distance-based specialization index (DSI) (Jorge et al. 2014, 2017). The DSI metric is a Z-score comparing the mean phylogenetic distance (MPD) between all host species on which an ASV was found weighted by their relative abundance on each host, to a null distribution obtained by randomly sampling hosts from the host pool (n=999). The metric is then scaled by the maximum or minimum value obtained for that ASV to obtain DSI*, a metric ranging between -1 (fully generalist ASV) to +1 (fully specialist ASV). This scaling enables a more accurate comparison of scores among ASV and datasets with different sample sizes. While the DSI* metric controls to some extent for the number of host species available, it tends to be less stable with fewer species of hosts (Jorge et al. 2017), so we performed these analyses at the site rather than the plot-level, to include as many species as possible in the calculations. In doing so, we first selected only the 8 species that were the most

abundant at each site, to standardize the number of host species across sites. For each of these site and species combinations, we used the sample from the plot where the species was the most abundant. To eliminate diversity effects due to variation in the size of the ASV pool among sites, we performed these analyses using only the 500 most abundant ASVs per site. We tested for an effect of site in driving variation in bacterial specialization using an analysis of variance, followed by Tukey's tests to test for differences in mean specialization among pairs of sites.

4.3.3.3 Drivers of bacterial specialization across the landscape

We evaluated the role of stress gradients, host functional diversity and a focal host local and regional abundances in driving bacterial specialization across sites. In order to compare microbial specialization with characteristics of the host, we calculated weighted averages of ASV specialization scores for each host species at each site, which we used in subsequent analyses. We defined both abiotic and host-associated stress gradients among host species and sites. Abiotic stress was defined as the gradient in mean annual temperature among sites. Host-associated stress was defined two-fold. Firstly, we evaluated the strength of bacterial selection by the host, in terms of the phylogenetic clustering of the bacterial community. Using the same samples for which specialization was measured, we calculated for each host species within each site the mean phylogenetic distance (MPD) between all ASVs of its bacterial community and compared this distance to a null model where the ASV labels were shuffled across a bacterial phylogeny of all ASVs recorded at that site to obtain a Z-score (n=999). We performed these analyses using R package 'picante' (Kembel et al. 2013). Secondly, we were interested in linking this proxy of host-associated stress to trait variation among hosts, since host traits determine the type of resources available for the microbes on their leaves (e.g. Ca or P) or impediments to colonization (e.g. the thicker wax layer associated with leaves with lower specific leaf area).

We used the specialization scores calculated above and the abiotic and host-associated stress proxies to evaluate how the prevalence and strength of ASV specialization on their host would vary across host species and sites. We first tested whether bacterial specialization varied as a function of mean annual temperature using a general linear model. We then added species identity as a variable to this model to evaluate the extent to which variation of bacterial specialization along the temperature gradient was related with a turnover in host species identity among sites. We compared the variance explained by temperature in the two models using analyses of variance.

We used individual regression models to evaluate the effect of host individual- and site-level trait variation on bacterial specialization. We then built a general linear model to explain variation in the mean specialization score of all ASVs on a given host by the filtering on that host species (ASV phylogenetic clustering). We tested the role of host functional diversity at the site level in driving specialization patterns using a similar model, with site-level functional diversity as an explanatory variable.

We lastly evaluated the effect of host local and regional abundance on specialization patterns in a focal host species (*A. saccharum*). First, we reran the calculations of the specialization metrics at each site using only the ASVs that were found on sugar maple. In order to get specialization estimates for each sugar maple sample while keeping the ASV pool from other host species constant, we reran these calculations at each site using in turn every sugar maple sample from that site. For the following analyses, we only used ASVs that were recorded at each site. We used analyses of variance of general linear models to test the influence of sugar maple relative abundance and bacterial ASV identity on specialization of sugar maple associated bacteria. We ran one such model using plot relative abundance, and another using regional relative abundance of sugar maple to evaluate their respective effects. We also tested the interactions between relative abundance and ASV identity to account for intraspecific variation in ASV specialization as a function of host relative

abundance. All ecological analyses were performed using R version 3.6.1 (R Core Team 2013).

4.4 Results

4.4.1 Drivers of bacterial community composition across the landscape

Bacterial community composition of the tree phyllosphere varied as a function of host taxonomy, plot-level host relative abundances and site-level host relative abundances, with only a small portion of this variation being jointly explained by covariation of tree species abundances with mean annual temperature across the gradient (Table 4.2, Fig. 4.2a, Fig. 4.3a-c). Similar patterns were observed when considering sugar maple samples alone (Table 4.2, Fig. 4.3d-f). One of the major gradients in bacterial community composition among samples was a differentiation between gymnosperm (top-left) and angiosperm (bottom) microbiota (Fig. 4.3a). This transition was also observed among sites, across the transition from old-growth deciduous and mixed forests with large abundances of sugar maple and red oak (*Quercus rubra* (L.)) to the boreal forest dominated by balsam fir (*Abies balsamea* (L.)) and paper birch (*Betula papyrifera* (Marshall)) (Fig. 4.3c-f). Across this gradient, species belonging to the birch family (Betulaceae) appeared as intermediates (Fig. 4.3a), with more northern plots dominated by yellow birch (*Betula alleghaniensis* (Britt.)) tending to draw the composition of co-occurring species further from conifer-type microbiota (Fig. 4.3b,e).

Variation in the traits of hosts, particularly within and among host species and among site-weighted means, explained a significant portion of the variation in bacterial community composition among hosts (Fig. 4.2b-d, Fig. 4.3). A gradient in specific leaf area and wood density was strongly associated with bacterial community turnover between gymnosperms and gymnosperm-dominated communities and the angiosperms

and angiosperm-dominated communities (Fig. 4.3a-c), while leaf phosphorus concentration was strongly correlated with bacterial community turnover among hosts and host communities dominated by the maple (Sapindaceae) and the oak (Fagaceae) family to the birch family (Betulaceae) (Fig. 4.3a-c).

Table 4.2 Variation in bacterial community composition as a function of sample structure and host taxonomy. Two sets of models were built, the first for all samples and the second for sugar maple samples only. Sample structure includes variation among plots, among plot types (hi- or low- abundance of sugar maple) and among sites. For each model, variation explained by different variables was evaluated using a PERMANOVA test on Bray-Curtis dissimilarity (n=999 permutations). F represents a pseudo-F ratio used to estimate statistical significance of the different variables. Significance levels for each variable are given by: *p < 0.05; **p < 0.01; ***p < 0.001.

All samples

Model	Variables	Bray-Curtis dissimilarities					
		df	Sum of Sqs	R^2 (%)	F	Pr (>F)	
Sample structure	Plot (nested within Type and Site)	33	8.331	11.4	1.270	0.001	***
	Type (nested within Site)	9	4.173	5.7	2.332	0.001	***
	Site	8	8.331	15.9	7.259	0.001	***
	Residuals	245	48.72				
	Total	295	72.772				
Host taxonomy	Species (nested within Genus and Family)	12	6.305	8.7	2.695	0.001	***
	Genus (nested within Family)	10	4.427	6.1	2.271	0.001	***
	Family	10	10.771	14.8	5.525	0.001	***
	Residuals	263	51.270				
	Total	295	72.772				

Sugar maple samples

Model	Variables	Bray-Curtis dissimilarities					
		df	Sum of Sqs	R^2 (%)	F	Pr (>F)	
Sample structure	Type (nested within Site)	9	1.815	18.5	1.056	0.21	
	Site	8	2.093	21.3	1.371	0.001	**
	Residuals	31	5.919				
	Total	48	9.828				

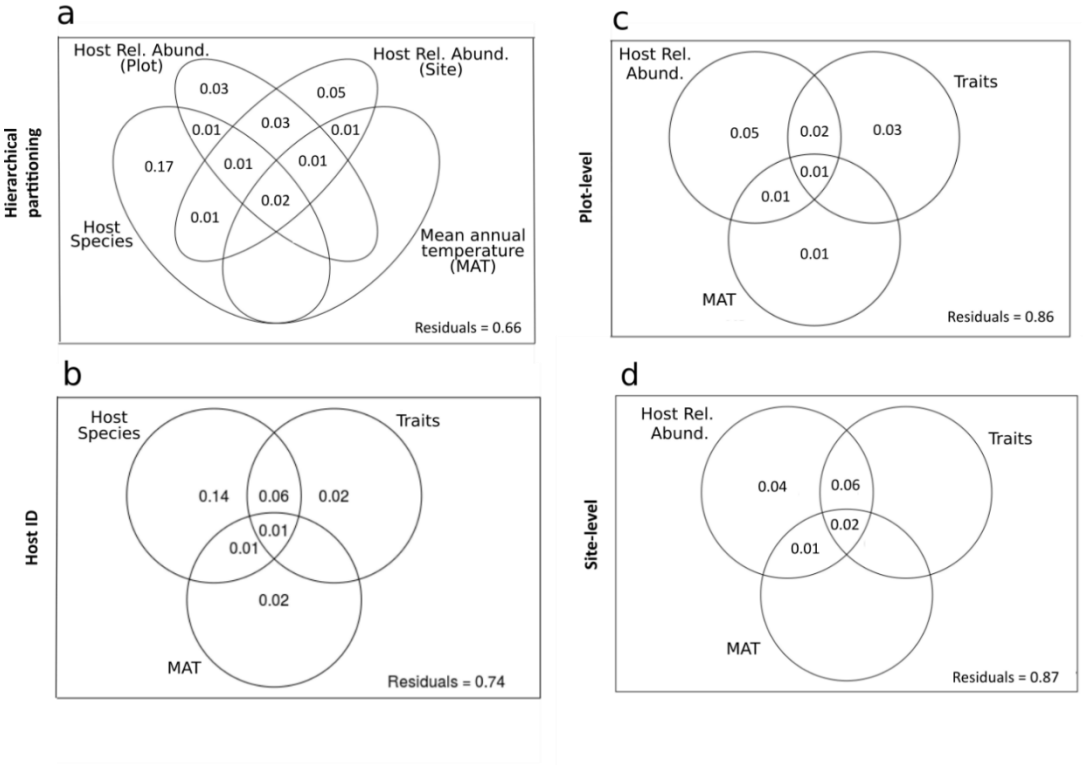

Figure 4.2 Variation partitioning of bacterial community composition (proportion of variance explained) as a function of host identity, host community composition, host traits and mean annual temperature (MAT). Panel a shows variation partitioning based on the host species identity, host community composition and MAT alone. Panels a-d present variation partitioning as a function of host identity or relative abundance of the host and their traits aggregated at three different hierarchical levels: individual host species within plot and site (b), plot-level within site (c) and site-level (d). Only significant portions ($p<0.05$) are shown.

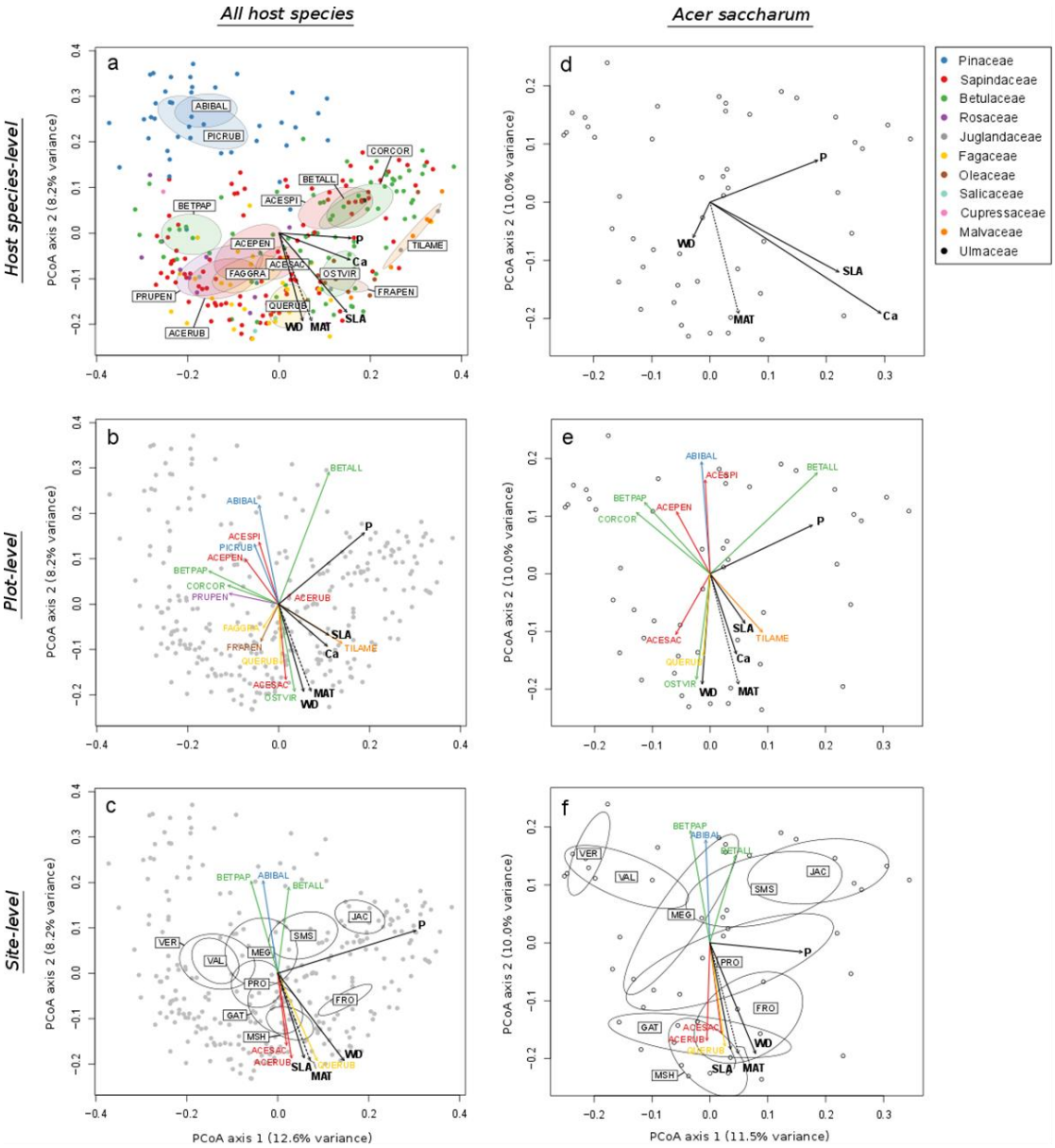

Figure 4.3 Principal components analyses (PCoA) of bacterial community composition of the phyllosphere of 33 tree host species. Two PCoA were calculated using Bray-Curtis distances among samples, one using samples from all species (a-c) and the other using samples from the sugar maple host only (d-f). Each point represents a bacterial community sample from a host species at a given plot and site. Samples are color-coded by the taxonomic family of the host. Ellipses representing the standard error of the average of scores are drawn for host species for which he had at least 5 samples (panel a) and for sites (panels c,f). We fitted host traits and relative abundances,

as well as mean annual temperature (MAT) onto each PCoA in a way that maximises the correlation between these variables and the configuration of the PCoA and plotted them on the graphs. We performed this fitting respectively for individual- (a,d), plot-level (b,e) and site-level (c,f) average host traits (black lines), as well as plot-level and site-level host relative abundances (lines colored by host taxonomic family). Only the host species occurring in more than 3 sites were used for plotted in site-level analyses. Species codes are the following: *Abies balsamea* (L.) [ABIBAL]; *Acer pensylvanicum* (L.) [ACEPEN]; *Acer rubrum* (L.) [ACERUB]; *Acer saccharum* (Marshall) [ACESAC], *Acer spicatum* (Lam.) [ACESPI], *Betula alleghaniensis* (Britt.) [BETALL], *Betula papyrifera* (Marshall) [BETPAP], *Corylus cornuta* (Marshall) [CORCOR], *Fagus grandifolia* (Ehrh.) [FAGGRA], *Fraxinus pennsylvanica* (Marshall) [FRAPEN], *Ostrya virginiana* (Mill.) K. Koch [OSTVIR], *Picea rubens* (Sarg.) [PICRUB], *Prunus pensylvanica* (L.f.) [PRUPEN], *Quercus rubra* (L.) [QUERUB], *Tilia americana* (L.) [TILAME]. Host trait codes are the following: specific leaf area ($mm^2 \cdot mg^{-1}$) [SLA]; wood density ($g \cdot cm^{-3}$) [Wood.dens]; leaf calcium concentration ($mg \cdot g^{-1}$) [Ca]; leaf phosphorus concentration ($mg \cdot g^{-1}$) [P]. Site codes are presented in Table 4.1.

4.4.2 Drivers of bacterial specialization on their host across the landscape

Phyllosphere bacteria showed a tendency for specialization (DSI* > 0) as opposed to generalization (DSI* < 0) (Fig. 4.4). Bacterial specialization on individual hosts varied significantly among host species (Type III ANOVA: $F(10,43) = 2.584$, MSE = 0.022, $p = 0.015$) and host families (Type III ANOVA: $F(3,50) = 5.925$, MSE = 0.022, $p = 0.002$). Phyllosphere microbiota of host species from the maple and birch family showed the highest specialization, while those of the pine and the oak families showed the lowest levels (Fig. 4.4A). Variation in specialization of bacteria on hosts was also observed across sites (Type II ANOVA: $F(8,4491) = 14.254$, MSE = 0.213, $p < 0.001$) (Fig. 4.4B). While we observed a negative correlation between bacterial specialization and mean annual temperature among sites, this relationship was mostly encompassed by turnover in the mean specialization levels of host species across the climatic gradient (Table 4.3, Fig. 4.5a).

The strength of bacterial filtering by a host was not correlated with the level of bacterial specialization on that host (Fig. 4.5b). However, individual host traits were good predictors of its average bacterial specialization (Fig. 4.6). We observed a positive relationship between specific leaf area, calcium concentration and phosphorus with the average bacterial specialization on a host (Annex F: Table S4.2a). In opposition to the patterns observed with bacterial community composition, bacterial specialization was not influenced by the functional characteristics of the host community. Mean site weighted traits did not explain variation in specialization across host species and among sites (Annex F: Table S4.2b), nor did host functional diversity (Fig. 4.5d).

Lastly, we tested the role of local and regional host abundance in driving host specialization in sugar maple. We find that plot-level abundance of sugar-maple did not have any impact on the level of specialization of their leaf microbiota. We however show that regional abundance of sugar maple was positively correlated with the specialization of its phyllosphere bacterial communities, even when controlling for

variation in mean annual temperature (Fig. 4.5c). This effect was linked to differences in specialization among, but also within ASVs long the sugar maple abundance gradient (Table 4.4).

Table 4.3 Variation in mean bacterial specialization (DSI*) on host species among sites along a temperature gradient. The DSI* metric is a standardized score characterizing the phylogenetic range of hosts an ASV interacts with. We used linear regression models to evaluate the influence of mean annual temperature on mean bacterial specialization on hosts per site while controlling or not for variation in host species identity. We evaluated sums of squares explained by each variable using analyses of variance. Only the 8 most abundant host species and the 500 most abundant ASVs per site were included in the model to control for differences in host species and ASV diversity among sites. Significance levels for each variable are given by: *p < 0.05; **p < 0.01; ***p < 0.001.

Model	Variables	df	Sum Sq	Mean Sq	F	Pr (>F)	
One factor model	Mean annual temperature	1	0.213	0.213	7.48	0.008	**
	Residuals	70	1.994				
Two factor model	Mean annual temperature	1	0.024	0.024	1.138	0.292	
	Host species	25	1.051	0.042	2.004	0.021	*
	Residuals	45	0.943				

Table 4.4 Variation in bacterial specialization (DSI*) on sugar maple (Acer saccharum) per site as a function of sugar maple abundance, bacterial ASV identity and their interaction. The effect of sugar maple abundance at the local and at the regional scales were examined in two separate models. The DSI* metric is a standardized score characterizing the phylogenetic range of hosts an ASV interacts with. We used an analysis of variance to determine the variance in specialization of individual ASVs present on sugar maple explained by each factor and their interaction. Only ASVs that were present at all sites across the gradient were included. Significance levels for each variable are given by: *p < 0.05; **p < 0.01; ***p < 0.001.

Model	Variables	df	Sum Sq	Mean Sq	F	Pr (>F)	
Local	Local abundance of sugar maple (SM)	1	0.09	0.14	0.653	0.419	
	Bacterial ASV identity (ASV)	113	182.45	1.61	17.202	<0.0001	***
	SM*ASV	113	4.84	0.04	0.457	1.000	
	Residuals	3328	435.78	0.13			
	Total	3555	623.16				
Regional	Regional abundance of sugar maple (SM)	1	3.40	3.40	29.430	<0.0001	***
	Bacterial ASV identity (ASV)	113	182.74	1.62	19.512	<0.0001	***
	SM*ASV	113	52.62	0.47	5.620	<0.0001	***
	Residuals	3328	384.69	0.12			
	Total	3555	623.45				

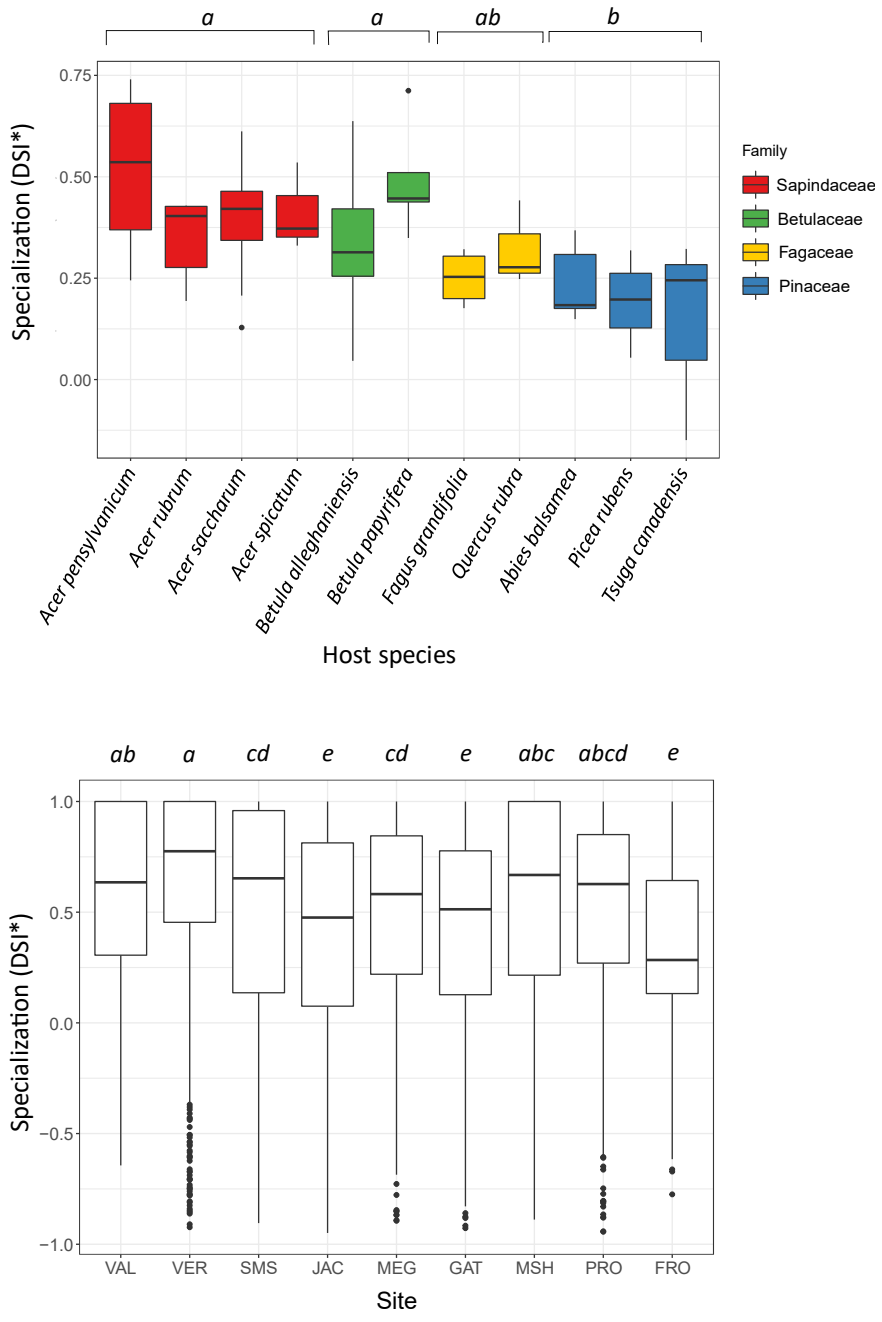

Figure 4.4 Variation in specialization (DSI*) of bacterial phyllosphere communities among the most abundant tree host species (a) and among sites (b). The DSI* metric is a standardized score characterizing the phylogenetic range of hosts an ASV interacts with varies between -1 (fully generalist) and +1 (fully specialist ASVs).

In panel a, data points are mean specialization scores of ASVs per host species per site, weighted by the ASV abundance on the host. We only show host species that were present in at least 3 sites. Different letters among taxonomic families indicate significant differences in their average level of specialization as indicated by Tukey's HSD test on an analysis of variance. Host species are color-coded by the host taxonomic family. In panel b, data points are specialization scores of individual ASVs at each site. Different letters among sites indicate significant differences in their average level of specialization as indicated by Tukey's HSD test on an analysis of variance. Sites were ordered in the figure from coldest (left) to warmest (right). Site codes are indicated in Table 4.1.

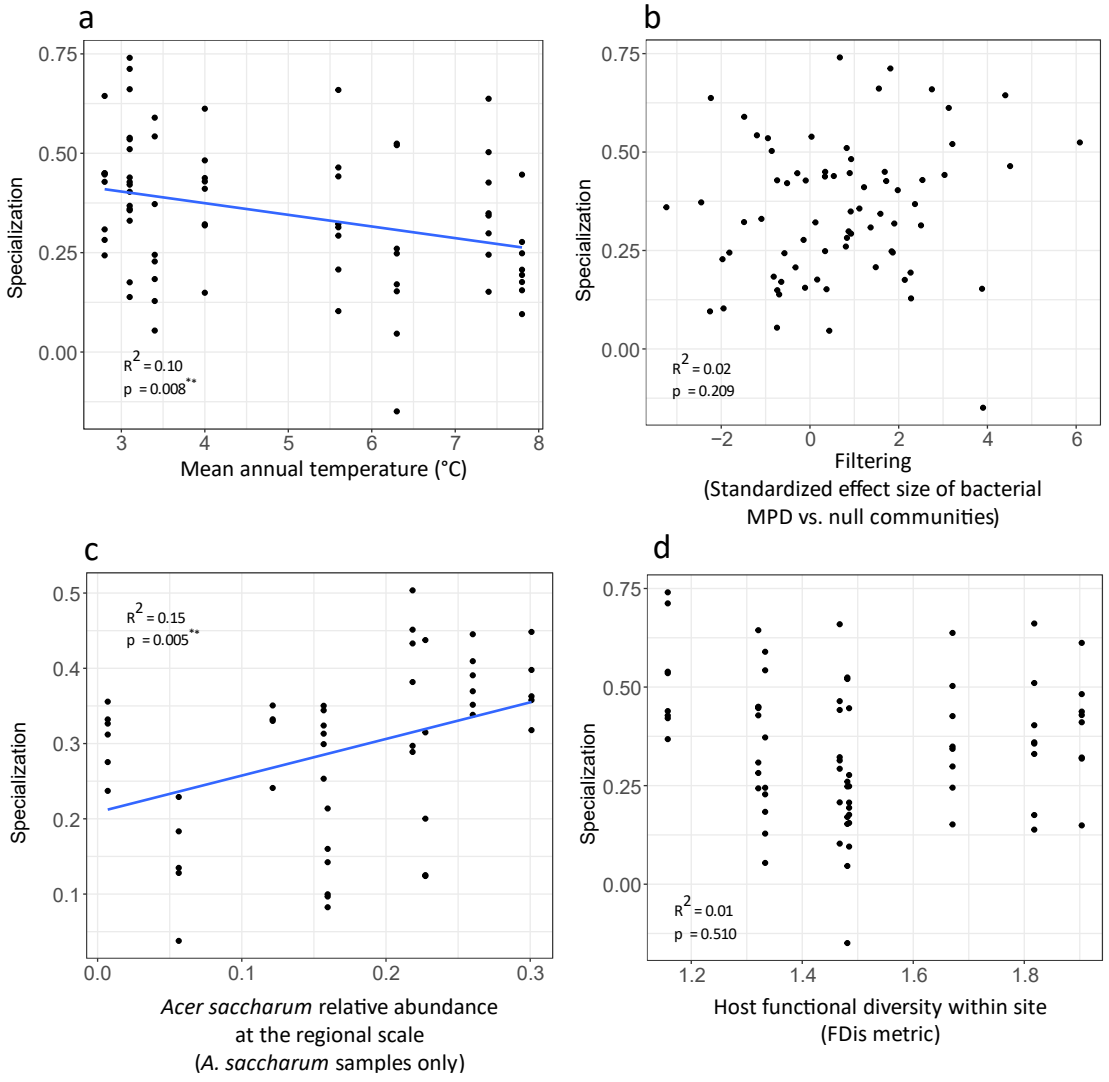

Figure 4.5 Variation in mean specialization (DSI*) of ASV on their host across sites, as a function of mean annual temperature (a), phylogenetic clustering of bacterial communities (b), host regional abundance (c) and host functional diversity (d). The DSI* metric is a standardized score characterizing the phylogenetic range of hosts an ASV interacts with and varies between -1 (fully generalist) and +1 (fully specialist ASVs). Each point represents the mean specialization of ASVs on a host species at a site. In panel b, phylogenetic clustering was calculated as a standardized effect size of mean pairwise phylogenetic distances between all members of the bacterial community as compared to null communities (n=999). The standardized scores vary from negative (less clustered than expected by chance) to positive (more clustered than expected by chance). Panels a, b and d include one sugar maple sample per site, while panel c only includes sugar maple samples, for which specialization scores were calculated for

every sample at every site. In panel d, the FDis metric represents a measure of functional dispersion among all host species at a given site, with a larger metric indicating a larger functional diversity.

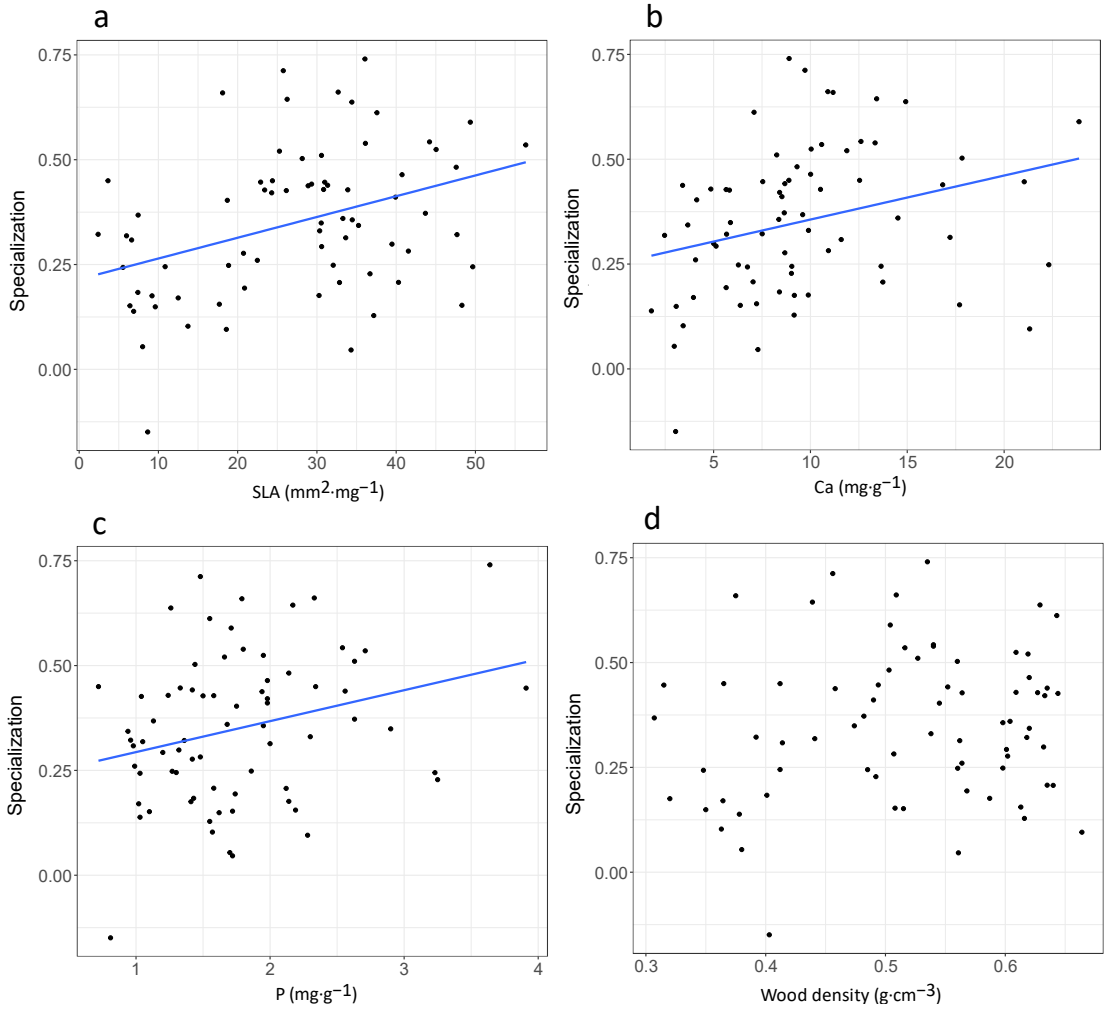

Figure 4.6 Variation in mean specialization (DSI*) of ASVs on their host across sites as a function of individual host traits. The DSI* metric varies between -1 (fully generalist) and +1 (fully specialist ASVs). Each point represents the mean specialization of ASVs on a host species at a site. Each panel represents a different trait. Blue solid slopes indicate significant relationship between specialization and the trait in linear regression models. Detailed results for these models are presented in Annex F: Table S4.2a.

4.5 Discussion

4.5.1 Role for host community structure in driving bacterial community composition

Our results show a role for both bacterial species sorting and mass effects in determining phyllosphere bacterial community composition. Indeed, bacterial community turnover was observed across plant host species and families as a function of their functional traits such as specific leaf area. Still, we also observe a role for plot-level and site-level host community composition and traits in driving bacterial community composition, suggesting that the presence of alternate hosts is affecting the assembly of bacterial communities on a given host. Similar differences in bacterial community composition of sugar maple among samples taken from sites where sugar maple, balsam fir or yellow birch were dominant also support a role for mass effects from abundant host species in phyllosphere community assembly. As such, our results support the presence of a "core" microbiota determined by host taxonomy and traits in the phyllosphere bacterial community, supplemented by a "peripheral" microbiome likely acquired from neighbour host species.

The role of host taxonomy or traits vs. horizontal transmission from contact with other host species in driving symbiont community assembly has been explored in the study of animal gut microbiomes. A few studies have provided evidence suggesting that the sharing of habitats among different animal species could make their microbiota more similar through direct or indirect contact (e.g. van Veelen et al. 2017, Perofsky et al. 2019, but see Ivens et al. 2018). While differences in the microbiota of tree leaves of host species among different sites have been reported (Finkel et al. 2011, Laforest-Lapointe et al. 2016), it was unclear if these patterns result from bacterial sorting in response to abiotic gradients such as temperature or precipitation, or differences in the ecological context experienced by the host. Here, we provide some of the first evidence

that a host's niche overlap with other abundant host species influences community composition of its associated phyllosphere bacteria.

4.5.2 Bacterial specialization across a heterogeneous landscape

We expected fewer opportunities for specialization in epiphytic bacteria from the leaf surface than in other host-microbe systems, because of greater opportunities for movement among hosts and less costly barriers to overcome to interact with the exposed surfaces of the host, both of which should favor generalization. Here however, we show high average levels of specialization of bacteria to hosts in the phyllosphere (specialization is always in the upper part of the specialization gradient ($DSI^* > 0$)), even when evaluating only the most abundant bacterial strains at each site, suggesting that there would still be potential benefits for a microbe to specialize on resources provided by a given host species in the phyllosphere.

Specialization of phyllosphere bacteria on their hosts was largely determined by tree species identity and linked to variation in their traits, with an increase in specialization on species with higher SLA and higher leaf calcium concentrations. Despite this influence of individual host traits, we did not find that specialization was linked to variation in the strength of filtering of hosts on their symbionts. These results suggest that the selection of phylogenetically similar bacteria by host plants is not an important driver of bacterial specialization on their host and that reciprocal specialization is therefore unlikely to be a strong driver of host-symbiont pairing across the gradient. The association between bacterial specialization and host traits suggests that niche partitioning and opportunities for specialization do not depend on phylogenetic distances among bacteria (Dolan et al. 2017), but rather on the opportunities for niche partitioning provided by leaf morphological and physiological characteristics. For example, higher leaching of resources on the surface of leaves with higher SLA (and thinner cuticles) could explain the higher propensity for specialization

of their bacteria through resource competition (Lindow and Brandl 2003, Brockhurst et al. 2006).

Variation in the level of specialization of different host species also led to a turnover in mean levels of specialization among sites along the climatic stress gradient. While conifer species and their less specialized microbiota were present throughout the climate gradient, the deciduous species observed at the colder end of the gradient hosted more specialized communities than their southern counterparts, such that we observed a general increase in average specialization with a decrease in temperature across the gradient. While deciduous tree species hosting more specialist bacteria were found at the northern part of our climatic gradient, specialization on sugar maple tended to be lower at these sites where conditions were colder and sugar maple was less regionally abundant. This effect was principally associated with turnover in ASV along the gradient, and to a lesser extent with variation in specialization within ASVs. These observations suggest on the one hand that the mass effects leading to a greater resemblance between microbiota of sugar maple and coniferous species at the northern part of the sugar maple range would also increase the proportion of generalist species coming from the surrounding conifers in its microbiota. On the other, it means that specialization on sugar maple could be harder to maintain in plots where it is not as abundant, such that even individual ASVs would tend to become more generalist.

The variation in specialization across sites that we observed is not entirely consistent with the expanded Stress Gradient Hypothesis (O'Brien et al. 2018). While average specialization of bacteria among deciduous host species follows the expectation that associations between hosts and symbionts should be more specialized in more stressful sites, within a deciduous host species, bacterial specialization appears to be decreasing with stress. A possible explanation would be that bacteria taking advantage of the greater nutrient availability on deciduous trees would benefit more from specializing on deciduous hosts that are consistently present in harsher climates.

However, bacteria specialized on more southern species would not be resistant to climate stress, and thus average specialization would be reduced towards a host's northern range limit where it is less abundant. These results are consistent with observed constraints to sugar maple establishment at the northern portion of their range as a result of a lower availability of their associated mycorrhizal fungi (Brown and Vellend 2014, Carteron et al. 2020), and suggest that such constraints might also exist at the level of the phyllosphere and be affecting the northern expansion of tree species.

Lastly, in contrast with what was observed for community composition, characteristics of the host community at the site-level, namely weighted average host traits and functional diversity did not explain variation in levels of specialization among sites. With average bacterial specialization levels being strongly determined by host species identity, characteristics of the population of that host and its relative abundance in the landscape appear to be more important in determining the extent of specialization of its bacteria than the characteristics of other hosts. Since specialization is higher when the host is more abundant, it seems that the opportunity to encounter a host encourages the maintenance of bacterial specialization. Whether specialization in phyllosphere bacteria is likely to lead to co-diversification with their hosts remains to be investigated. Still, the absence of correlation between average levels of specialization of bacterial communities and the extent of phylogenetic clustering of these communities among hosts suggest it is unlikely to be prevalent in this system.

4.6 Conclusions

Using field data from more than 30 host tree species sampled across a wide latitudinal gradient in eastern North America, we have provided evidence for context-dependence in the associations between plant hosts and their bacteria in the phyllosphere. We show that bacterial community composition on temperate tree leaves was determined both by the abiotic environmental context and the identity of the focal

host, but also by the taxonomic and functional traits of co-occurring hosts across the landscape. The level of specialization of bacteria on their hosts also varied as a function of its regional abundance, suggesting an importance for co-occurring host species in altering the match between host trees and their microbiota. Overall, our results represent a first major examination of the drivers of bacterial composition and specialization to life on diverse tree hosts in natural communities. They also represent a major step in trying to predict potential mismatches between tree host species and their microbiota under environmental change.

4.7 Acknowledgements

We thank the following institutions for access to sampling sites: Société des établissements de plein air du Québec (Parc du Mont-Mégantic, Parc des Monts-Valins, Parc de la Jacques-Cartier, Parc de la Vérendrye), Université du Québec à Montréal (Centre écologique La Huardière), McGill University (Gault Nature Reserve), Commission de la Capitale Nationale (Parc de la Gatineau), Ontario Parks (Frontenac Park) and University of Vermont (Proctor Maple Research Center). We are grateful to Ariane Lafrenière, Dominique Tardif and Geneviève Bourret for their technical assistance in the field and in the lab. We acknowledge funding from the Natural Sciences and Engineering Research Council of Canada (GL, SWK) and the Canada Research Chairs (SWK).

4.8 Data and code availability statement

Data used in this study will be deposited on FigShare upon acceptance of the manuscript. Code used for sequence data processing and ecological analyses is available in GitHub repository: https://github.com/glajoie1/host_neighbourhood.

CONCLUSION

Understanding the diversity of microbial life is crucial to addressing several fundamental questions in ecology. Through the study of microbes, we are starting to evaluate the universality of ecological processes driving species distributions across the tree of life. We are also improving our appreciation of the role for biotic interactions in driving coexistence and the evolution of co-occurring species across trophic levels. With numerous studies linking the composition of human and plant microbial communities to their host health or productivity, applied research in microbial ecology is also having tremendous impacts on our understanding of plant and animal fitness. My thesis research focused on improving our understanding of the structure of microbial diversity at a global scale using trait-based approaches, and of the way microbial diversity is maintained through their association with hosts. I used diverse research methodologies, including a conceptual review, a meta-analysis and a field-based study to address these fundamental questions. Specific contributions of this thesis to the field of ecology, and more specifically microbial ecology, along with caveats and future directions for this research program are outlined below.

5.1 Trait-based approaches in microbial ecology

While the prediction of microbial functions based on gene sequences are being more routinely added to studies of taxonomic turnover among microbial communities, the use of trait-based data continues to fulfill a largely descriptive role in microbial ecology. In the first chapter, we identify recent contributions of trait-based approaches

to knowledge development in microbial ecology and outline specific challenges that are hampering a transition to a more predictive and integrative microbial trait-based ecology. Such challenges include a lack of a working definition of a microbial trait, a large diversity of microbial lifestyles limiting the potential for "universal" methodologies for studying traits, and incomplete and biased databases on microbial traits. In response to these challenges, we first make the case for initiating new discussions on what constitutes a valuable microbial trait to study. We then argue that the identification of major microbial adaptive strategies, combining numerous covarying traits, could facilitate the integration of trait-based results among research teams studying trait variation at different biological levels and with different methodologies. We finally propose grounds for better data and theory sharing between all practitioners of microbial ecology to facilitate integration and generalization of research results.

Altogether, this opinion piece comes at a time when there is a great interest in applying trait-based approaches to microbial ecology, but when the question of how and why to do it is still lagging behind. By anchoring our discussion in a framework derived from philosophy and epistemology, we broaden the scope of trait-based approaches to microbial ecology from one oriented mostly around explanation towards improving the predictive and integrative potential of these approaches. We use case studies from macro-organismal ecology to show concretely how these goals for knowledge development can be fulfilled and propose precise solutions, adapted to the biological reality of microbes, to make the most of recent advancements in the measurement of microbial phenotypes and traits.

Despite such stimulating prospects for microbial trait-based approaches and an increased interest in generating functional predictions from sequence data, important caveats to these approaches still remain to be solved and keep on applying to most genetic trait-based studies. These caveats therefore also apply to Chapter 2. Namely,

the presence of genes encoding for a certain protein in a microbial genome does not necessarily mean that these genes are expressed in the environment from which the sequence was collected, such that we may get at the fundamental niche of microbes, but not the realized niche. The still massive challenge that characterization of the proteome or transcriptome of microbes poses suggests that the description of realized niches may not be widely used for yet another several years. Also, protein-prediction from gene sequences are still biased to those observed from model organisms of narrow taxonomic breadth and economic importance (i.e. human-associated). As such, the functions of environmental microbes remain largely uncharacterized. Assignment of functions to genes also relies on the study of culturable organisms, which has yet excluded a great portion of microbial life (but see Martiny 2019). However, with improvements in the capacity to culture microorganisms regardless of habitat, the description of phenotypes from culture may represent the best way forward in characterizing new microbial traits.

5.2 Fundamental bacterial strategies across clades and ecosystems

The search for the fundamental axes of functional trait variation among living organisms has been a key question in ecology and evolutionary biology for decades (David and Alm 2011, Levine 2015, Díaz et al. 2016). In this article, we use a trait screening approach based on genomic and metagenomic data to identify the key functional strategies of bacteria across the tree of life and across ecosystems. We identify three main axes of correlated functional genes that explain functional turnover both among bacterial clades and habitats, namely 1) DNA metabolism, 2) metabolism of secondary compounds, and 3) signalling and attachment to hosts. Importantly, we show that these strategies are mostly linked with biotic interactions rather than resource use (e.g. carbon compounds), as is usually considered in microbial trait-based approaches (e.g. Krause et al. 2014, Malik et al. 2019). We also show an important role for hosts in explaining bacterial turnover among habitats worldwide. These results

represent a major advance in our understanding of the origin and maintenance of biological diversity, expanding on related findings for plants (Wright et al. 2004) and birds (Pigot et al. 2020).

Overall, our study is the first to use a data-driven approach to identify microbial ecological strategies across both genomes and environmental metagenomes. Our novel approach to studying the structure of bacterial functional diversity allows us to quantify the role of evolutionary processes in structuring microbial ecological differences among ecosystems. By reducing the high dimensionality of trait variation observed among microorganisms around a small number of fundamental axes of trait covariation, we make a significant step towards generalization of the drivers of biological diversity in microbes but also across study systems.

The quality of functional inference possible from metagenomic datasets represents a caveat to most metagenomic studies of environmental microbiota. With current datasets used for microbial protein predictions, we are still only capable of annotating a very low portion of sequences in metagenomic datasets (~3%). As a result, even with relatively high depth of sequencing it remains difficult to confirm the presence of whole functional pathways for most members of the microbial community. In this context of our analyses, we therefore assumed that the presence of a functional gene was evidence for the presence of the functional pathways in which it is known to participate. Still, genes can participate in several pathways and it may not thus be possible to tell which one of these pathways is truly represented in the dataset, such that there might be false positives in our datasets. Still, by setting a minimum of 74,000 functional annotations for inclusion in our study, we expect to have controlled for this issue. In a similar vein, metagenomic and genomic approaches to the identification of adaptive strategies are still limited by the precision of the fitness metric they are using in identifying traits of ecological importance. That is, we had to rely on the relative abundance of a functional gene in the community as an indicator of its fitness

consequences, but this metric may not be perfectly correlated with a gene's impact on growth and survival in individual organisms. More precisely linking traits to function in sequencing studies may require more extensive coupling of sequencing approaches with culture-based methods.

An important motivation for investigating functional strategies was to improve integration and generalization and while our work represents an important first step in doing this, this project could radiate in different ways to contribute to these goals. It would namely be relevant to ask whether these axes of variation are scalable among biological levels, meaning whether they explain variation within ecosystems and bacterial communities, and among bacterial populations. In the third chapter of this thesis, we found that the functions that explained most variation among bacterial phyllosphere samples from a neotropical forest were those linked to signalling and attachment, as well as metabolism of terpenoids and polyketides, which is in line with the functional strategies found to be important among ecosystems as well.

5.3 Adaptive matching between phyllosphere bacteria and their tree host

The phyllosphere is an important microbial habitat but we have a limited understanding of how plant hosts drive the composition of their associated leaf microbial communities and whether taxonomic associations between plants and phyllosphere microbes arises through adaptive matching. In this paper, we asked how plant hosts are shaping community assembly of their phyllosphere bacteria through their role as biotic filters. We first describe a core functional microbiome of the phyllosphere of 17 neotropical tree species. We next uncover a role for trees in constraining the composition of traits of their phyllosphere communities along a gradient of leaf trait properties among host species, with specific microbial adaptations pertaining to the biosynthesis and degradation of secondary compounds appearing to play a role in responding to these host gradients. We further find that several microbial

traits driving functional turnover among communities were conserved in the host phylogeny, strengthening the proposed role for plants as selective forces on microbial community assembly in the phyllosphere.

This study is among the first field-based studies to investigate the structure and drivers of phyllosphere functional diversity across multiple host species using metagenomic shotgun sequencing. Our study considerably improves the mechanistic understanding of plant-microbe interactions by using a trait-based approach to evaluate the nature and extent of adaptive matching between microbes and their tree hosts. It also provides an important first step in unraveling the main adaptive axes of microbial communities in the phyllosphere and in finding specific microbial functions that could be routinely used to describe such communities.

Some questions however remain unanswered regarding the emergence of these trait correlation patterns between hosts and bacteria. Namely, we were not able to determine whether this trait matching arose through adaptation of bacteria to trait variation in their tree hosts, or whether they are the result of reciprocal adaptation between both types of partners potentially leading to coevolutionary dynamics. Evidence for broad cophylogenetic patterns between phyllosphere bacteria and their tree hosts was uncovered in the same study system using samples from 57 tree species, with several bacterial phyla being consistently associated with specific plant families (Kembel et al. 2014). Evaluating evolutionary correlation between host and bacterial traits would represent a valuable step in testing the mechanistic bases of these patterns (e.g. Adams and Nason 2018). It would also be relevant to test in controlled conditions whether trait matches are the result of reciprocal adaptation, leading to individual partners exhibiting higher fitness through these traits in the presence of each other (e.g. Bassar et al. 2017).

It would also be worthwhile to evaluate how the main axes of variation uncovered in phyllosphere bacterial traits in this tropical forest compare with other phyllosphere habitats that may present different abiotic and biotic filters for both microbes and plants. While we provided evidence for a role for the host community in driving bacterial community composition on individual hosts across climatic gradients (Chapter 4), we do not know whether their functional composition is similarly affected. For example, we could expect droughts to represent greater selective pressure on the traits of phyllosphere microbes and their hosts in temperate relative to tropical forests, with consequences for the quality of trait matching between them. Evaluating the generality of the drivers of bacterial community assembly on their host in the phyllosphere remains a standing objective for the field.

5.4 Bacterial community assembly and specialization in a multi-host landscape

Turnover in the composition of microbial communities among host species and genotypes supports a role for hosts in structuring microbial diversity. While such patterns have mostly been explained with host taxonomy, little is understood of the role of neighbouring hosts in influencing the match between a focal host and its microbiome. In this article, we show that tree hosts structure bacterial diversity on their leaves through selection by their leaf traits, but also through horizontal transmission between co-occurring host species. More specifically, we show that the abundance of other tree species in the locality and region of a focal tree population influences the composition of its bacterial microbiota across its range. The relative abundance of the focal host species in the landscape is similarly a good predictor of bacterial specialization on that host, suggesting that the transmission of generalized bacteria from alternative host species may affect host-symbiont matching in the phyllosphere.

All in all, this study is among the first to characterize the influence of co-occurring host species in driving community assembly of plant-associated microbes. It

also represents one of the first characterization and investigation of the drivers of phyllosphere microbial specialization in multi-species host communities. Characterizing variation in microbial specialization is an important first step in evaluating the role of hosts in the evolution of phyllosphere microbes in natural settings. It also provides important insight for predicting host-symbiont mismatches with variation in the distribution of host species as a result of climate change.

One of the main challenges remaining in the study of host-symbiont associations is to determine to which extent partners of co-occurrence translate into interactions. It is possible that several of the microbial occurrences on host leaves represent passenger microbes rather than resident microbes. While we could not confirm this without extensive laboratory manipulations, the consistent association patterns found between host species and their microbiome appear to support the presence of interactions (whether commensalist, mutualist or antagonistic) between hosts and their microbiota in our system. To better understand the pairing mechanisms behind these associations it would also be relevant to obtain data on the functions expressed by microbes on the host leaves, for example through metagenomic sequencing of bacterial samples. Another interesting venture for expanding this project would be to assess whether specialization of bacteria on different host species across the landscape is associated with co-diversification with their host. Linking trait evolution in hosts and diversification of their bacteria would be an important next step in understanding the role of hosts in structuring bacterial diversity in the wild.

5.5 Final remarks

In this thesis, I have argued that the use of trait-based approaches represents a promising way forward in investigating mechanisms of adaptation of microbes to environmental gradients, but also in attempting to build a more integrative practice of microbial ecology. I namely showed that the investigation of major bacterial strategies

represented an informative tool for drawing more general understanding of the processes at play in structuring the composition and evolution of bacterial communities worldwide. I also provided empirical evidence of the role of individual plant hosts and their neighbourhoods in structuring bacterial diversity across spatial scales. Overall, this thesis supports adaptation to life with hosts as an important axis of ecological variation in bacteria, both globally and within a forest ecosystem. It also lays foundations for an improved use of bacterial and host traits in understanding the origin and maintenance of bacterial diversity.

ANNEX A

SUPPLEMENTARY FIGURES – CHAPTER II

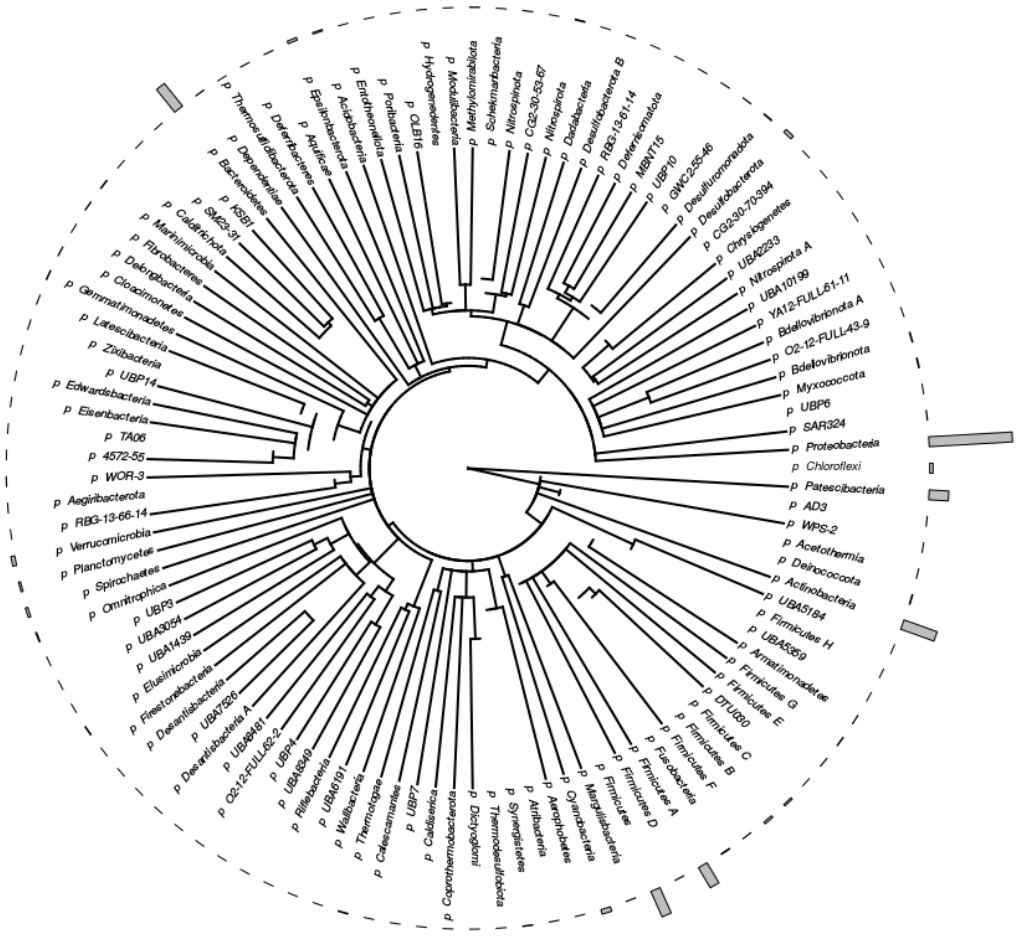

Figure S2.1. Taxonomic composition of the genomic dataset at the phylum level. Bar height indicates the relative abundance of each phylum.

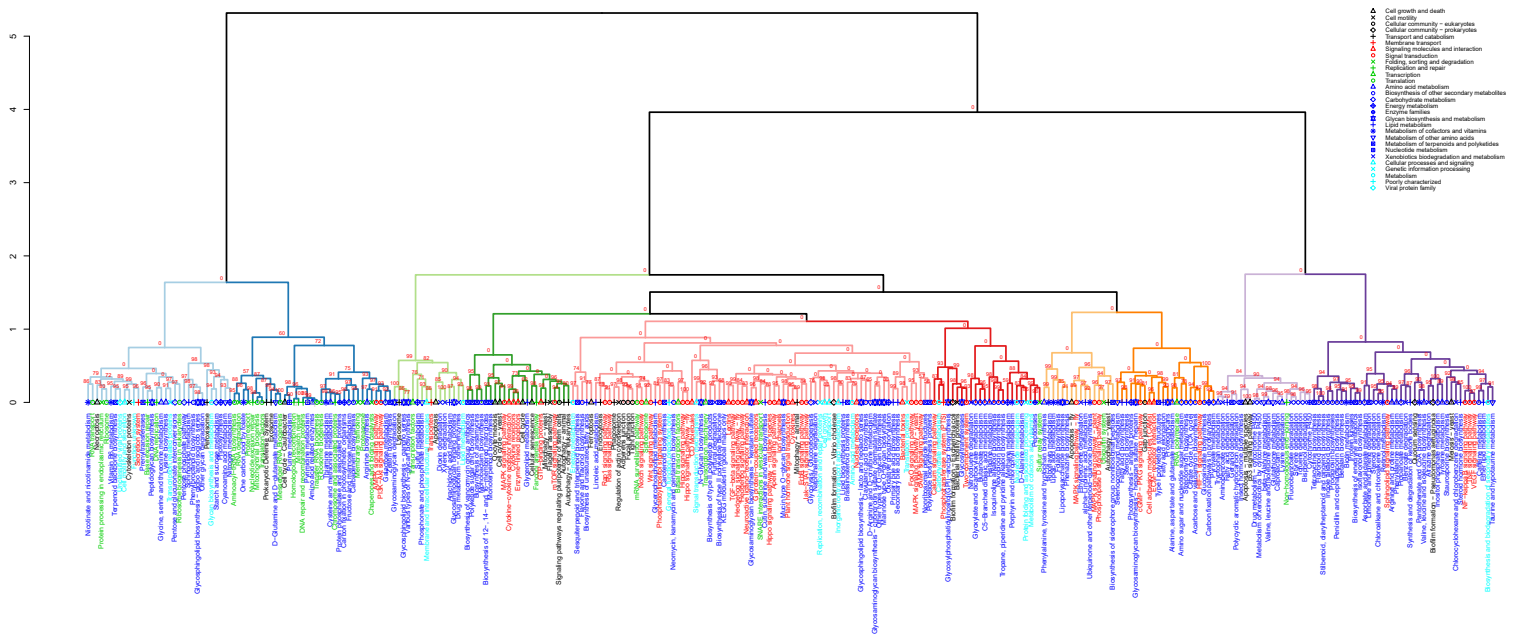

Figure S2.2. Detailed dendrogram indicating correlations among bacterial functional pathways across 69 metagenomic samples. Main clusters of correlated traits are color-coded and labelled (C1-C10) in the dendrogram. Colors and shapes of leaves indicate the KEGG functional category to which the functional pathways belong. Red numbers at each node represent the Approximately Unbiased p-value obtained from multiscale bootstrap resampling of the dendrogram (n=10 000).

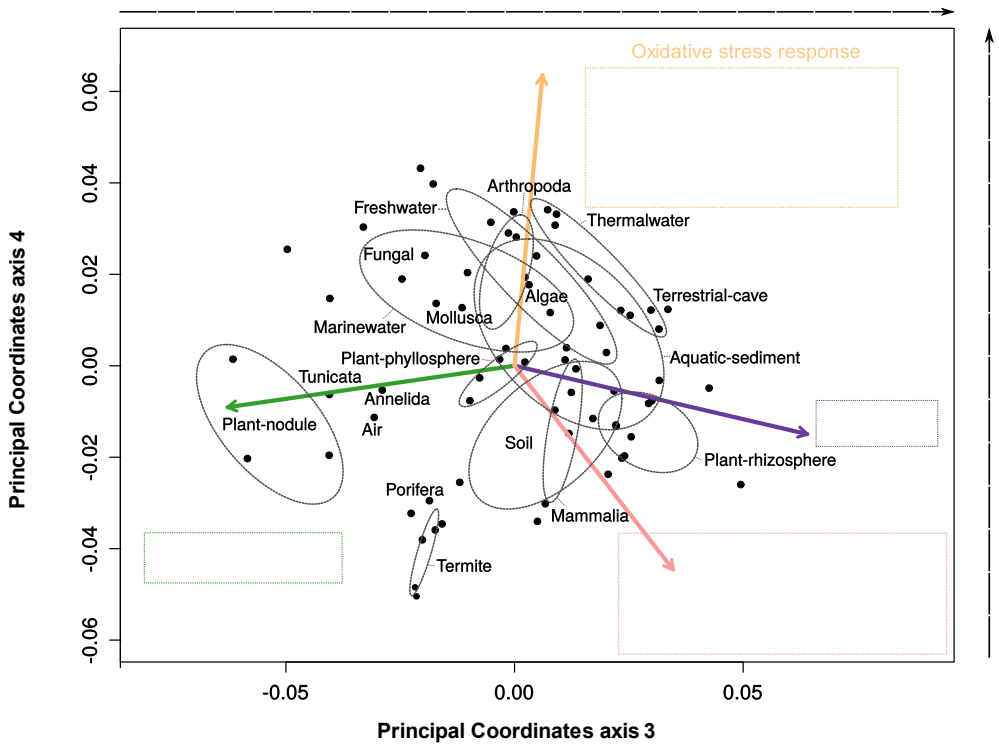

Figure S2.3. Principal coordinates analyses of the metagenomic functional dataset. Axes 3 and 4 are shown. Colored arrows represent the mean position of the traits contributing most to variance across these dimensions by functional cluster (as depicted in Fig. 2.1). These most important traits are indicated in colored boxes next to the corresponding arrow. Bacterial phyla that correlate the most with the axes are indicated on the outer portion of the graph, along with the direction of the correlation. Ellipses define the average position of the points in each environmental group.

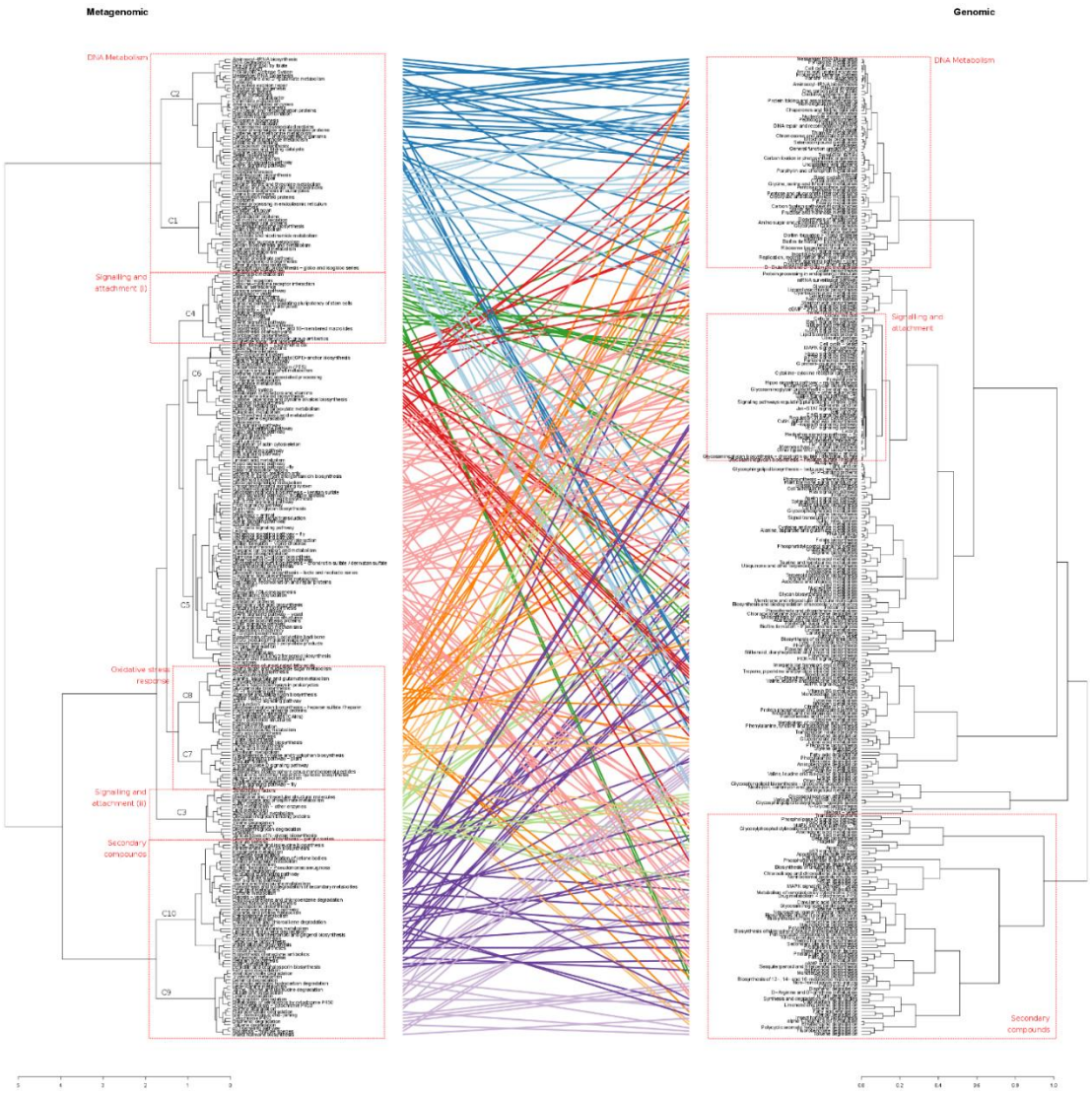

Figure S2.4. Detailed tanglegram comparing functional clusters for bacteria based on annotation of metagenomic (left) and genomic (right) datasets. Functional pathways are color coded by metagenomic clusters (see details in Fig. 2.1). Lines connect the same functional pathways in each data set. Red boxes indicate the main functional strategies identified for bacterial genomes (see details in Fig. 2.1, Fig. 2.2).

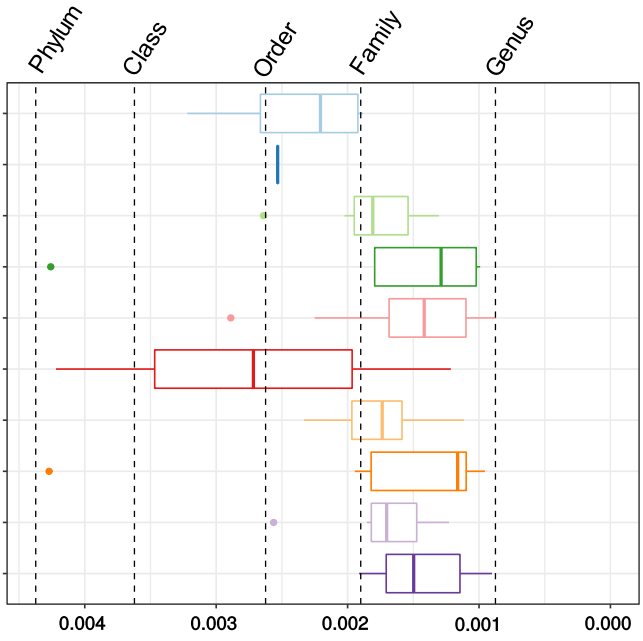

Figure S2.5. Mean phylogenetic depth at which bacterial functional traits are conserved across clades, for each bacterial functional trait cluster identified from metagenomic data (Fig. S2.2). Only traits that had non-random mean depths (i.e. with an observed mean depth as high or higher than 95% of those obtained through the null model) are shown. The median depth at which each taxonomic level varies in this dataset is represented by the dashed black lines.

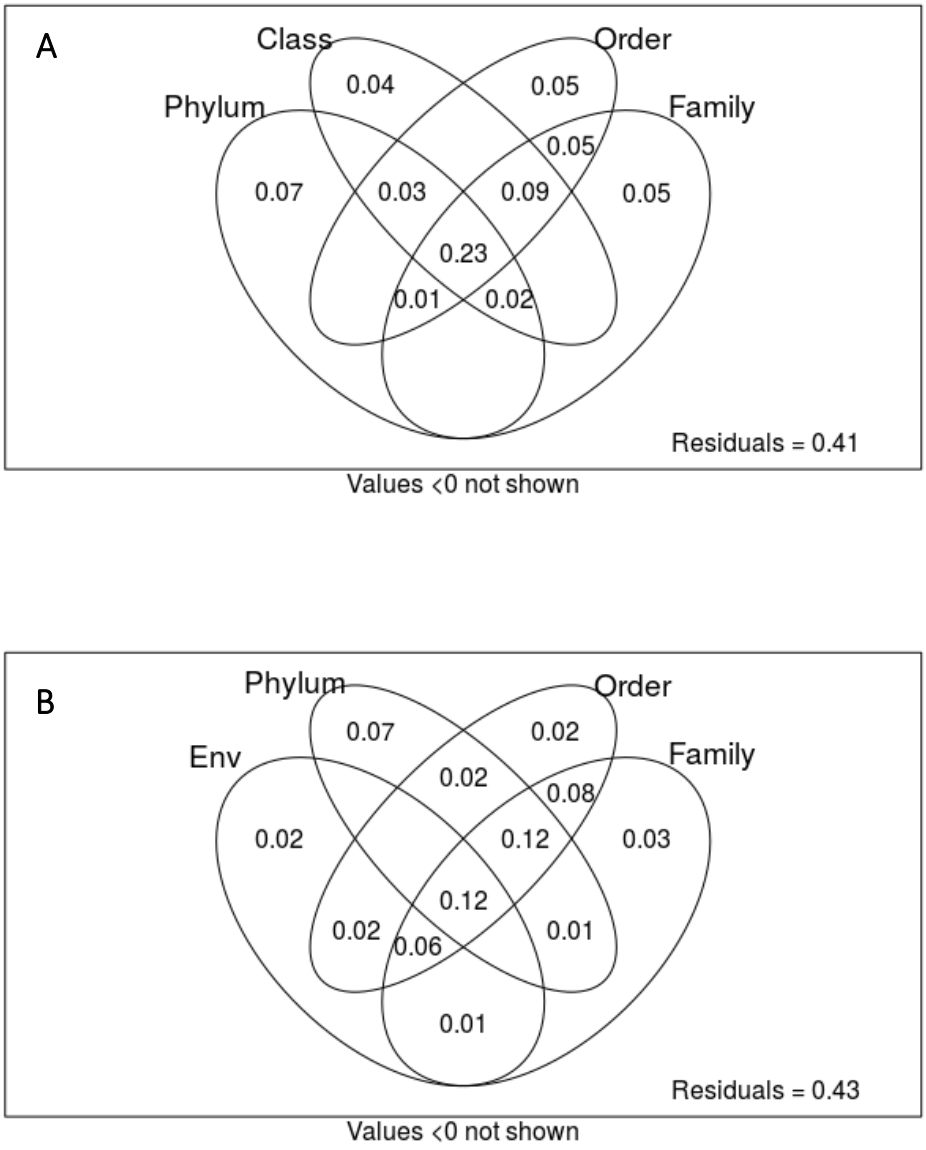

Figure S2.6. Variation partitioning of the functional composition of metagenomic samples, as explained by taxonomic levels (A), and taxonomic levels in conjunction with environmental variables (B). Values indicate the proportion of total variation in functional composition explained by each factor.

ANNEX B

SUPPLEMENTARY TABLES – CHAPTER II

Table S2.1. Metagenomics datasets used in this study.

3300000097	Passalidae beetle gut microbial communities from Costa Rica - Adult (4MA+4BA+4MSA)	Gs0050939	Host-associated	Arthropoda	Digestive system	Arthropoda	Costa Rica: Quebrada Gonzales Sector, Braulio Carrillo National Park
3300000110	Microbial communities from multiple species of Shipworm: Sample from Bankia setacea gill BSg2	Gs0063438	Host-associated	Mollusca	Respiratory system	Mollusca	USA: Puget Sound, Washington
3300000333	Honey bee gut microbial communities from New Haven, Connecticut, USA - Honey Bee colony	Gs0067856	Host-associated	Arthropoda	Digestive system	Arthropoda	USA: New Haven, Connecticut
3300000385	Marine microbial community from Cabo Rojo, Puerto Rico - PR CR 10% Liquid 1	Gs0053056	Environmental	Aquatic	Marine	Marinewater	Puerto Rico: Cabo Rojo
3300000401	Marine microbial community from La Parguera, Puerto Rico - BB Mangrove B Liquid	Gs0053056	Environmental	Aquatic	Marine	Marinewater	Puerto Rico: Bioluminescent Bay, La Paraguera
3300000488	Salt pond sediment microbial community from Fremont, CA, USA - Pond A23 Sediment 3	Gs0053056	Environmental	Aquatic	Non-marine saline and alkaline	Saltwater-sediment	USA: Alviso Ponds, San Francisco, California
3300000734	Tropical forest soil microbial communities from Luquillo Experimental Forest, Puerto Rico - Sample 81	Gs0075432	Environmental	Terrestrial	Soil	Soil	Puerto Rico: Luquillo Experimental Forest Soil
3300000793	Forest soil microbial communities from Amazon forest - 2010 replicate II A001	Gs0067860	Environmental	Terrestrial	Soil	Soil	Brazil: Amazon forest, Fazenda Nova Vida, City of Ariquemes, State of Rondonia
3300001384	Arctic peat soil from Barrow, Alaska - NGEE Surface sample 53-3 shallow-072012	Gs0084162	Environmental	Terrestrial	Soil	Soil	USA: Barrow Environmental Observatory site, Alaska
3300001401	Arctic peat soil from Barrow, Alaska - NGEE Surface sample 53-3 deep-072012	Gs0084162	Environmental	Terrestrial	Soil	Soil	USA: Barrow Environmental Observatory site, Alaska
3300001539	Ecteinascidia turbinata endosymbiont from Florida, USA - Sample 2	Gs0090291	Host-associated	Tunicata	Endophyte	Tunicata	USA: Key West, Florida

3300002125	Cubitermes P4 segment microbial communities from Max Planck Institute, Germany - Cu122P4	Gs0084161	Host-associated	Arthropoda	Digestive system	Termite	Kenya: Kakamega
3300002147	Host-associated microbial community of the marine sponge Aplysina aerophoba from Gulf of Piran - sponge mesohyl, lysed by freeze-thaw cycling	Gs0099546	Host-associated	Porifera	Endophyte	Porifera	Italy: Gulf of Piran, Adriatic Sea
3300002308	Nasutitermes corniger P4 segment microbial communities from Max Planck Institute, Germany - Nc150P4	Gs0084161	Host-associated	Arthropoda	Digestive system	Termite	USA: Davie, Florida
3300002405	Earthworm egg capsule microbial community from the University of Washington, USA - E. fetida Yelm	Gs0060821	Host-associated	Annelida	Reproductive system	Annelida	USA: University of Washington
3300002504	Termite gut P4 segment microbial communities from Max Planck Institute, Germany - Nt197	Gs0084161	Host-associated	Arthropoda	Digestive system	Termite	Germany: Marburg
3300003170	Upper troposphere microbial communities - SDPR-005	Gs0110167	Environmental	Air	Outdoor Air	Air	USA
3300003203	Tabebuia heterophylla rhizosphere microbial communities from the University of Puerto Rico - S4T2R2	Gs0103004	Host-associated	Plants	Rhizosphere	Plant-rhizosphere	Puerto Rico, University of Puerto Rico, San Juan
3300005453	Anoxygenic and chlorotrophic microbial mat microbial communities from Yellowstone National Park, USA - YNP MS-T MetaG	Gs0046783	Environmental	Aquatic	Thermal springs	Thermalwater	USA: Wyoming, Yellowstone National Park
3300005578	Corn rhizosphere microbial communities from Kellogg Biological Station, Michigan, USA - KBS Corn C4-2	Gs0090294	Host-associated	Plants	Rhizosphere	Plant-rhizosphere	USA: Michigan, Kellogg Biological Station
3300006046	Grasslands soil microbial communities from the Angelo Coastal Reserve, California, USA - Sample Angelo_101	Gs0110119	Environmental	Terrestrial	Soil	Soil	USA: California, Angelo Coastal Reserve
3300006353	Populus root and rhizosphere microbial communities from Tennessee, USA - Endosphere MetaG P. TD hybrid TD303-5	Gs0103573	Host-associated	Plants	Rhizosphere	Plant-rhizosphere	USA: Tennessee

3300006619	Arctic peat soil microbial communities from the Barrow Environmental Observatory site, Barrow, Alaska, USA - NGEE Permafrost159B-4B	Gs0084162	Environmental	Terrestrial	Soil	Soil	USA: Barrow Environmental Observatory site, Alaska
3300006893	Iron sulfur acid spring bacterial and archeal communities from Banff, Canada, to study Microbial Dark Matter (Phase II) - Paint Pots PPA 5.5 metaG	Gs0111485	Environmental	Terrestrial	Soil	Soil	Canada: Banff
3300006894	Agricultural soil microbial communities from Utah to study Nitrogen management - NC Control	Gs0114436	Environmental	Terrestrial	Soil	Soil	USA: Utah
3300006903	Populus root and rhizosphere microbial communities from Tennessee, USA - Soil MetaG P. TD hybrid SBSTD5	Gs0103573	Environmental	Terrestrial	Soil	Plant-rhizosphere	USA: Tennessee
3300006941	Root nodule microbial communities of legume samples collected from California, USA - Siratro red BW	Gs0114676	Host-associated	Plants	Nodule	Plant-nodule	USA: California
3300006944	Root nodule microbial communities of legume samples collected from California, USA - Cow pea red BW	Gs0114676	Host-associated	Plants	Nodule	Plant-nodule	USA: California
3300009102	Deep subsurface microbial communities from Mariana Trench to uncover new lineages of life (NeLLi) - CR04 metaG	Gs0118434	Environmental	Aquatic	Marine	Marinewater	Mariana Trench
3300009446	Marine algal microbial communities from Maine, USA - Maine_Asex2 metaG	Gs0019863	Host-associated	Algae	Epiphyte	Algae	USA: Maine
3300009488	Deep subsurface microbial communities from Indian Ocean to uncover new lineages of life (NeLLi) - Sumatra_00607 metaG	Gs0118434	Environmental	Aquatic	Marine	Marinewater	North Sumatra
3300009507	Pelagic marine microbial communities from North Sea - COGITO_mtgs_120607	Gs0084160	Environmental	Aquatic	Marine	Marinewater	Atlantic Ocean: North Sea, Helgoland
3300009510	Host-associated microbial communities from peat moss isolated from Minnesota, USA - S1T2_Fd - Sphagnum fallax MG	Gs0118677	Host-associated	Bryophyta	Whole-organism	Bryophyta	USA: Minnesota
3300009594	Groundwater microbial communities from Devils Hole, Nevada to study Microbial Dark Matter (Phase II) - Devils Hole	Gs0111485	Environmental	Aquatic	Freshwater	Freshwater	Nevada: Devil's Hole

3300009697	Host-associated microbial communities from peat moss isolated from Minnesota, USA - S1T2_Fd - Sphagnum magellanicum MG	Gs0118677	Host-associated	Bryophyta	Whole-organism	Bryophyta	USA: Minnesota
3300009765	Root nodule microbial communities of legume samples collected from Mexico - Turtle bean Mexico pink nodule	Gs0114676	Host-associated	Plants	Nodule	Plant-nodule	Mexico: Nepantla
3300009774	Glacier valley bacterial and archeal communities from Borup Fiord, Nunavut, Canada, to study Microbial Dark Matter (Phase II) - lysozymeSSSS metaG	Gs0111485	Environmental	Aquatic	Freshwater	Freshwater	Canada: Borup Fiord, Nunavut
3300010262	Eastern black-and-white colobus group fecal microbial communities from Wisconsin, USA - Cm1105 metagenome	Gs0120398	Host-associated	Mammalia	Digestive system	Mammalia	USA: Wisconsin
3300010264	Marine hydrothermal vent microbial communities from Guaymas Basin, Gulf of California to study Microbial Dark Matter (Phase II) - Marker 14 Mat core 4571-4 33-36 cm metaG	Gs0111485	Environmental	Aquatic	Marine	Marinewater	Mexico: Guaymas Basin, Gulf of California
3300010278	Western lowland gorrila individual fecal microbial communities from Wisconsin, USA - Go1022B metagenome	Gs0120398	Host-associated	Mammalia	Digestive system	Mammalia	USA: Wisconsin
3300010279	Orangutan group fecal microbial communities from fecal samples from Wisconsin, USA - O1105 metagenome	Gs0120398	Host-associated	Mammalia	Digestive system	Mammalia	USA: Wisconsin
3300010282	Capybara group fecal microbial communities from Wisconsin, USA - P1105 metagenome	Gs0120398	Host-associated	Mammalia	Digestive system	Mammalia	USA: Wisconsin
3300010284	Hot spring microbial mat communities from California, USA to study Microbial Dark Matter (Phase II) - Cone Pool mat layer H metaG	Gs0111485	Environmental	Aquatic	Thermal springs	Thermalwater	USA: California
3300010313	Hot spring microbial communities from South Africa to study Microbial Dark Matter (Phase II) - Sagole hot spring metaG	Gs0111485	Environmental	Aquatic	Thermal springs	Thermalwater	South Africa: Limpopo
3300010324	Lake sediment bacterial and archeal communities from Gulf of Boni, Indonesia to study Microbial Dark Matter (Phase II) - ?I18A1 metaG	Gs0111485	Environmental	Aquatic	Freshwater	Freshwater-sediment	Indonesia: Gulf of Boni

3300010343	Bog forest soil microbial communities from Calvert Island, British Columbia, Canada - Bog Forest MetaG ECP23OM1	Gs0110174	Environmental	Terrestrial	Soil	Soil	Canada: Calvert Island, British Columbia
3300011013	Deep subsurface microbial communities from Kolumbo volcano to uncover new lineages of life (NeLLi) - 4SBTROV10_white metaG	Gs0118434	Environmental	Aquatic	Marine	Marinewater	Aegean Sea: Kolumbo volcano
3300012264	Freshwater sediment bacterial and archeal communities from Indian Creek, Illinois, USA to study Microbial Dark Matter (Phase II) - Sed-PBS metaG	Gs0111485	Environmental	Aquatic	Freshwater	Freshwater-sediment	USA: Indian Creek, Illinois
3300012266	Freshwater bacterial and archeal communities from Indian Creek, Illinois, USA to study Microbial Dark Matter (Phase II) - JTO19cm metaG	Gs0111485	Environmental	Aquatic	Freshwater	Freshwater	USA: Indian Creek, Illinois
3300012809	Enriched millipede-associated microbial communities from UW Madison campus, WI, USA - HID1971M_E11 MG	Gs0121620	Host-associated	Arthropoda	Digestive system	Arthropoda	USA: Madison, Wisconsin
3300012824	Enriched pill bug-associated microbial communities from UW Madison campus, WI, USA - HID1972M_E11 MG	Gs0121620	Host-associated	Arthropoda	Digestive system	Arthropoda	USA: Madison, Wisconsin
3300012830	Enriched soil microbial communities from UW Madison campus, WI, USA - DID2934_E24_Xylan MG	Gs0121620	Environmental	Terrestrial	Soil	Soil	USA: Madison, Wisconsin
3300012839	Enriched mosquito-associated microbial communities from UW Madison campus, WI, USA - HID1973M_E11 MG	Gs0121620	Host-associated	Arthropoda	Digestive system	Arthropoda	USA: Madison, Wisconsin
3300012995	Fungus gardens microbial communities from leaf cutter ant in Ribeir?o Preto, State of S?o Paulo, Brazil - Atta laevigata ALBM2	Gs0121620	Host-associated	Fungi	Mycelium	Fungal	Brazil: Ribeiro Preto, State of Sao Paulo
3300012996	Fungus gardens microbial communities from leaf cutter ant in Botucatu, State of S?o Paulo, Brazil - Atta capiguara ACBM2	Gs0121620	Host-associated	Fungi	Mycelium	Fungal	Brazil: Botucatu, State of Sao Paulo
3300013091	Freshwater microbial communities from Lake Kivu, Western Province, Rwanda to study Microbial Dark Matter (Phase II) - Kivu_220m	Gs0111485	Environmental	Aquatic	Freshwater	Freshwater	Rwanda: Western Province

3300013297	Miscanthus rhizosphere microbial communities from Kellogg Biological Station, Michigan, USA - M6-5 metaG	Gs0090294	Host-associated	Plants	Rhizosphere	Plant-rhizosphere	USA: Michigan
3300014325	Switchgrass rhizosphere microbial communities from Kellogg Biological Station, Michigan, USA - S6-5 metaG	Gs0090294	Host-associated	Plants	Rhizosphere	Plant-rhizosphere	USA: Michigan
3300014488	Bulk soil microbial communities from Mexico - San Felipe (SF) metaG	Gs0053055	Environmental	Terrestrial	Soil	Soil	Mexico: San Luis Potosi
3300017788	Freshwater microbial communities from Lake Kivu, Western Province, Rwanda to study Microbial Dark Matter (Phase II) - Kivu_15m_20L	Gs0111485	Environmental	Aquatic	Freshwater	Freshwater	Rwanda: Western Province
3300019360	White microbial mat communities from a lava cave in the Kipuka Kanohina Cave System on the Island of Hawaii, USA - GBC170108-1 metaG	Gs0118434	Environmental	Terrestrial	Cave	Terrestrial-cave	USA: the Island of Hawaii
3300019458	Bio-ooze microbial communities from a basaltic lava cave in the Kipuka Kanohina Cave System on the Island of Hawaii, USA - MA170107-3 metaG	Gs0118434	Environmental	Terrestrial	Cave	Terrestrial-cave	USA: the Island of Hawaii
3300020185	Pelagic subsurface seawater microbial communities from Kabeltonne, Helgoland, North Sea - Helgoland_Spring_Bloom_20160517_1	Gs0084160	Environmental	Aquatic	Marine	Marinewater	Atlantic Ocean: North Sea, Helgoland
3300024038	Enriched microbial communities from leaf-cutter ant dump, University of Wisconsin, Madison, United States - 3A200A	Gs0121620	Host-associated	Arthropoda	Ant dump	Arthropoda	USA: Wisconsin
3300025100	Hot spring sediment microbial communities from Zodletone spring, Oklahoma to study Microbial Dark Matter (Phase II) - Zodletone Spring source 0.5m metaG (SPAdes)	Gs0111485	Environmental	Aquatic	Hot (42-90C)	Thermalwater-sediment	USA: Oklahoma, Zodletone Spring
3300028599	Marine sediment microbial communities from subtidal zone of North Sea - Hel_20160524 (Illumina Assembly)	Gs0084160	Environmental	Aquatic	Marine	Marinewater-sediment	Atlantic Ocean: North Sea, Helgoland

3300028887	Bovine rumen microbial communities from tropical cattle in Woodstock, Queensland, Australia - Gonzalo_02	Gs0133408	Host-associated	Mammalia	Digestive system	Mammalia	Australia: Woodstock, Queensland
3300028888	Sheep rumen microbial communities from Palmerston North, Manawatu-Wanganui, New Zealand - 1728 DNA GHGlow gp2	Gs0133408	Host-associated	Mammalia	Digestive system	Mammalia	New Zealand: Palmerston North, Manawatu-Wanganui
3300030501	Agave microbial communities from Guanajuato, Mexico - Mg.Sf.e (v2)	Gs0053055	Host-associated	Plants	Phylloplane	Plant-phyllosphere	Mexico: Guanajuato

Table S2.2. Predictions of the ecological and evolutionary significance of groups of functional pathways based on their clustering in the metagenomic and the genomic datasets.

		Genomic dataset	
	Group of traits is :	Clustered	Not clustered, or clustered but unsupported
Metagenomic dataset	Clustered	The cluster is a conserved strategy of current ecological importance. (There is selection and/or constraints on the evolution of this group of traits.)	The cluster represents a strategy selected in the species pool, regardless of phylogenetic identity.
	Not clustered, or clustered but unsupported	The cluster may have been selected for under previous ecological constraints that are no longer important in driving the distribution of organisms among ecosystems, or that explain ecological variation at another scale.	These traits do not participate in a general strategy. They might be labile traits that do not have strong constraints of evolution. They might be important for occupying rare niches.

ANNEX C

SUPPLEMENTARY FIGURES – CHAPTER III

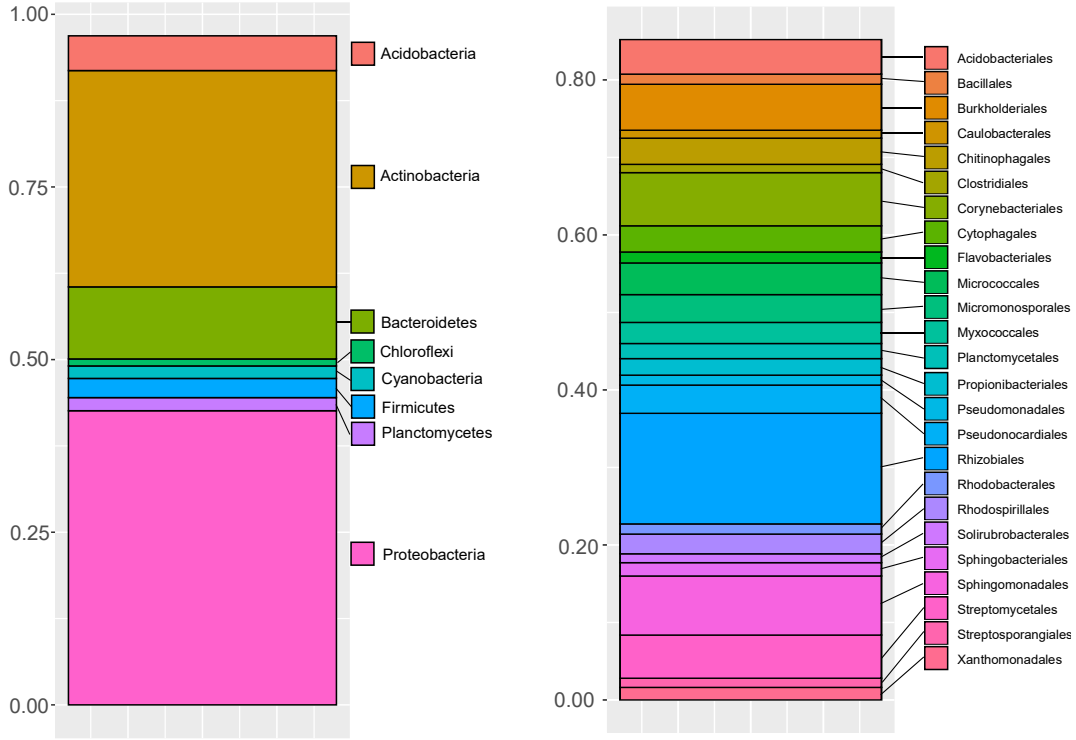

Figure S3.1. Relative abundance of phyla (A) and orders (B) across 24 leaf bacterial communities in a neotropical forest. Only the taxa that made up more than 1% of the total abundance of bacteria are indicated in the barchart.

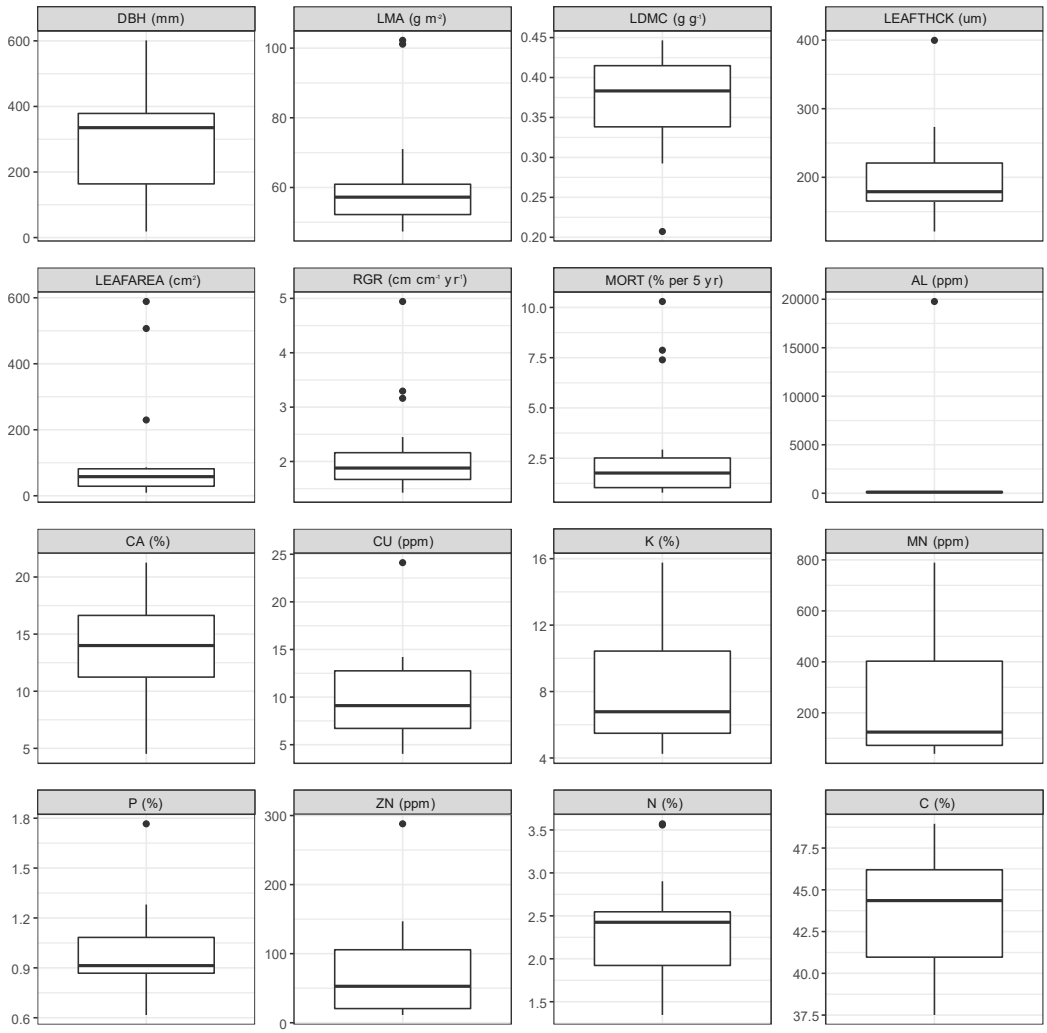

Figure S3.2. Distribution of trait values for 16 traits across 17 tree species from a neotropical forest. See caption of Fig. 2 for a description of trait abbreviations.

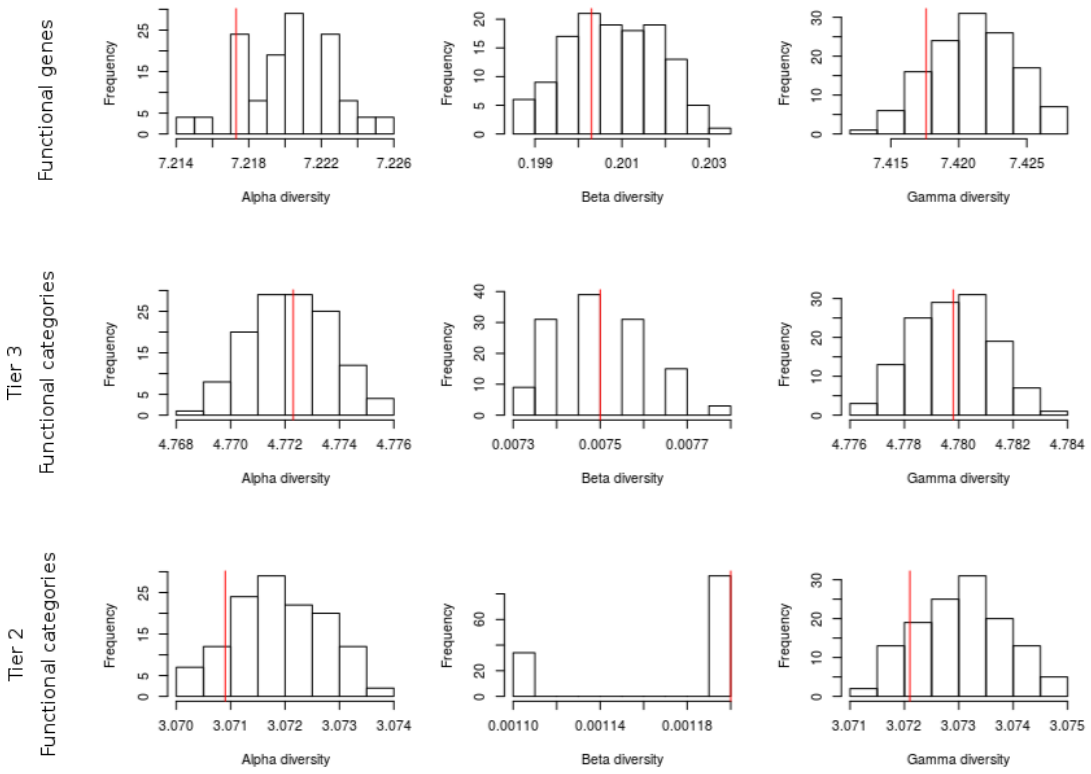

Figure S3.3. Distribution of alpha, beta and gamma diversities generated from 128 subsampling of the metagenomic functional dataset to include only one sample per tree species. Despite variation observed among bootstraps, the relative importance of alpha vs. gamma diversity stayed constant at 97.3% alpha diversity and 2.7% beta-diversity for all subsamples. The red vertical line indicates the observed value.

ANNEX D

SUPPLEMENTARY TABLES – CHAPTER III

Table S3.1. Tier 3 functional categories that are perfectly correlated across samples. The category on the left-hand column was kept for all analyses, while the correlated categories in the right-hand column were discarded.

Functional category in dataset	Correlated functional category removed from dataset
Adherens junction	Focal adhesion Hippo signaling pathway -fly Hippo signaling pathway Phagosome Rap1 signaling pathway Regulation of actin cytoskeleton Tight junction
Biosynthesis of 12-, 14- and 16-membered macrolides	Type I polyketide structures
Biosynthesis of type II polyketide backbone	Tetracycline biosynthesis
Endocytosis	Ras signaling pathway
Glycosphingolipid biosynthesis - ganglio series	Various types of N-glycan biosynthesis
NF-kappa B signaling pathway	TNF signaling pathway VEGF signaling pathway
Notch signaling pathway	Wnt signaling pathway

Table S3.2. The 25 Tier 3 functions contributing the most to variation among samples. Columns correspond to functional categories defined by the Kegg hierarchy (see Methods).

Tier 1 Functional category	Tier 2 Functional category	Tier 3 Functional category
Cellular Processes	Cell growth and death	Apoptosis
Cellular Processes	Cellular community - eukaryotes	Adherens junction
Cellular Processes	Transport and catabolism	Exosome
Environmental Information Processing	Signal transduction	cGMP - PKG signaling pathway
Environmental Information Processing	Signal transduction	MAPK signaling pathway - fly
Environmental Information Processing	Signal transduction	Two-component system
Genetic Information Processing	Replication and repair	DNA repair and recombination proteins
Genetic Information Processing	Transcription	Transcription factors
Metabolism	Amino acid metabolism	Phenylalanine, tyrosine and tryptophan biosynthesis
Metabolism	Amino acid metabolism	Valine, leucine and isoleucine biosynthesis
Metabolism	Biosynthesis of other secondary metabolites	Betalain biosynthesis
Metabolism	Biosynthesis of other secondary metabolites	Indole alkaloid biosynthesis
Metabolism	Biosynthesis of other secondary metabolites	Isoquinoline alkaloid biosynthesis
Metabolism	Biosynthesis of other secondary metabolites	Novobiocin biosynthesis
Metabolism	Biosynthesis of other secondary metabolites	Phenazine biosynthesis
Metabolism	Biosynthesis of other secondary metabolites	Phenylpropanoid biosynthesis
Metabolism	Biosynthesis of other secondary metabolites	Tropane, piperidine and pyridine alkaloid biosynthesis
Metabolism	Carbohydrate metabolism	C5-Branched dibasic acid metabolism
Metabolism	Carbohydrate metabolism	Citrate cycle (TCA cycle)
Metabolism	Carbohydrate metabolism	Starch and sucrose metabolism

Metabolism	Energy metabolism	Oxidative phosphorylation
Metabolism	Energy metabolism	Photosynthesis
Metabolism	Energy metabolism	Sulfur metabolism
Metabolism	Metabolism of cofactors and vitamins	Pantothenate and CoA biosynthesis
Metabolism	Metabolism of terpenoids and polyketides	Zeatin biosynthesis
Unclassified	Genetic information processing	Replication, recombination and repair proteins
Unclassified	Viral protein family	Unclassified viral proteins

ANNEX E

SUPPLEMENTARY FIGURES – CHAPTER IV

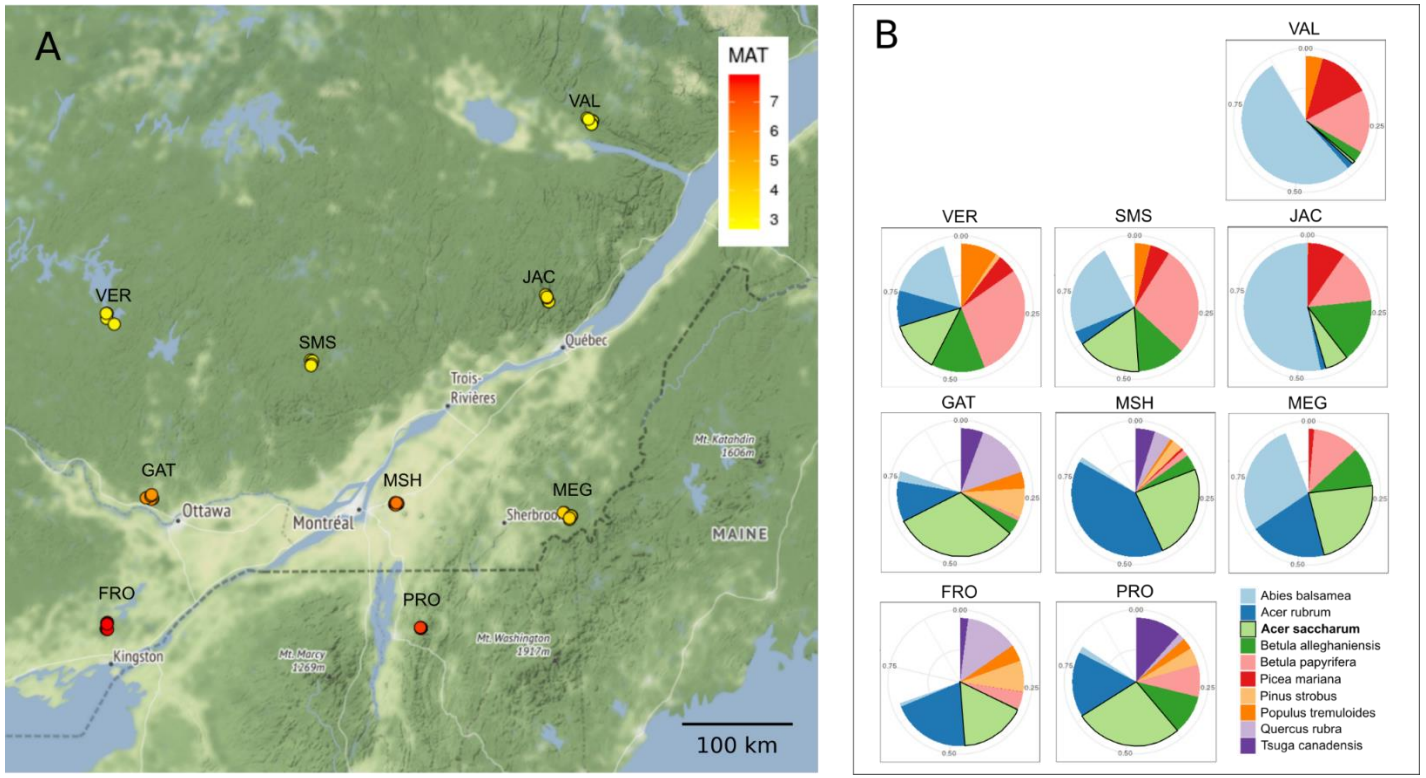

Figure S4.1. Geographical position (panel a) and tree composition (panel a) of sampling sites across North-Eastern America. Mean annual temperatures at each site is color-coded using a heatmap. Abbreviations for sampling sites are presented in Table 4.1. Tree composition was evaluated in a 25 km radius from the center of the site using governmental tree inventories from field surveys and aerial photography. Relative abundances of the ten most abundant tree species across the study area are shown for each site.

ANNEX F

SUPPLEMENTARY TABLES – CHAPTER IV

Table S4.1. Species names and taxonomic classification of tree host species sampled in this study.

Species name	Genus	Family	Order	Abbreviation
Abies balsamea (L.)	Abies	Pinaceae	Pinales	ABIBAL
Acer pensylvanicum (L.)	Acer	Sapindaceae	Sapindales	ACEPEN
Acer rubrum (L.)	Acer	Sapindaceae	Sapindales	ACERUB
Acer saccharum (Marshall)	Acer	Sapindaceae	Sapindales	ACESAC
Acer spicatum (Lam.)	Acer	Sapindaceae	Sapindales	ACESPI
Alnus incana subsp. *rugosa* (Du Roi) R.T.Clausen	Alnus	Betulaceae	Fagales	ALNRUG
Amelanchier laevis (Wiegand)	Amelanchier	Rosaceae	Rosales	AMELAE
Amelanchier stolonifera (Wiegand)	Amelanchier	Rosaceae	Rosales	AMESTO
Betula alleghaniensis (Britt.)	Betula	Betulaceae	Fagales	BETALL
Betula papyrifera (Marshall)	Betula	Betulaceae	Fagales	BETPAP
Carpinus caroliniana (Walter)	Carpinus	Betulaceae	Fagales	CARCAR
Carya cordiformis (Wangenh.) K. Koch	Carya	Juglandaceae	Fagales	CARCOR
Carya ovata (Mill.) K. Koch	Carya	Juglandaceae	Fagales	CAROVA
Corylus cornuta (Marshall)	Corylus	Betulaceae	Fagales	CORCOR
Fagus grandifolia (Ehrh.)	Fagus	Fagaceae	Fagales	FAGGRA
Fraxinus americana (L.)	Fraxinus	Oleaceae	Lamiales	FRAAME
Fraxinus nigra (Marshall)	Fraxinus	Oleaceae	Lamiales	FRANIG
Fraxinus pennsylvanica (Marshall)	Fraxinus	Oleaceae	Lamiales	FRAPEN
Ostrya virginiana (Mill.) K. Koch	Ostrya	Betulaceae	Fagales	OSTVIR
Picea glauca (Moench) Voss	Picea	Pinaceae	Pinales	PICGLA
Picea mariana (Mill.) Britton, Sterns & Poggenburg	Picea	Pinaceae	Pinales	PICMAR
Picea rubens (Sarg.)	Picea	Pinaceae	Pinales	PICRUB
Pinus strobus (L.)	Pinus	Pinaceae	Pinales	PINSTR

Populus grandidentata (Michx.)	Populus	Salicaceae	Malpighiales	POPGRA
Populus tremuloides (Michx.)	Populus	Salicaceae	Malpighiales	POPTRE
Prunus pensylvanica (L.f.)	Prunus	Rosaceae	Rosales	PRUPEN
Quercus alba (L.)	Quercus	Fagaceae	Fagales	QUEALB
Quercus rubra (L.)	Quercus	Fagaceae	Fagales	QUERUB
Sorbus americana (Marshall)	Sorbus	Rosaceae	Rosales	SORAME
Thuja occidentalis (L.)	Thuya	Cupressaceae	Pinales	THUOCC
Tilia americana (L.)	Tilia	Malvaceae	Malvales	TILAME
Tsuga canadensis (L.) Carrière	Tsuga	Pinaceae	Pinales	TSUCAN
Ulmus americana (L.)	Ulmus	Ulmaceae	Rosales	ULMAME

Table S4.2. Variation in mean specialization of phyllosphere bacterial communities on their tree host as a function of individual (a) and site-weighted mean (b) host traits. Results from linear regression models are presented for each trait : specific leaf area (SLA), leaf calcium concentration (Ca), leaf phosphorus concentration (P) and wood density. β represents the standardized regression coefficient. Significance is indicated as the following: *$p < .05$, **$p < .01$, ***$p < .001$.

a. Individual host traits

	β	R^2	p	
SLA	0.066	0.14	0.001	***
Ca	0.051	0.08	0.014	*
P	0.048	0.08	0.019	*
Wood density	0.025	0.02	0.230	

b. Site-weighted mean host traits

	β	R^2	p	
SLA	-0.021	0.01	0.318	
Ca	-0.005	0.00	0.803	
P	-0.034	0.04	0.108	
Wood density	-0.043	0.06	0.040	*

REFERENCES

Abby, S. S., E. Tannier, M. Gouy, and V. Daubin. 2012. Lateral gene transfer as a support for the tree of life. Proceedings of the National Academy of Sciences 109:4962–4967.

Ackerly, D. D., and P. B. Reich. 1999. Convergence and correlations among leaf size and function in seed plants: comparative tests using independent contrast. American Journal of Botany 86:1272–1281.

Adams, D. C., and J. D. Nason. 2018. A phylogenetic comparative method for evaluating trait coevolution across two phylogenies for sets of interacting species. Evolution 72:234–243.

Agrawal, A. A. 2004. Resistance and susceptibility of milkweek: Competition, root herbivory, and plant genetic variation. Ecology 85:2118–2133.

Agrawal, A. A. 2006. Community heterogeneity and the evolution of interactions between plants and insect herbivores. The Quarterly Review of Biology 81:349–376.

Ainsworth, T. D., L. Krause, T. Bridge, G. Torda, J. B. Raina, M. Zakrzewski, R. D. Gates, J. L. Padilla-Gamiño, H. L. Spalding, C. Smith, E. S. Woolsey, D. G. Bourne, P. Bongaerts, O. Hoegh-Guldberg, and W. Leggat. 2015. The coral core microbiome identifies rare bacterial taxa as ubiquitous endosymbionts. ISME Journal 9:2261–2274.

Akinsanya, M. A., J. K. Goh, S. P. Lim, and A. S. Y. Ting. 2015. Metagenomics study of endophytic bacteria in Aloe vera using next-generation technology. Genomics Data 6:159–163.

Anderegg, L. D. L., L. T. Berner, G. Badgley, M. L. Sethi, B. E. Law, and J.

HilleRisLambers. 2018. Within-species patterns challenge our understanding of the leaf economics spectrum. Ecology Letters 21:734–744.

Anderson, B., J. S. Terblanche, and A. G. Ellis. 2010. Predictable patterns of trait mismatches between interacting plants and insects. BMC Evolutionary Biology 10:204.

Ashoori, M., and A. Saedisomeolia. 2014. Riboflavin (vitamin B2) and oxidative stress: A review. British Journal of Nutrition 111:1985–1991.

Atamna-Ismaeel, N., O. M. Finkel, F. Glaser, I. Sharon, R. Schneider, A. F. Post, J. L. Spudich, C. von Mering, J. A. Vorholt, D. Iluz, O. Béjà, and S. Belkin. 2012. Microbial rhodopsins on leaf surfaces of terrestrial plants. Environmental Microbiology 14:140–146.

Bahram, M., F. Hildebrand, S. K. Forslund, J. L. Anderson, N. A. Soudzilovskaia, P. M. Bodegom, J. Bengtsson-Palme, S. Anslan, L. P. Coelho, H. Harend, J. Huerta-Cepas, M. H. Medema, M. R. Maltz, S. Mundra, P. A. Olsson, M. Pent, S. Põlme, S. Sunagawa, M. Ryberg, L. Tedersoo, and P. Bork. 2018. Structure and function of the global topsoil microbiome. Nature 560:233–237.

Bai, Y., D. B. Müller, G. Srinivas, R. Garrido-Oter, E. Potthoff, M. Rott, N. Dombrowski, P. C. Münch, S. Spaepen, M. Remus-Emsermann, B. Hüttel, A. C. McHardy, J. A. Vorholt, and P. Schulze-Lefert. 2015. Functional overlap of the *Arabidopsis* leaf and root microbiota. Nature 528:364–369.

Bailey, J. K., R. Deckert, J. A. Schweitzer, B. J. Rehill, R. L. Lindroth, C. Gehring, and T. G. Whitham. 2005. Host plant genetics affects hidden ecological players: links among *Populus*, condensed tannins, and fungal endophyte infection. Canadian Journal of Botany 83:356–361.

Bar-On, Y. M., R. Phillips, and R. Milo. 2018. The biomass distribution on Earth. Proceedings of the National Academy of Sciences 115:6506–6511.

Barberan, A., J. Henley, N. Fierer, and E. O. Casamayor. 2014. Structure, inter-annual recurrence, and global-scale connectivity of airborne microbial communities. Science of the Total Environment 487:187–195.

Barberán, A., J. Ladau, J. W. Leff, K. S. Pollard, H. L. Menninger, R. R. Dunn, and N. Fierer. 2015. Continental-scale distributions of dust-associated bacteria and fungi.

Proceedings of the National Academy of Sciences 112:5756–61.

Barott, K. L., B. Rodriguez-Brito, J. Janouškovec, K. L. Marhaver, J. E. Smith, P. Keeling, and F. L. Rohwer. 2011. Microbial diversity associated with four functional groups of benthic reef algae and the reef-building coral *Montastraea annularis*. Environmental Microbiology 13:1192–1204.

Basenko, E. Y., J. A. Pulman, A. Shanmugasundram, O. S. Harb, K. C. Id, D. S. Id, S. Warrenfeltz, C. Aurrecoechea, C. J. S. Jr, J. C. Kissinger, D. S. Roos, and C. Hertz-fowler. 2018. FungiDB : An integrated bioinformatic resource for fungi and oomycetes. Journal of Fungi 4:jof4010039.

Bashiardes, S., G. Zilberman-Schapira, and E. Elinav. 2016. Use of metatranscriptomics in microbiome research. Bioinformatics and Biology Insights 10:19–25.

Bassar, R. D., T. Simon, W. Roberts, J. Travis, and D. N. Reznick. 2017. The evolution of coexistence: Reciprocal adaptation promotes the assembly of a simple community. Evolution 71:373–385.

Berg, G., M. Grube, M. Schloter, and K. Smalla. 2014. The plant microbiome and its importance for plant and human health. Frontiers in Microbiology 5:491.

Bergey, D. H. 1923. Bergey's Manual of Determinative Bacteriology. Williams & Wilkins Co, Baltimore.

Bertness, M. D., and R. Callaway. 1994. Positive interactions in communities. Trends in Ecology & Evolution 9:191–193.

Bier, R. L., E. S. Bernhardt, C. M. Boot, E. B. Graham, E. K. Hall, J. T. Lennon, D. R. Nemergut, B. B. Osborne, C. Ruiz-González, J. P. Schimel, M. P. Waldrop, and M. D. Wallenstein. 2015. Linking microbial community structure and microbial processes: An empirical and conceptual overview. FEMS Microbiology Ecology 91:1–11.

Bissett, A., A. E. Richardson, G. Baker, S. Wakelin, and P. H. Thrall. 2010. Life history determines biogeographical patterns of soil bacterial communities over multiple spatial scales. Molecular Ecology 19:4315–4327.

Björk, J. R., C. Díez-Vives, R. Coma, M. Ribes, and J. M. Montoya. 2013. Specificity and temporal dynamics of complex bacteria – sponge symbiotic interactions. Ecology 94:2781–2791.

Blumenstein, K., B. R. Albrectsen, J. A. Martin, M. Hultberg, T. N. Sieber, M. Helander, and J. Witzell. 2015. Nutritional niche overlap potentiates the use of endophytes in biocontrol of a tree disease. BioControl 60:655–667.

Bochner, B. 1989. "Breathprints" at the microbial level. ASM News 55:536–539.

Bodenhausen, N., M. Bortfeld-Miller, M. Ackermann, and J. A. Vorholt. 2014. A synthetic community approach reveals plant genotypes affecting the phyllosphere microbiota. PLoS Genetics 10:e1004283.

Bose, T., M. M. Haque, C. Reddy, and S. S. Mande. 2015. COGNIZER: A framework for functional annotation of metagenomic datasets. PLoS ONE 10:e0142102.

Brigandt, I. 2013. Integration in biology: Philosophical perspectives on the dynamics of interdisciplinarity. Studies in History and Philosophy of Biological and Biomedical Sciences 44:461–465.

Bringel, F., and I. Couée. 2015. Pivotal roles of phyllosphere microorganisms at the interface between plant functioning and atmospheric trace gas dynamics. Frontiers in Microbiology 6:486.

Brockhurst, M. A., M. E. Hochberg, T. Bell, and A. Buckling. 2006. Character displacement promotes cooperation in bacterial biofilms. Current Biology 16:2030–2034.

Brown, C. D., and M. Vellend. 2014. Non-climatic constraints on upper elevational plant range expansion under climate change. Proceedings of the Royal Society B: Biological Sciences 281:20141779.

Bruns, T. D. 2018. The developing relationship between the study of fungal communities and community ecology theory. Fungal Ecology 39:393–402.

Bryant, J. A., C. Lamanna, H. Morlon, A. J. Kerkhoff, B. J. Enquist, and J. L. Green. 2008. Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. Proceedings of the National Academy of Sciences 105:11505–

11511.

Bulgarelli, D., K. Schlaeppi, S. Spaepen, E. van Themaat, and P. Schulze-Lefert. 2013. Structure and functions of the bacterial microbiota of plants. Annual Reviews of Plant Biology 64:807–838.

Burke, C., P. Steinberg, D. Rusch, S. Kjelleberg, and T. Thomas. 2011. Bacterial community assembly based on functional genes rather than species. Proceedings of the National Academy of Sciences 108:14288–14293.

Cadotte, M. W. 2017. Functional traits explain ecosystem function through opposing mechanisms. Ecology Letters 20:989–996.

Callahan, B. J., P. J. Mcmurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes. 2016. DADA2 : High-resolution sample inference from Illumina amplicon data. Nature Methods 13:581–583.

Callaway, R. M., R. W. Brooker, P. Choler, Z. Kikvidze, C. J. Lortie, R. Michalet, L. Paolini, F. I. Pugnaire, B. Newingham, E. T. Aschehoug, C. Armas, D. Kikodze, and B. J. Cook. 2002. Positive interactions among alpine plants increase with stress. Nature 417:844–848.

Calow, P. 1987. Towards a definition of functional ecology. Functional Ecology 1:57–61.

Caporaso, J. G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Peña, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, and R. Knight. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7:335–336.

Carteron, A., V. Parasquive, F. Blanchard, X. Guilbeault-Mayers, B. L. Turner, M. Vellend, and E. Laliberté. 2020. Soil abiotic and biotic properties constrain the establishment of a dominant temperate tree into boreal forests. Journal of Ecology In Press.

Carvalho, S. D., and J. A. Castillo. 2018. Influence of light on plant–phyllosphere interaction. Frontiers in Plant Science 9:1482.

Chagnon, P. L., R. L. Bradley, H. Maherali, and J. N. Klironomos. 2013. A trait-based framework to understand life history of mycorrhizal fungi. Trends in Plant Science 18:484–491.

Chamberlain, S. A., J. L. Bronstein, and J. A. Rudgers. 2014. How context dependent are species interactions ? Ecology Letters 17:881–890.

Charlop-Powers, Z., J. G. Owen, B. V. B. Reddy, M. A. Ternei, and S. F. Brady. 2014. Chemical-biogeographic survey of secondary metabolism in soil. Proceedings of the National Academy of Sciences 111:3757–3762.

Chave, J. 2005. Measuring wood density for tropical forest trees. Pages 1–7 Wood density measurement protocol.

Chen, I. A., K. Chu, K. Palaniappan, M. Pillay, A. Ratner, J. Huang, M. Huntemann, N. Varghese, J. R. White, R. Seshadri, T. Smirnova, E. Kirton, S. P. Jungbluth, T. Woyke, N. N. Ivanova, and N. C. Kyrpides. 2019. IMG/M v.5.0 : an integrated data management and comparative analysis system for microbial genomes and microbiomes. Nucleic Acids Research 47:666–677.

Chen, I. M. A., V. M. Markowitz, K. Chu, K. Palaniappan, E. Szeto, M. Pillay, A. Ratner, J. Huang, E. Andersen, M. Huntemann, N. Varghese, M. Hadjithomas, K. Tennessen, T. Nielsen, N. N. Ivanova, and N. C. Kyrpides. 2017. IMG/M: Integrated genome and metagenome comparative data analysis system. Nucleic Acids Research 45:D507–D516.

Chen, P. E., and B. J. Shapiro. 2015. The advent of genome-wide association studies for bacteria. Current Opinion in Microbiology 25:17–24.

Chibucos, M. C., A. E. Zweifel, J. C. Herrera, W. Meza, S. Eslamfam, P. Uetz, D. a Siegele, J. C. Hu, and M. G. Giglio. 2014. An ontology for microbial phenotypes. BMC Microbiology 14:294.

Chisholm, C., Z. Lindo, and A. Gonzalez. 2011. Metacommunity diversity depends on connectivity and patch arrangement in heterogeneous habitat networks. Ecography 34:415–424.

Choi, J., F. Yang, R. Stepanauskas, E. Cardenas, A. Garoutte, R. Williams, J. Flater, J. M. Tiedje, K. S. Hofmockel, B. Gelder, and A. Howe. 2016. Strategies to improve reference databases for soil microbiomes. The ISME Journal 11:829–834.

Clark, J. W., S. Hossain, C. a Burnside, and S. Kambhampati. 2001. Coevolution between a cockroach and its bacterial endosymbiont: a biogeographical perspective. Proceedings of the Royal Society B: Biological Sciences 268:393–398.

Clooney, A. G., F. Fouhy, R. D. Sleator, and A. O. Driscoll. 2016. Comparing apples and oranges ?: Next generation sequencing and its impact on microbiome analysis. PLoS ONE 11:e0148028.

Coleman-Derr, D., D. Desgarennes, C. Fonseca-Garcia, S. Gross, S. Clingenpeel, T. Woyke, G. North, A. Visel, L. P. Partida-Martinez, and S. Tringe. 2016. Plant compartment and biogeography affect microbiome composition in cultivated and native *Agave* species. New Phytologist 209:798–811.

Condit, R., P. S. Ashton, N. Manokaran, J. V Lafrankie, S. P. Hubbell, and R. B. Foster. 1999. Dynamics of the forest communities at Pasoh and Barro Colorado: comparing two 50-ha plots. Philosophical Transactions of the Royal Society B: Biological Sciences 354:1739–1748.

Cornwell, W. K., J. H. C. Cornelissen, K. Amatangelo, E. Dorrepaal, V. T. Eviner, O. Godoy, S. E. Hobbie, B. Hoorens, H. Kurokawa, N. Pérez-Harguindeguy, H. M. Quested, L. S. Santiago, D. A. Wardle, I. J. Wright, R. Aerts, S. D. Allison, P. Van Bodegom, V. Brovkin, A. Chatain, T. V. Callaghan, S. Díaz, E. Garnier, D. E. Gurvich, E. Kazakou, J. A. Klein, J. Read, P. B. Reich, N. A. Soudzilovskaia, M. V. Vaieretti, and M. Westoby. 2008. Plant species traits are the predominant control on litter decomposition rates within biomes worldwide. Ecology Letters 11:1065–1071.

Cornwell, W. K., H. Flores-Moreno, and A. Zanne. 2018. fungaltraits v.0.0.3:http://doi.org/10.5281/zenodo.1216257.

Cox, C. E., M. T. Brandl, M. H. de Moraes, S. Gunasekera, and M. Teplitski. 2018. Production of the plant hormone auxin by Salmonella and its role in the interactions with plants and animals. Frontiers in Microbiology 8:2668.

Danchin, A., and G. Fang. 2016. Unknown unknowns : essential genes in quest for function. Microbial Biotechnology 9:530–540.

Darling, A. E., G. Jospin, E. Lowe, F. A. Matsen, H. M. Bik, and J. A. Eisen. 2014. PhyloSift: phylogenetic analysis of genomes and metagenomes. PeerJ 2:e243.

David, L. A., and E. J. Alm. 2011. Rapid evolutionary innovation during an Archaean genetic expansion. Nature 469:93–96.

DeLeon-Rodriguez, N., T. L. Lathem, L. M. Rodriguez-R, J. M. Barazesh, B. E. Anderson, A. J. Beyersdorf, L. D. Ziemba, M. Bergin, A. Nenes, and K. T. Konstantinidis. 2013. Microbiome of the upper troposphere: species composition and prevalence, effects of tropical storms, and atmospheric implications. Proceedings of the National Academy of Sciences 110:2575–80.

Delgado-Baquerizo, M., A. M. Oliverio, T. E. Brewer, A. Benavent-Gonzalez, D. J. Eldridge, R. D. Bardgett, F. T. Maestre, B. K. Singh, and N. Fierer. 2018. A global atlas of the dominant bacteria found in soil. Science 359:320–325.

Delmotte, N., C. Knief, S. Chaffron, G. Innerebner, B. Roschitzki, R. Schlapbach, C. von Mering, and J. A. Vorholt. 2009. Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. Proceedings of the National Academy of Sciences 106:16428–16433.

Díaz, S., J. Kattge, J. H. C. Cornelissen, I. J. Wright, S. Lavorel, S. Dray, B. Reu, M. Kleyer, C. Wirth, I. C. Prentice, E. Garnier, G. Bönisch, M. Westoby, H. Poorter, P. B. Reich, A. T. Moles, J. Dickie, A. N. Gillison, A. E. Zanne, J. Chave, S. J. Wright, S. N. Sheremet'ev, H. Jactel, B. Christopher, B. Cerabolini, S. Pierce, B. Shipley, D. Kirkup, F. Casanoves, J. S. Joswig, A. Günther, V. Falczuk, N. Rüger, M. D. Mahecha, and L. D. Gorné. 2016. The global spectrum of plant form and function. Nature 529:167–171.

Dinsdale, E. A., R. A. Edwards, D. Hall, F. Angly, M. Breitbart, J. M. Brulc, M. Furlan, C. Desnues, M. Haynes, L. Li, L. McDaniel, M. A. Moran, K. E. Nelson, C. Nilsson, R. Olson, J. Paul, B. R. Brito, Y. Ruan, B. K. Swan, R. Stevens, D. L. Valentine, R. V. Thurber, L. Wegley, B. A. White, and F. Rohwer. 2008. Functional metagenomic profiling of nine biomes. Nature 452:629–632.

Direction des inventaires forestiers du Ministère des Forêts Faune et Parcs. 2015. Carte écoforestière à l'échelle 1 : 20 000. https://www.donneesquebec.ca/recherche/fr/dataset/carte-ecoforestiere-pdf.

Dobzhansky, T. 1950. Evolution in the tropics. American Scientist 8:209–221.

Dolan, K. L., J. Peña, S. D. Allison, and J. B. H. Martiny. 2017. Phylogenetic conservation of substrate use specialization in leaf litter bacteria. PLoS ONE

12:e0174472.

Drenovsky, R., K. Steenwerth, L. Jackson, and K. M. Scow. 2010. Land use and climatic factors structure regional patterns in soil microbial communities. Global Ecology and Biogeography 19:27–39.

Dudhagara, P., S. Bhavsar, C. Bhagat, A. Ghelani, S. Bhatt, and R. Patel. 2015. Web resources for metagenomics studies. Genomics, Proteomics & Bioinformatics 13:296–303.

Eck, J. L., S. M. Stump, C. S. Delavaux, S. A. Mangan, and L. S. Comita. 2019. Evidence of within-species specialization by soil microbes and the implications for plant community diversity. Proceedings of the National Academy of Sciences 116:7371–7376.

Ereshefsky, M., and M. Pedroso. 2013. Biological individuality: The case of biofilms. Biology and Philosophy 28:331–349.

ESRI. 2011. ArcGIS Desktop: Release 10.5. Redlands, CA: Environmental Systems Research Institute.

Evans, S. E., and M. D. Wallenstein. 2014. Climate change alters ecological strategies of soil bacteria. Ecology Letters 17:155–164.

Fadrosh, D. W., P. G. Bing Ma, N. Sengamalay, S. Ott, R. M. Brotman, and J. Ravel. 2014. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. Microbiome 2:6.

Fierer, N., M. A. Bradford, and R. B. Jackson. 2007. Toward an ecological classification of soil bacteria. Ecology 88:1354–1364.

Fierer, N., J. W. Leff, B. J. Adams, U. N. Nielsen, S. Thomas, C. L. Lauber, S. Owens, J. A. Gilbert, D. H. Wall, and J. G. Caporaso. 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proceedings of the National Academy of Sciences 109:21390–21395.

Finkel, O. M., A. Y. Burch, S. E. Lindow, A. F. Post, and S. Belkin. 2011. Geographical location determines the population structure in phyllosphere microbial communities of a salt-excreting desert tree. Applied and Environmental

Microbiology 77:7647–7655.

Finkel, O. M., T. O. Delmont, A. F. Post, and S. Belkin. 2016. Metagenomic signatures of bacterial adaptation to life in the phyllosphere of a salt-secreting desert tree. Applied and Environmental Microbiology 82:2854–2861.

Firrincieli, A., M. Khorasani, A. C. Frank, and S. L. Doty. 2020. Influences of climate on phyllosphere endophytic bacterial communities of wild poplar. Frontiers in Plant Science 11:203.

Follows, M. J., S. Dutkiewicz, S. Grant, and S. Chisholm. 2007. Emergent biogeography of microbial communities in a model ocean. Science 315:1843–1846.

Frenette-Dussault, C., B. Shipley, D. Meziane, and Y. Hingrat. 2013. Trait-based climate change predictions of plant community structure in arid steppes. Journal of Ecology 101:484–492.

Futuyma, D. J. 2010. Evolutionary constraint and ecological consequences. Evolution 64:1865–1884.

Galili, T. 2015. dendextend: an R package for visualizing, adjusting, and comparing trees of hierarchical clustering. Bioinformatics 31:3718–3720.

Gandon, S. 2004. Evolution of multihost parasites. Evolution 58:455–469.

Garcia-Pichel, F., V. Loza, Y. Marusenko, P. Mateo, and R. M. Potrafka. 2013. Temperature drives the continental-scale distribution of key microbes in topsoil communities. Science 340:1574–1577.

Gawad, C., W. Koh, and S. R. Quake. 2016. Single-cell genome sequencing: current state of the science. Nature Reviews Genetics 17:175–188.

Glassman, S. I., C. Weihe, J. Li, M. B. N. Albright, C. I. Looby, and A. C. Martiny. 2018. Decomposition responses to climate depend on microbial community composition. Proceedings of the National Academy of Sciences 115:11994–11999.

Goberna, M., and M. Verdú. 2016. Predicting microbial traits with phylogenies. The

ISME Journal 10:959–967.

Gouvernement du Canada. 2017. Normales climatiques canadiennes. http://climat.meteo.gc.ca/climate_normals/index_f.html.

Grime, J. P. 1977. Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. The American Naturalist 111:1169–1194.

Haggerty, J. M., and E. A. Dinsdale. 2017. Distinct biogeographical patterns of marine bacterial taxonomy and functional genes. Global Ecology and Biogeography 26:177–190.

Hall, E. K., E. S. Bernhardt, R. L. Bier, M. A. Bradford, C. M. Boot, J. B. Cotner, P. A. del Giorgio, S. E. Evans, E. B. Graham, S. E. Jones, J. T. Lennon, K. Locey, D. R. Nemergut, B. Osborne, J. D. Rocca, J. S. Schimel, M. P. Waldrop, and M. W. Wallenstein. 2018. Understanding how microbiomes influence the systems they inhabit. Nature Microbiology 3:977–982.

Handa, I. T., R. Aerts, F. Berendse, M. P. Berg, A. Bruder, O. Butenschoen, E. Chauvet, M. O. Gessner, J. Jabiol, M. Makkonen, B. G. McKie, B. Malmqvist, E. T. H. M. Peeters, S. Scheu, B. Schmid, J. van Ruijven, V. C. A. Vos, and S. Hättenschwiler. 2014. Consequences of biodiversity loss for litter decomposition across biomes. Nature 509:218–221.

Hartmann, A., M. Schmid, D. van Tuinen, and G. Berg. 2009. Plant-driven selection of microbes. Plant and Soil 321:235–257.

Heintz-Buschart, A., and P. Wilmes. 2018. Human gut microbiome: Function matters. Trends in Microbiology 26:563–574.

Hendrickson, H. L., D. Barbeau, R. Ceschin, and J. G. Lawrence. 2018. Chromosome architecture constrains horizontal gene transfer in bacteria. PLoS Genetics 14:e1007421.

Hentschel, U., J. Hopke, M. Horn, A. B. Friedrich, M. Wagner, J. Hacker, and B. S. Moore. 2002. Molecular evidence for a uniform microbial community in sponges from different oceans. Applied and Environmental Microbiology 68:4431–4440.

Hirano, S. S., and C. D. Upper. 2000. Bacteria in the leaf ecosystem with emphasis on Pseudomonas syringae - a pathogen, ice nucleus, and epiphyte. Microbiology and Molecular Biology Reviews 64:624–653.

Ho, A., F. M. Kerckhof, C. Luke, A. Reim, S. Krause, N. Boon, and P. L. E. Bodelier. 2013. Conceptualizing functional traits and ecological characteristics of methane-oxidizing bacteria as life strategies. Environmental Microbiology Reports 5:335–345.

Hoffman, A. A., and M. J. Hercus. 2000. Environmental stress as an evolutionary force. BioScience 50:217–226.

Högbom, M., and R. Ihalin. 2017. Functional and structural characteristics of bacterial proteins that bind host cytokines. Virulence 8:1592–1601.

Horner-Devine, M. C., and B. J. M. Bohannan. 2006. Phylogenetic clustering and overdispersion in bacterial communities. Ecology 87:100–108.

Human, T., and M. Project. 2012. Structure, function and diversity of the healthy human microbiome. Nature 486:207–214.

Hunter, P. J., P. Hand, D. Pink, J. M. Whipps, and G. D. Bending. 2010. Both leaf properties and microbe-microbe interactions influence within-species variation in bacterial population diversity and structure in the lettuce (*Lactuca* species) phyllosphere. Applied and Environmental Microbiology 76:8117–8125.

Huttenhower, C., J. Fah Sathirapongsasuti, N. Segata, D. Gevers, A. M. Earl, M. G. Fitzgerald, S. K. Young, Q. Zeng, E. J. Alm, L. Alvarado, S. Anderson, H. M. Arachchi, T. Bloom, D. M. Ciulla, R. L. Erlich, M. Feldgarden, S. Fisher, D. C. Friedrich, G. Giannoukos, J. M. Goldberg, A. Griggs, S. Gujja, B. J. Haas, T. A. Hepburn, C. Howarth, K. H. Huang, C. Kells, N. Lennon, T. Mehta, C. Nusbaum, M. Pearson, M. E. Priest, C. Russ, N. Shenoy, S. M. Sykes, D. G. Tabbaa, D. V. Ward, C. Yandava, J. D. Zucker, B. W. Birren, R. Knight, J. C. Clemente, C. A. Lozupone, D. McDonald, S. Abubucker, A. T. Chinwalla, R. S. Fulton, K. Hallsworth-Pepin, E. A. Lobos, V. Magrini, J. C. Martin, M. Mitreva, E. J. Sodergren, A. M. Wollam, E. Appelbaum, V. Bhonagiri, L. Chen, S. W. Clifton, K. D. Delehaunty, D. J. Dooling, C. N. Farmer, C. C. Fronick, L. L. Fulton, H. Gao, B. Herter, K. C. Kota, E. R. Mardis, K. A. Mihindukulasuriya, P. J. Minx, M. Oglaughlin, C. Pohl, C. M. Tomlinson, J. Walker, Z. Wang, W. Warren, K. M. Wylie, T. Wylie, L. Ye, Y. Zhou, G. M. Weinstock, R. K. Wilson, J. H. Badger,

R. Madupu, M. Bihan, D. A. Busam, A. Scott Durkin, L. Foster, J. Goll, K. Li, J. M. McCorrison, J. R. Miller, Y.-H. H. Rogers, R. K. Sanka, I. Singh, G. G. Sutton, M. Thiagarajan, M. Torralba, B. A. Methé, K. E. Nelson, H. H. Creasy, M. G. Giglio, J. R. Wortman, O. O. Abolude, C. A. Arze, B. L. Cantarel, J. Crabtree, N. J. Davidovics, V. M. Felix, C. Jordan, A. A. Mahurkar, J. Orvis, J. Ravel, L. Schriml, J. R. White, O. White, D. M. Muzny, K. C. Worley, C. J. Buhay, Y. Ding, S. P. Dugan, M. E. Holder, H. Jiang, V. Joshi, C. L. Kovar, S. L. Lee, L. Lewis, Y. Liu, I. Newsham, X. Qin, J. G. Reid, K. Wilczek-Boney, Y. Wu, L. Zhang, Y. Zhu, R. A. Gibbs, S. K. Highlander, J. F. Petrosino, J. Versalovic, K. M. Aagaard, E. Allen-Vercoe, G. L. Andersen, G. Armitage, T. Ayvaz, W. A. Keitel, M. C. Ross, B. P. Youmans, C. C. Baker, L. Begg, T. Belachew, J. L. Campbell, C. Deal, V. Di Francesco, C. Giblin, M. Y. Giovanni, M. J. Blaser, V. Bonazzi, S. Chhibba, J. McEwen, J. Peterson, L. M. Proctor, J. A. Schloss, L. Wang, C. Wellington, K. A. Wetterstrand, J. Paul Brooks, G. A. Buck, M. C. Rivera, N. U. Sheth, S. R. Canon, P. S. G. Chain, C.-C. C. Lo, M. Scholz, N. C. Kyrpides, K. Liolios, V. M. Markowitz, K. Mavromatis, I. Pagani, I.-M. A. M. A. Chen, K. Chu, K. Palaniappan, M. A. Cutting, H. A. Hamilton, E. L. Harris, R. Dwayne Lunsford, P. McInnes, C. C. Davis, T. Z. Desantis, F. E. Dewhirst, J. Izard, K. P. Lemon, E. Deych, P. S. La Rosa, W. D. Shannon, W. Michael Dunne, M. A. Watson, R. C. Edgar, R. M. Farrell, R. R. Sharp, K. Faust, J. Raes, A. A. Fodor, L. J. Forney, J. Friedman, C. S. Smillie, N. Garcia, A. Gonzalez, D. Knights, S. Kinder Haake, D. E. Hoffmann, S. M. Huse, J. K. Jansson, J. A. Katancik, S. T. Kelley, B. Rodriguez-Mueller, N. B. King, H. H. Kong, O. Koren, R. E. Ley, S. Koren, B. Liu, M. Pop, D. D. Sommer, C. M. Lewis, P. Spicer, T. Madden, P. J. Mannon, A. L. McGuire, S. M. Patel, M. Podar, T. A. Vishnivetskaya, K. S. Pollard, T. J. Sharpton, R. M. Truty, M. Rho, Y. Ye, R. Rhodes, K. P. Riehle, P. Sankar, P. D. Schloss, A. M. Schubert, T. M. Schmidt, G. A. Simone, J. D. Sobel, T. J. Treangen, S. Yooseph, L. Zoloth, S. Conlan, J. A. Segre, A. T. Chinwalla, H. H. Creasy, A. M. Earl, M. G. Fitzgerald, R. S. Fulton, M. G. Giglio, K. Hallsworth-Pepin, E. A. Lobos, R. Madupu, V. Magrini, J. C. Martin, M. Mitreva, D. M. Muzny, E. J. Sodergren, J. Versalovic, A. M. Wollam, K. C. Worley, J. R. Wortman, S. K. Young, Q. Zeng, K. M. Aagaard, O. O. Abolude, E. Allen-Vercoe, E. J. Alm, L. Alvarado, G. L. Andersen, S. Anderson, E. Appelbaum, H. M. Arachchi, G. Armitage, C. A. Arze, T. Ayvaz, C. C. Baker, L. Begg, T. Belachew, V. Bhonagiri, M. Bihan, M. J. Blaser, T. Bloom, V. Bonazzi, J. Paul Brooks, G. A. Buck, C. J. Buhay, D. A. Busam, J. L. Campbell, S. R. Canon, B. L. Cantarel, P. S. G. Chain, I.-M. A. M. A. Chen, L. Chen, S. Chhibba, K. Chu, D. M. Ciulla, J. C. Clemente, S. W. Clifton, S. Conlan, J. Crabtree, M. A. Cutting, N. J. Davidovics, C. C. Davis, T. Z. Desantis, C. Deal, K. D. Delehaunty, F. E. Dewhirst, E. Deych, Y. Ding, D. J. Dooling, S. P. Dugan, W. Michael Dunne, A. Scott Durkin, R. C. Edgar, R. L. Erlich, C. N. Farmer, R. M. Farrell, K. Faust, M. Feldgarden, V. M. Felix, S. Fisher, A. A. Fodor, L. J. Forney, L. Foster, V. Di Francesco, J. Friedman, D. C. Friedrich, C. C. Fronick, L. L. Fulton, H. Gao, N. Garcia, G. Giannoukos, C.

Giblin, M. Y. Giovanni, J. M. Goldberg, J. Goll, A. Gonzalez, A. Griggs, S. Gujja, S. Kinder Haake, B. J. Haas, H. A. Hamilton, E. L. Harris, T. A. Hepburn, B. Herter, D. E. Hoffmann, M. E. Holder, C. Howarth, K. H. Huang, S. M. Huse, J. Izard, J. K. Jansson, H. Jiang, C. Jordan, V. Joshi, J. A. Katancik, W. A. Keitel, S. T. Kelley, C. Kells, N. B. King, D. Knights, H. H. Kong, O. Koren, S. Koren, K. C. Kota, C. L. Kovar, N. C. Kyrpides, P. S. La Rosa, S. L. Lee, K. P. Lemon, N. Lennon, C. M. Lewis, L. Lewis, R. E. Ley, K. Li, K. Liolios, B. Liu, Y. Liu, C.-C. C. Lo, C. A. Lozupone, R. Dwayne Lunsford, T. Madden, A. A. Mahurkar, P. J. Mannon, E. R. Mardis, V. M. Markowitz, K. Mavromatis, J. M. McCorrison, D. McDonald, J. McEwen, A. L. McGuire, P. McInnes, T. Mehta, K. A. Mihindukulasuriya, J. R. Miller, P. J. Minx, I. Newsham, C. Nusbaum, M. O'Laughlin, J. Orvis, I. Pagani, K. Palaniappan, S. M. Patel, M. Pearson, J. Peterson, M. Podar, C. Pohl, K. S. Pollard, M. Pop, M. E. Priest, L. M. Proctor, X. Qin, J. Raes, J. Ravel, J. G. Reid, M. Rho, R. Rhodes, K. P. Riehle, M. C. Rivera, B. Rodriguez-Mueller, Y.-H. H. Rogers, M. C. Ross, C. Russ, R. K. Sanka, P. Sankar, J. Fah Sathirapongsasuti, J. A. Schloss, P. D. Schloss, T. M. Schmidt, M. Scholz, L. Schriml, A. M. Schubert, N. Segata, J. A. Segre, W. D. Shannon, R. R. Sharp, T. J. Sharpton, N. Shenoy, N. U. Sheth, G. A. Simone, I. Singh, C. S. Smillie, J. D. Sobel, D. D. Sommer, P. Spicer, G. G. Sutton, S. M. Sykes, D. G. Tabbaa, M. Thiagarajan, C. M. Tomlinson, M. Torralba, T. J. Treangen, R. M. Truty, T. A. Vishnivetskaya, J. Walker, L. Wang, Z. Wang, D. V. Ward, W. Warren, M. A. Watson, C. Wellington, K. A. Wetterstrand, J. R. White, K. Wilczek-Boney, Y. Wu, K. M. Wylie, T. Wylie, C. Yandava, L. Ye, Y. Ye, S. Yooseph, B. P. Youmans, L. Zhang, Y. Zhou, Y. Zhu, L. Zoloth, J. D. Zucker, B. W. Birren, R. A. Gibbs, S. K. Highlander, B. A. Methé, K. E. Nelson, J. F. Petrosino, G. M. Weinstock, R. K. Wilson, and O. White. 2012. Structure, function and diversity of the healthy human microbiome. Nature 486:207–214.

Ilhan, Z. E., A. K. Marcus, D.-W. Kang, B. E. Rittmann, and R. Krajmalnik-Brown. 2017. pH-mediated microbial and metabolic interactions in fecal enrichment cultures. mSphere 2:e00047-17.

Inácio, J., P. Pereira, M. De Carvalho, Á. Fonseca, M. T. Amaral-Collaço, and I. Spencer-Martins. 2002. Estimation and diversity of phylloplane mycobiota on selected plants in a Mediterranean-type ecosystem in Portugal. Microbial Ecology 44:344–353.

Inkpen, S. A., G. M. Douglas, T. D. P. Brunet, K. Leuschen, W. F. Doolittle, and M. G. I. Langille. 2017. The coupling of taxonomy and function in microbiomes. Biology & Philosophy 32:1225–1243.

Ivens, A. B. F., A. Gadau, E. T. Kiers, and D. J. C. Kronauer. 2018. Can social partnerships influence the microbiome? Insights from ant farmers and their trophobiont mutualists. Molecular Ecology 27:1898–1914.

Jackson, S. T., editor. 2009. Essay on the Geography of Plants. University of Chicago Press.

Jacobs, J. L., T. L. Carroll, and G. W. Sundin. 2005. The role of pigmentation, ultraviolet radiation tolerance, and leaf colonization strategies in the epiphytic survival of phyllosphere bacteria. Microbial Ecology 49:104–113.

Janda, J. M., and S. L. Abbott. 2002. Bacterial identification for publication : When is enough enough ? Journal of Clinical Microbiology 40:1887–1891.

Jessup, C. M., R. Kassen, S. E. Forde, B. Kerr, A. Buckling, P. B. Rainey, and B. J. M. Bohannan. 2004. Big questions, small worlds: Microbial model systems in ecology. Trends in Ecology and Evolution 19:189–197.

Johnson, Z. I., E. R. Zinser, A. Coe, N. P. Mcnulty, E. S. Malcolm, S. W. Chisholm, E. M. S. Woodward, and S. W. Chisholm. 2006. Partitioning among *Prochlorococcus* ecotypes along environmental gradients. Science 311:1737–1740.

Johnstone, I. M., and D. M. Titterington. 2009. Statistical challenges of high-dimensional data. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 367:4237–4253.

Jombart, T., and S. Dray. 2010. adephylo: exploratory analyses for the phylogenetic comparative method. Bioinformatics 26:1907–1909.

Jombart, T., S. Pavoine, S. Devillard, and D. Pontier. 2010. Putting phylogeny into the analysis of biological traits : a methodological approach. Journal of Theoretical Biology 264:693–701.

Jorge, L. R., V. Novotny, S. T. Segar, G. D. Weiblen, S. E. Miller, Y. Basset, and T. M. Lewinsohn. 2017. Phylogenetic trophic specialization: a robust comparison of herbivorous guilds. Oecologia 185:551–559.

Jorge, L. R., P. I. Prado, M. Almeida-Neto, and T. M. Lewinsohn. 2014. An integrated

framework to improve the concept of resource specialisation. Ecology Letters 17:1341–1350.

Kaczmarczyk, A., R. Hochstrasser, J. A. Vorholt, and A. Francez-Charlot. 2014. Complex two-component signaling regulates the general stress response in Alphaproteobacteria. Proceedings of the National Academy of Sciences 111:E5196–E5204.

Kamke, J., A. Sczyrba, N. Ivanova, P. Schwientek, C. Rinke, K. Mavromatis, T. Woyke, and U. Hentschel. 2013. Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. ISME Journal 7:2287–2300.

Kanehisa, M., S. Goto, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe. 2014. Data, information, knowledge and principle: Back to metabolism in KEGG. Nucleic Acids Research 42:199–205.

Kattge, J., S. Díaz, S. Lavorel, I. C. Prentice, P. Leadley, G. Bönisch, E. Garnier, M. Westoby, P. B. Reich, I. J. Wright, J. H. C. Cornelissen, C. Violle, S. P. Harrison, P. M. Van Bodegom, M. Reichstein, B. J. Enquist, N. A. Soudzilovskaia, D. D. Ackerly, M. Anand, O. Atkin, M. Bahn, T. R. Baker, D. Baldocchi, R. Bekker, C. C. Blanco, B. Blonder, W. J. Bond, R. Bradstock, D. E. Bunker, F. Casanoves, J. Cavender-Bares, J. Q. Chambers, F. S. Chapin, J. Chave, D. Coomes, W. K. Cornwell, J. M. Craine, B. H. Dobrin, L. Duarte, W. Durka, J. Elser, G. Esser, M. Estiarte, W. F. Fagan, J. Fang, F. Fernández-Méndez, A. Fidelis, B. Finegan, O. Flores, H. Ford, D. Frank, G. T. Freschet, N. M. Fyllas, R. V. Gallagher, W. A. Green, A. G. Gutierrez, T. Hickler, S. I. Higgins, J. G. Hodgson, A. Jalili, S. Jansen, C. A. Joly, A. J. Kerkhoff, D. Kirkup, K. Kitajima, M. Kleyer, S. Klotz, J. M. H. Knops, K. Kramer, I. Kühn, H. Kurokawa, D. Laughlin, T. D. Lee, M. Leishman, F. Lens, T. Lenz, S. L. Lewis, J. Lloyd, J. Llusià, F. Louault, S. Ma, M. D. Mahecha, P. Manning, T. Massad, B. E. Medlyn, J. Messier, A. T. Moles, S. C. Müller, K. Nadrowski, S. Naeem, Ü. Niinemets, S. Nöllert, A. Nüske, R. Ogaya, J. Oleksyn, V. G. Onipchenko, Y. Onoda, J. Ordoñez, G. Overbeck, W. A. Ozinga, S. Patiño, S. Paula, J. G. Pausas, J. Peñuelas, O. L. Phillips, V. Pillar, H. Poorter, L. Poorter, P. Poschlod, A. Prinzing, R. Proulx, A. Rammig, S. Reinsch, B. Reu, L. Sack, B. Salgado-Negret, J. Sardans, S. Shiodera, B. Shipley, A. Siefert, E. Sosinski, J. F. Soussana, E. Swaine, N. Swenson, K. Thompson, P. Thornton, M. Waldram, E. Weiher, M. White, S. White, S. J. Wright, B. Yguel, S. Zaehle, A. E. Zanne, and C. Wirth. 2011. TRY - a global database of plant traits. Global Change Biology 17:2905–2935.

Keddy, P. A. 1992. A pragmatic approach to functional ecology. Functional Ecology 6:621–626.

Keegan, K., E. Glass, and F. Meyer. 2016. MG-RAST, a metagenomics service for analysis of microbial community structure and function. Methods in Molecular Biology 1399:207–233.

Kembel, A. S. W., D. D. Ackerly, S. P. Blomberg, W. K. Cornwell, P. D. Cowan, M. R. Hel-, H. Morlon, and C. O. Webb. 2013. Package ‘ picante .’

Kembel, S., and R. Mueller. 2014. Plant traits and taxonomy drive host associations in tropical phyllosphere fungal communities. Botany 92:303–311.

Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26:1463–1464.

Kembel, S. W., T. K. O’Connor, H. K. Arnold, S. P. Hubbell, S. J. Wright, and J. L. Green. 2014. Relationships between phyllosphere bacterial communities and plant functional traits in a neotropical forest. Proceedings of the National Academy of Sciences 111:13715–20.

Kerdraon, L., V. Laval, and F. Suffert. 2019. Microbiomes and pathogen survival in crop residues, an ecotone between plant and soil. Phytobiomes Journal 3:246–255.

Kim, M., D. Singh, A. Lai-hoe, J. Chun, and J. M. Adams. 2012. Distinctive phyllosphere bacterial communities in tropical trees. Microbial Ecology 63:674–681.

Klassen, J. L. 2018. Defining microbiome function. Nature Microbiology 3:864–869.

Knief, C., A. Ramette, L. Frances, C. Alonso-Blanco, and J. A. Vorholt. 2010. Site and plant species are important determinants of the Methylobacterium community composition in the plant phyllosphere. ISME Journal 4:719–728.

Konno, M., S. Iwamoto, and K. Seiwa. 2011. Specialization of a fungal pathogen on host tree species in a cross-inoculation experiment. Journal of Ecology 99:1394–1401.

Koskella, B., L. Hall, and J. Metcalf. 2017. The microbiome beyond the horizon of ecological and evolutionary theory. Nature Ecology and Evolution 1:1606–1615.

Krause, S., P. M. Van Bodegom, W. K. Cornwell, and P. L. E. Bodelier. 2014. Weak phylogenetic signal in physiological traits of methane-oxidizing bacteria. Journal of Evolutionary Biology 27:1240–1247.

Krohn-Molt, I., B. Wemheuer, M. Alawi, A. Poehlein, S. Güllert, C. Schmeisser, A. Pommerening-Röser, A. Grundhoff, R. Daniel, D. Hanelt, and W. R. Streit. 2013. Metagenome survey of a multispecies and alga-associated biofilm revealed key elements of bacterial-algal interactions in photobioreactors. Applied and Environmental Microbiology 79:6196–6206.

Kunstler, G., D. Falster, D. A. Coomes, F. Hui, M. Kooyman, Robert, D. C. Laughlin, L. Poorter, M. Vanderwel, G. Vieilledent, S. J. Wright, M. Aiba, C. Baraloto, J. Caspersen, J. H. C. Cornelissen, S. Gourlet-Fleury, M. Hanewinkel, B. Herault, J. Kattge, H. Kurokawa, Y. Onoda, J. Penuelas, H. Poorter, M. Uriarte, S. Richardson, and P. Ruiz-Benito. 2016. Plant functional traits have globally consistent effects on competition. Nature 529:1–15.

Kuussaari, M., M. Singer, and I. Hanski. 2001. Local specialization and landscape-level influence on host use in an herbivorous insect. Ecology 81:2177–2187.

Laforest-Lapointe, I., C. Messier, and S. W. Kembel. 2016. Host species identity, site and time drive temperate tree phyllosphere bacterial community structure. Microbiome 4:27.

Laforest-Lapointe, I., A. Paquette, C. Messier, and S. Kembel. 2017. Leaf bacterial diversity mediates plant diversity and ecosystem function relationships. Nature 546:145–147.

Lagier, J.-C., S. Edouard, I. Pagnier, O. Mediannikov, M. Drancourt, and D. Raoult. 2015. Current and past strategies for bacterial culture in clinical microblogy. Clinical Microbiology Reviews 28:208–236.

Lajoie, G., and S. W. Kembel. 2019. Making the most of trait-based approaches for microbial ecology. Trends in Microbiology 27:814–823.

Lajoie, G., R. Maglione, and S. W. Kembel. 2020. Drivers of phyllosphere microbial functional diversity in a neotropical forest. Microbiome In Press.

Laliberté, E. 2017. Below-ground frontiers in trait-based plant ecology. New Phytologist 213:1597–1603.

Laliberté, E., and P. Legendre. 2010. A distance-based framework for measuring functional diversity from multiple traits. Ecology 91:299–305.

Lambais, M. R., A. R. Lucheta, and D. E. Crowley. 2014. Bacterial community assemblages associated with the phyllosphere, dermosphere, and rhizosphere of tree species of the Atlantic forest are host taxon dependent. Microbial Ecology 68:567–574.

Laughlin, D. C., J. J. Leppert, M. M. Moore, and C. H. Sieg. 2010. A multi-trait test of the leaf-height-seed plant strategy scheme with 133 species from a pine forest flora. Functional Ecology 24:493–501.

Lê, S., J. Josse, and F. Husson. 2008. FactoMineR: An R package for multivariate analysis. Journal of Statistical Software 25:1–18.

Leff, J. W., S. E. Jones, S. M. Prober, A. Barberán, E. T. Borer, J. L. Firn, W. S. Harpole, S. E. Hobbie, K. S. Hofmockel, J. M. H. Knops, R. L. McCulley, K. La Pierre, A. C. Risch, E. W. Seabloom, M. Schütz, C. Steenbock, C. J. Stevens, and N. Fierer. 2015. Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. Proceedings of the National Academy of Sciences 112:10967–72.

Lei, B. R., and K. J. Olival. 2014. Contrasting patterns in mammal-bacteria coevolution: *Bartonella* and *Leptospira* in bats and rodents. PLoS Neglected Tropical Diseases 8:1–11.

Lennon, J. T. 2007. Diversity and metabolism of marine bacteria cultivated on dissolved DNA. Applied and Environmental Microbiology 73:2799–2805.

Lennon, J. T., Z. T. Aanderud, B. K. Lehmkuhl, and D. R. Schoolmaster. 2012. Mapping the niche space of soil microorganisms using taxonomy and traits. Ecology 93:1867–1879.

Lennon, J. T., and B. K. Lehmkuhl. 2016. A trait-based approach to bacterial biofilms in soil. Environmental microbiology 18:2732–2742.

Leonelli, S. 2013. Integrating data to acquire new knowledge : Three modes of integration in plant science. Studies in History and Philosophy of Biological & Biomedical Sciences 44:503–514.

Levine, J. M. 2015. A trail map for trait-based studies. Nature 529:163–164.

Ley, R. E., M. Hamady, C. Lozupone, P. J. Turnbaugh, R. R. Ramey, J. S. Bircher, M. L. Schlegel, T. A. Tucker, M. D. Schrenzel, R. Knight, and J. I. Gordon. 2008. Evolution of mammals and their gut microbes. Science 320:1647–1651.

Li, H., B. Liu, M. L. Mccormack, Z. Ma, and D. Guo. 2017. Diverse belowground resource strategies underlie plant species coexistence and spatial distribution in three grasslands along a precipitation gradient. New Phytologist 216:1140–1150.

Li, Y., and B. Shipley. 2017. An experimental test of CSR theory using a globally calibrated ordination method. PLoS ONE 12:e0175404.

Lindemann, J., and C. Upper. 1985. Aerial dispersal of epiphytic bacteria over bean plants. Applied and Environmental Microbiology 50:1229–1232.

Lindow, S. E., and M. T. Brandl. 2003. Microbiology of the phyllosphere. Applied and Environmental Microbiology 69:1875–1883.

Lindström, E. S., X. M. Feng, W. Granéli, and E. S. Kritzberg. 2010. The interplay between bacterial community composition and the environment determining function of inland water bacteria. Limnology and Oceanography 55:2052–2060.

Louca, S., and M. Doebeli. 2018. Efficient comparative phylogenetics on large trees. Bioinformatics 34:1053–1055.

Louca, S., L. Wegener Parfrey, and M. Doebeli. 2016. Decoupling function and taxonomy in the global ocean microbiome. Science 353:1272–1277.

Ma, L., B. Song, T. Curran, N. Phong, E. Dressaire, M. Roper, and M. Roper. 2016a. Defining individual size in the model filamentous fungus *Neurospora crassa*. Proceedings of the Royal Society B: Biological Sciences 283:20152470.

Ma, Z., B. Sandel, and J. Svenning. 2016b. Phylogenetic assemblage structure of North American trees is more strongly shaped by glacial–interglacial climate variability

in gymnosperms than in angiosperms. Ecology and Evolution 6:3092–3106.

MacArthur, R., and E. Wilson. 1967. The theory of island biogeography. Princeton University Press.

Madani, N., J. S. Kimball, A. P. Ballantyne, D. L. R. Affleck, P. M. Van Bodegom, P. B. Reich, J. Kattge, A. Sala, M. Nazeri, M. O. Jones, M. Zhao, and S. W. Running. 2018. Future global productivity will be affected by plant trait response to climate. Scientific Reports 8:2870.

Mailloux, R. J., R. Bériault, J. Lemire, R. Singh, D. R. Chénier, R. D. Hamel, and V. D. Appanna. 2007. The tricarboxylic acid cycle, an ancient metabolic network with a novel twist. PLoS ONE 2:e690.

Malik, A. A., J. B. H. Martiny, E. L. Brodie, A. C. Martiny, K. K. Treseder, and S. D. Allison. 2020. Defining trait-based microbial strategies with consequences for soil carbon cycling under climate change. The ISME Journal 14:1–9.

Malik, A. A., B. Thomson, A. Whiteley, M. Bailey, and R. Griffiths. 2017. Bacterial physiological adaptations to contrasting edaphic conditions identified using landscape-scale metagenomics. mBio 8:e00799-17.

Marcon, E., and B. Hérault. 2015. entropart : An R package to measure and partition diversity. Journal of Statistical Software 67:1–26.

De Mares, M. C., D. Sipkema, S. Huang, B. Bunk, J. Overmann, and J. D. van Elsas. 2017. Host specificity for bacterial, archaeal and fungal communities determined for high- and low-microbial abundance sponge species in two genera. Frontiers in Microbiology 8:2560.

Maris, V., P. Huneman, A. Coreau, S. Kéfi, R. Pradel, and V. Devictor. 2018. Prediction in ecology: promises, obstacles and clarifications. Oikos 127:171–183.

Markussen, T., and R. L. Marvig. 2014. Environmental heterogeneity drives within-host diversification and evolution of *Pseudomonas aeruginosa*. mBio 5:e01592-14.

Maron, P. A., L. Ranjard, C. Mougel, and P. Lemanceau. 2007. Metaproteomics: A new approach for studying functional microbial ecology. Microbial Ecology

53:486–493.

Martiny, A. C. 2019. High proportions of bacteria are culturable across major biomes. ISME Journal 13:2125–2128.

Martiny, A. C., A. P. K. Tai, D. Veneziano, F. Primeau, and S. W. Chisholm. 2009. Taxonomic resolution, ecotypes and the biogeography of Prochlorococcus. Environmental Microbiology 11:823–832.

Martiny, A. C., K. Treseder, and G. Pusch. 2013. Phylogenetic conservatism of functional traits in microorganisms. The ISME Journal 7:830–8.

Martiny, J. B. H., S. E. Jones, J. T. Lennon, and A. C. Martiny. 2015. Microbiomes in light of traits: A phylogenetic perspective. Science 350:aac9323.

Matallana-Surget, S., C. Villette, L. Intertaglia, F. Joux, M. Bourrain, and P. Lebaron. 2012. Response to UVB radiation and oxidative stress of marine bacteria isolated from South Pacific Ocean and Mediterranean Sea. Journal of Photochemistry and Photobiology B: Biology 117:254–261.

Mayr, E. 1982. The Growth of Biological Thought. Belknap Press, Cambridge, MA.

McGill, B. J., B. J. Enquist, E. Weiher, and M. Westoby. 2006. Rebuilding community ecology from functional traits. Trends in Ecology and Evolution 21:178–185.

McQuilton, P., A. Gonzalez-Beltran, P. Rocca-Serra, M. Thurston, A. Lister, E. Maguire, and S. A. Sansone. 2016. BioSharing: curated and crowd-sourced metadata standards, databases and data policies in the life sciences. Database : the journal of biological databases and curation 2016:baw075.

Mendler, K., H. Chen, D. H. Parks, B. Lobb, L. A. Hug, and A. C. Doxey. 2019. AnnoTree : visualization and exploration of a functionally annotated microbial tree of life. Nucleic Acids Research 47:4442–4448.

Menzel, P., K. L. Ng, and A. Krogh. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nature Communications 7:11257.

Meyer, K. M., and J. H. J. Leveau. 2012. Microbiology of the phyllosphere: A playground for testing ecological concepts. Oecologia 168:621–629.

Mihaljevic, J. R. 2012. Linking metacommunity theory and symbiont evolutionary ecology. Trends in Ecology and Evolution 27:323–329.

Miyake, S., D. K. Ngugi, and U. Stingl. 2016. Phylogenetic diversity, distribution, and cophylogeny of giant bacteria (*Epulopiscium*) with their surgeonfish hosts in the Red Sea. Frontiers in Microbiology 7:285.

Moretti, M., A. T. Dias, F. Bello, F. Altermatt, S. L. Chown, F. M. J. R. B. Azcárate, B. Fournier, M. Hedde, J. Hortal, S. Ibanez, E. Öckinger, J. P. Sousa, J. Ellers, and M. P. Berg. 2016. Handbook of protocols for standardized measurement of terrestrial invertebrate functional traits. Functional Ecology 31:558–567.

Morrison-Whittle, P., and M. R. Goddard. 2015. Quantifying the relative roles of selective and neutral processes in defining eukaryotic microbial communities. ISME Journal 9:2003–2011.

Morrissey, E. M., R. L. Mau, E. Schwartz, J. G. Caporaso, P. Dijkstra, N. van Gestel, B. J. Koch, C. M. Liu, M. Hayer, T. A. McHugh, J. C. Marks, L. B. Price, and B. A. Hungate. 2016. Phylogenetic organization of bacterial activity. The ISME Journal 10:2336–2340.

Moyes, A. B., L. M. Kueppers, J. Pett-Ridge, D. L. Carper, N. Vandehey, J. O'Neil, and A. C. Frank. 2016. Evidence for foliar endophytic nitrogen fixation in a widely distributed subalpine conifer. New Phytologist 210:657–668.

Muir, C. D. 2015. Making pore choices: repeated regime shifts in stomatal ratio. Proceedings of the Royal Society B: Biological Sciences 282:20151498.

Müller, D. B., O. T. Schubert, H. Röst, R. Aebersold, and J. A. Vorholt. 2016. Systems-level proteomics of two ubiquitous leaf commensals reveals complementary adaptive traits for phyllosphere colonization. Molecular & Cellular Proteomics 15:3256–3269.

National Atmospheric and Oceanic Administration. 2017. 1981-2010 U.S. Climate normals. https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/climate-normals/1981-2010-normals-data.

National Research Council (US) Committee on Metagenomics: Challenges and Functional Applications. 2007. The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet. National Academies Press (US), Washington

(DC).

Nayfach, S., and K. S. Pollard. 2016. Toward accurate and quantitative comparative metagenomics. Cell 166:1103–1116.

Nelson, M. B., A. C. Martiny, and J. B. H. Martiny. 2016. Global biogeography of microbial nitrogen-cycling traits in soil. Proceedings of the National Academy of Sciences 113:8033–8040.

Nguyen, N. H., Z. Song, S. T. Bates, S. Branco, L. Tedersoo, J. Menke, J. S. Schilling, and P. G. Kennedy. 2016. FUNGuild : An open annotation tool for parsing fungal community datasets by ecological guild. Fungal Ecology 20:241–248.

Nielsen, J. L., and P. H. Nielsen. 2005. Advances in microscopy: Microautoradiography of single cells. Methods in Enzymology 397:237–256.

Nilsson, R. H., K. Larsson, A. F. S. Taylor, J. Bengtsson-palme, T. S. Jeppesen, D. Schigel, P. Kennedy, K. Picard, F. Oliver, L. Tedersoo, I. Saar, and K. Urmas. 2019. The UNITE database for molecular identification of fungi : handling dark taxa and parallel taxonomic classifications. Nucleic Acids Research 47:D259–D264.

O'Brien, A. M., R. J. H. Sawers, J. Ross-Ibarra, and S. Y. Strauss. 2018. Evolutionary responses to conditionality in species interactions across environmental gradients. The American Naturalist 192:715–730.

O'Malley, M. A. 2013. When integration fails : Prokaryote phylogeny and the tree of life. Studies in History and Philosophy of Biological & Biomedical Sciences 44:551–562.

Ofek-Lalzar, M., N. Sela, M. Goldman-Voronov, S. J. Green, Y. Hadar, and D. Minz. 2014. Niche and host-associated functional signatures of the root surface microbiome. Nature Communications 5:4950.

Ogata, H., S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa. 1999. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research 27:29–34.

Oksanen, A. J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. Mcglinn, P. R. Minchin, R. B. O. Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and E.

Szoecs. 2013. Package ‘ vegan .’

Ontario Ministry of Natural Resources and Forestry. 2007. Forest Resources Inventory (FIM v2 2D) Packaged Product.

Österlund, T., V. Jonsson, and E. Kristiansson. 2017. HirBin : high-resolution identification of differentially abundant functions in metagenomes. BMC Genomics 18:316.

Overmann, J., B. Abt, and J. Sikorski. 2017. Present and future of culturing bacteria. Annual Review of Microbiology 71:711–730.

Packer, L., E. Witt, and H. Tritschler. 1995. Alpha-lipoic acid as a biological antioxidant. Free Radical Biology and Medicine 19:227–250.

Paradis, E., and K. Schliep. 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526–528.

Parkinson, J. A., and S. E. Allen. 1975. A wet oxidation procedure suitable for the determination of nitrogen and mineral nutrients in biological material. Communications in Soil Science and Plant Analysis 6:1–11.

Paslaru, V. 2009. Ecological explanation between manipulation and mechanism description. Philosophy of Science 76:821–837.

Paster, B. J., S. K. Boches, J. L. Galvin, R. E. Ericson, C. N. Lau, V. A. Levanos, A. Sahasrabudhe, and F. E. Dewhirst. 2001. Bacterial diversity in human subgingival plaque. Journal of Bacteriology 183:3770–3783.

Pérez-Harguindeguy, N., S. Diaz, E. Garnier, S. Lavorel, H. Poorter, P. Jaureguiberry, M. S. S. Bret-Harte, W. K. K. Cornwell, J. M. M. Craine, D. E. E. Gurvich, C. Urcelay, E. J. J. Veneklaas, P. B. B. Reich, L. Poorter, I. J. J. Wright, Etc., P. Ray, Etc., S. Díaz, S. Lavorel, H. Poorter, P. Jaureguiberry, M. S. S. Bret-Harte, W. K. K. Cornwell, J. M. M. Craine, D. E. E. Gurvich, C. Urcelay, E. J. J. Veneklaas, P. B. B. Reich, L. Poorter, I. J. J. Wright, P. Ray, L. Enrico, J. G. Pausas, A. C. de Vos, N. Buchmann, G. Funes, F. Quétier, J. G. Hodgson, K. Thompson, H. D. Morgan, H. ter Steege, M. G. A. van der Heijden, L. Sack, B. Blonder, P. Poschlod, M. V. Vaieretti, G. Conti, A. C. Staver, S. Aquino, and J. H. C. Cornelissen. 2013. New handbook for standardized measurement of plant functional traits worldwide. Australian Journal of Botany 61:167–234.

Pérez, M. J., E. Falqué, and H. Domínguez. 2016. Antimicrobial action of compounds from marine seaweed. Marine Drugs 14:1–38.

Perofsky, A. C., R. J. Lewis, and L. A. Meyers. 2019. Terrestriality and bacterial transfer: a comparative study of gut microbiomes in sympatric Malagasy mammals. ISME Journal 13:50–63.

Pey, B., J. Nahmani, A. Auclerc, Y. Capowiez, D. Cluzeau, J. Cortet, T. Decaëns, L. Deharveng, F. Dubs, S. Joimel, C. Briard, F. Grumiaux, M.-A. Laporte, A. Pasquet, C. Pelosi, C. Pernin, J.-F. Ponge, S. Salmon, and M. Hedde. 2014. Current use of and future needs for soil invertebrate functional traits in community ecology. Basic and Applied Ecology 15:194–206.

Pierce, S., D. Negreiros, B. E. L. Cerabolini, J. Kattge, D. Sandra, M. Kleyer, B. Shipley, S. J. Wright, N. A. Soudzilovskaia, G. Onipchenko, P. M. Van Bodegom, C. Frenette-dussault, E. Weiher, X. Pinho, J. H. C. Cornelissen, J. P. Grime, K. Thompson, P. J. Wilson, G. Buffa, O. C. Nyakunga, P. B. Reich, M. Caccianiga, F. Mangili, R. M. Ceriani, A. Luzzaro, G. Brusa, A. Siefert, N. P. U. Barbosa, F. Stuart, and C. Iii. 2017. A global method for calculating plant CSR ecological strategies applied across biomes world-wide. Functional Ecology 31:444–457.

Pigot, A. L., C. Sheard, E. T. Miller, T. Bregman, B. Freeman, N. Seddon, C. H. Trisos, B. Weeks, and J. . Tobias. 2020. Macroevolutionary convergence connects morphological form to ecological function in birds. Nature Ecology & Evolution In Press.

Pineda, A., M. Dicke, C. M. J. Pieterse, and M. J. Pozo. 2013. Beneficial microbes in a changing environment: Are they always helping plants to deal with insects? Functional Ecology 27:574–586.

Pita, L., S. López-Legentil, and P. M. Erwin. 2013. Biogeography and host fidelity of bacterial communities in Ircinia spp. from the Bahamas. Microbial Ecology 66:437–447.

Price, M. N., P. S. Dehal, and A. P. Arkin. 2010. FastTree 2 -- Approximately maximum-likelihood trees for large alignments. PLoS ONE 5:e9490.

Price, M. N., K. M. Wetmore, R. J. Waters, M. Callaghan, J. Ray, H. Liu, V. Jennifer, R. A. Melnyk, J. S. Lamson, Y. Suh, H. K. Carlson, Z. Esquivel, H. Sadeeshkumar, R. Chakraborty, G. M. Zane, B. E. Rubin, J. D. Wall, A. Visel, J. Bristow, M. J.

Blow, and P. Adam. 2018. Mutant phenotypes for thousands of bacterial genes of unknown function. Nature 557:503–509.

Prosser, J. I., B. J. M. Bohannan, T. P. Curtis, R. J. Ellis, M. K. Firestone, R. P. Freckleton, J. L. Green, L. E. Green, K. Killham, J. J. Lennon, a M. Osborn, M. Solan, C. J. van der Gast, and J. P. W. Young. 2007. The role of ecological theory in microbial ecology. Nature Reviews Microbiology 5:384–392.

Provorov, N. A., and N. I. Vorobyov. 2009. Host plant as an organizer of microbial evolution in the beneficial symbioses. Phytochemistry Reviews 8:519–534.

R Core Team. 2013. R: A language and environment for statistical computing. Vienna, Austria.

Raes, J., I. Letunic, T. Yamada, L. J. Jensen, and P. Bork. 2011. Toward molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data. Molecular Systems Biology 7:473.

Rainey, P. B., and M. Travisano. 1998. Adaptive radiation in a heterogeneous environment. Nature 394:69–72.

Ramírez-Flandes, S., B. González, and O. Ulloa. 2019. Redox traits characterize the organization of global microbial communities. Proceedings of the National Academy of Sciences 116:3630–3635.

Ran, J. H., T. T. Shen, M. M. Wang, and X. Q. Wang. 2018. Phylogenomics resolves the deep phylogeny of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms. Proceedings of the Royal Society B: Biological Sciences 285:20181012.

Rastogi, G., G. L. Coaker, and J. H. J. Leveau. 2013. New insights into the structure and function of phyllosphere microbiota through high-throughput molecular approaches. FEMS Microbiology Letters 348:1–10.

Rastogi, G., J. J. Tech, G. L. Coaker, and J. H. J. Leveau. 2010. A PCR-based toolbox for the culture-independent quantification of total bacterial abundances in plant environments. Journal of Microbiological Methods 83:127–132.

Redford, A. J., R. M. Bowers, R. Knight, Y. Linhart, and N. Fierer. 2010. The ecology

of the phyllosphere: geographic and phylogenetic variability in the distribution of bacteria on tree leaves. Environmental Microbiology 12:2885–2893.

Reed, D. C., C. K. Algar, J. A. Huber, and G. J. Dick. 2014. Gene-centric approach to integrating environmental genomics and biogeochemical models. Proceedings of the National Academy of Sciences 111:1879–1884.

Reimer, L. C., A. Vetcininova, J. Carbasse, C. Sohngen, D. Gleim, C. Ebeling, and J. Overmann. 2019. BacDive in 2019 : bacterial phenotypic data for High-throughput biodiversity analysis. Nucleic Acids Research 47:631–636.

Reznick, D., M. J. Bryant, and F. Bashey. 2002. r- and K-selection revisited : The role of population regulation in life-history evolution. Ecology 83:1509–1520.

Roane, T., and I. Pepper. 1999. Microbial responses to environmentally toxic cadmium. Microbial Ecology 38:358–364.

Rojo, D., C. Méndez-García, B. A. Raczkowska, R. Bargiela, A. Moya, M. Ferrer, and C. Barbas. 2017. Exploring the human microbiome from multiple perspectives: Factors altering its composition and function. FEMS Microbiology Reviews 41:453–478.

Rowe, S. L., J. S. Norman, and M. L. Friesen. 2018. Coercion in the evolution of plant-microbe communication: a perspective. Molecular Plant-Microbe Interactions 31:789–794.

Ruiz-González, C., J. P. Niño-García, and P. A. del Giorgio. 2015. Terrestrial origin of bacterial communities in complex boreal freshwater networks. Ecology Letters 18:1198–1206.

Ryffel, F., E. J. Helfrich, P. Kiefer, L. Peyriga, J. Portais, J. Piel, and J. A. Vorholt. 2015. Metabolic footprint of epiphytic bacteria on Arabidopsis thaliana leaves. The ISME Journal 10:632–643.

Saleem, M., N. Meckes, Z. H. Pervaiz, M. B. Traw, M. Saleem, and M. Saleem. 2017. Microbial interactions in the phyllosphere increase plant performance under herbivore biotic stress. Frontiers in Microbiology 8:41.

Sambles, C., K. Moore, T. M. Lux, K. Jones, G. R. Littlejohn, J. D. Gouveia, S. J. Aves,

D. J. Studholme, R. Lee, and J. Love. 2017. Metagenomic analysis of the complex microbial consortium associated with cultures of the oil-rich alga <i>Botryococcus braunii>/i>. MicrobiologyOpen 6:e482.

Santillan, E., H. Seshan, F. Constancias, and S. Wuertz. 2019. Trait-based life-history strategies explain succession scenario for complex bacterial communities under varying disturbance. Environmental Microbiology 21:3751–3764.

Sberro, H., N. Greenfield, G. Pavlopoulos, N. Kyrpides, and A. Bhatt. 2019. Large-scale analyses of human microbiomes reveal thousands of small, novel genes and their predicted functions. Cell 178:1245–1259.

Schilling, B., B. W. Gibson, and C. L. Hunter. 2017. Generation of high-quality SWATH acquisition data for label-free quantitative proteomics studies using tripleTOF mass spectrometers. Methods in Molecular Biology 1550:223–233.

Schmitt, S., P. Tsai, J. Bell, J. Fromont, M. Ilan, N. Lindquist, T. Perez, A. Rodrigo, P. J. Schupp, J. Vacelet, N. Webster, U. Hentschel, and M. W. Taylor. 2012. Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. ISME Journal 6:564–576.

Schweitzer, J. A., J. K. Bailey, D. G. Fischer, C. J. LeRoy, E. V. Lonsdorf, T. G. Whitham, and S. C. Hart. 2008. Plant-soil microorganism interactions: heritable relationship between plant genotype and microorganisms. Ecology 89:773–781.

Setiawan, N. N., M. Vanhellemont, L. Baeten, R. Gobin, P. De Smedt, W. Proesmans, E. Ampoorter, and K. Verheyen. 2016. Does neighbourhood tree diversity affect the crown arthropod community in saplings? Biodiversity and Conservation 25:169–185.

Severin, I., Ö. Östman, and E. S. Lindström. 2013. Variable effects of dispersal on productivity of bacterial communities due to changes in functional trait composition. PLoS ONE 8:e80825.

Shafquat, A., R. Joice, S. L. Simmons, and C. Huttenhower. 2014. Functional and phylogenetic assembly of microbial communities in the human microbiome. Trends in Microbiology 22:261–266.

Shapiro, B. J., and M. F. Polz. 2014. Ordering microbial diversity into ecologically and genetically cohesive units. Trends in Microbiology 22:235–247.

Shi, W., H. Qi, Q. Sun, G. Fan, S. Liu, J. Wang, B. Zhu, H. Liu, F. Zhao, X. Wang, X. Hu, W. Li, J. Liu, Y. Tian, L. Wu, and J. Ma. 2019. gcMeta : a Global Catalogue of Metagenomics platform to support the archiving, standardization and analysis of microbiome data. Nucleic Acids Research 47:637–648.

Smalla, K., G. Wieland, A. Buchner, A. Zock, J. Parzy, S. Kaiser, N. Roskot, H. Heuer, and G. Berg. 2001. Bulk and rhizosphere soil bacterial communities studied by denaturing gradient gel electrophoresis: plant-dependent enrichment and seasonal shifts revealed. Applied and Environmental Microbiology 67:4742–51.

Staley, C., T. J. Gould, P. Wang, J. Phillips, J. B. Cotner, and M. J. Sadowsky. 2014. Core functional traits of bacterial communities in the Upper Mississippi River show limited variation in response to land cover. Frontiers in Microbiology 5:414.

Sunagawa, S., L. P. Coelho, S. Chaffron, J. R. Kultima, K. Labadie, G. Salazar, B. Djahanschiri, G. Zeller, D. R. Mende, A. Alberti, F. M. Cornejo-castillo, P. I. Costea, C. Cruaud, F. Ovidio, S. Engelen, I. Ferrera, J. M. Gasol, L. Guidi, F. Hildebrand, F. Kokoszka, and C. Lepoivre. 2015. Structure and function of the global ocean microbiome. Science 348:1261359.

Sungpalee, W., A. Itoh, M. Kanzaki, and K. Sri-ngernyuang. 2009. Intra- and interspecific variation in wood density and fine-scale spatial distribution of stand-level wood density in a northern Thai tropical montane forest. Journal of Tropical Ecology 25:359–370.

Suzuki, R., and H. Shimodaira. 2006. Pvclust : an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22:1540–1542.

Tack, A. J. M., F. Horns, and A. L. Laine. 2014. The impact of spatial scale and habitat configuration on patterns of trait variation and local adaptation in a wild plant parasite. Evolution 68:176–189.

Taffner, J., A. Erlacher, A. Bragina, C. Berg, C. Moissl-Eichinger, and G. Berg. 2018. What is the role of Archaea in plants? New insights from the vegetation of alpine bogs. mSphere 3:1–14.

Tang, J. 2011. Microbial metabolomics. Current Genomics 12:391–403.

Tatusov, R. L., E. V Koonin, and D. J. Lipman. 1997. A genomic perspective on protein families. Science 278:631–637.

The Gene Ontology Consortium. 2019. The Gene Ontology Resource : 20 years and still GOing strong. Nucleic Acids Research 47:D330–D338.

The UniProt Consortium. 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Research 45:D158–D169.

Thompson. 2005. The geographic mosaic of coevolution. The University of Chicago Press, Chicago, USA.

Thompson, L. R., G. J. Williams, M. F. Haroon, A. Shibl, P. Larsen, J. Shorenstein, R. Knight, and U. Stingl. 2017. Metagenomic covariation along densely sampled environmental gradients in the Red Sea. The ISME Journal 11:138–151.

Thrall, P. H., M. E. Hochberg, J. J. Burdon, and J. D. Bever. 2007. Coevolution of symbiotic mutualists and parasites in a community context. Trends in Ecology and Evolution 22:120–126.

Tipton, L., J. L. Darcy, and N. A. Hynson. 2019. A developing symbiosis : enabling cross-talk between ecologists and microbiome scientists. Frontiers in Microbiology 10:292.

Treseder, K., and M. Allen. 2002. Direct nitrogen and phosphorus limitation of arbuscular mycorrhizal fungi: a model and field test. New Phytologist 155:507–515.

Tripathi, A., C. Marotz, A. Gonzalez, Y. Vázquez-Baeza, S. J. Song, A. Bouslimani, D. McDonald, Q. Zhu, J. G. Sanders, L. Smarr, P. C. Dorrestein, and R. Knight. 2018. Are microbiome studies ready for hypothesis-driven research? Current Opinion in Microbiology 44:61–69.

Tyc, O., C. Song, J. S. Dickschat, M. Vos, and P. Garbeva. 2017. The ecological role of volatile and soluble secondary metabolites produced by soil bacteria. Trends in Microbiology 25:280–292.

US Department of Agriculture - Forest Service - Northern Research Station. 2019. Forest Inventory and Analysis Database.

Vandeputte, D., R. Y. Tito, R. Vanleeuwen, G. Falony, and J. Raes. 2017. Practical considerations for large-scale gut microbiome studies. FEMS Microbiology

Reviews 41:S154–S167.

Veech, J. A., K. S. Summerville, T. O. Crist, and J. C. Gering. 2002. The additive partitioning of species diversity: Recent revival of an old idea. Oikos 99:3–9.

van Veelen, H. P. J., J. Falcao Salles, and B. I. Tieleman. 2017. Multi-level comparisons of cloacal, skin, feather and nest-associated microbiota suggest considerable influence of horizontal acquisition on the microbiota assembly of sympatric woodlarks and skylarks. Microbiome 5:156.

Vepsäläinen, K., and J. R. Spence. 2000. Generalization in ecology and evolutionary biology : From hypothesis to paradigm. Biology and Philosophy 15:211–238.

Violle, C., M. Navas, D. Vile, E. Kazakou, C. Fortunel, I. Hummel, E. Garnier, S. Oikos, N. May, C. Violle, M. Navas, D. Vile, E. Kazakou, and C. Fortunel. 2007. Let the concept of trait be functional ! Oikos 116:882–892.

Vokou, D., K. Vareli, E. Zarali, K. Karamanoli, H. I. A. Constantinidou, N. Monokrousos, J. M. Halley, and I. Sainis. 2012. Exploring biodiversity in the bacterial community of the Mediterranean phyllosphere and its relationship with airborne bacteria. Microbial Ecology 64:714–724.

Vorholt, J. A. 2012. Microbial life in the phyllosphere. Nature Reviews Microbiology 10:828–840.

Wagi, S., and A. Ahmed. 2017. Phyllospheric plant growth promoting bacteria. Journal of Bacteriology and Mycology 5:215–216.

van der Wal, A., and J. H. J. Leveau. 2011. Modelling sugar diffusion across plant leaf cuticles : the effect of free water on substrate availability to phyllosphere bacteria. Environmental Microbiology 13:792–797.

Wallenstein, M. D., and E. K. Hall. 2012. A trait-based framework for predicting when and where microbial adaptation to climate change will affect ecosystem functioning. Biogeochemistry 109:35–47.

Webb, C. O., D. D. Ackerly, M. a. McPeek, and M. J. Donoghue. 2002. Phylogenies and Community Ecology. Annual Review of Ecology and Systematics 33:475–505.

Webb, C. O., and M. J. Donoghue. 2005. Phylomatic : tree assembly for applied phylogenetics. Molecular Ecology Notes 5:181–183.

Werner, G. D. A., J. E. Strassmann, A. B. F. Ivens, D. J. P. Engelmoer, E. Verbruggen, D. C. Queller, R. Noë, N. C. Johnson, P. Hammerstein, and E. T. Kiers. 2014. Evolution of microbial markets. Proceedings of the National Academy of Sciences 111:1237–44.

Westoby, M. 1998. A leaf-height-seed (LHS) plant ecology strategy scheme. Plant and Soil 199:213–227.

Westoby, M., D. S. Falster, A. T. Moles, P. A. Vesk, and I. J. Wright. 2002. Plant ecological strategies: Some leading dimensions of variation between species. Annual Review of Ecology and Systematics 33:125–159.

Whipps, J. M., P. Hand, D. Pink, and G. D. Bending. 2008. Phyllosphere microbiology with special reference to diversity and plant genotype. Journal of Applied Microbiology 105:1744–1755.

Whitham, T. G., C. A. Gehring, L. J. Lamit, T. Wojtowicz, L. M. Evans, A. R. Keith, and D. S. Smith. 2012. Community specificity: life and afterlife effects of genes. Trends in Plant Science 17:271–281.

Widder, S., R. J. Allen, T. Pfeiffer, T. P. Curtis, C. Wiuf, W. T. Sloan, O. X. Cordero, S. P. Brown, B. Momeni, W. Shou, H. Kettle, H. J. Flint, A. F. Haas, B. Laroche, and J. Kreft. 2016. Challenges in microbial ecology : building predictive understanding of community function and dynamics. The ISME Journal 10:2557–2568.

Wiebe, W. J., W. M. Sheldon, and L. R. Pomeroy. 1992. Bacterial growth in the cold: Evidence for an enhanced substrate requirement. Applied and Environmental Microbiology 58:359–364.

Williams, M. 2001. Problems of Knowledge: A Critical Introduction to Epistemology. Oxford University Press, New York, USA.

Wolinska, J., and K. C. King. 2009. Environment can alter selection in host-parasite interactions. Trends in Parasitology 25:236–244.

Wright, I. J., P. B. Reich, M. Westoby, D. D. Ackerly, Z. Baruch, F. Bongers, J. Cavender-bares, T. Chapin, J. H. C. Cornelissen, M. Diemer, J. Flexas, E. Garnier, P. K. Groom, and J. Gulias. 2004. The worldwide leaf economics spectrum. Nature 428:821–827.

Wright, S. J., K. Kitajima, N. J. B. Kraft, P. B. Reich, I. J. Wright, D. E. Bunker, R. Condit, J. W. Dalling, S. J. Davies, S. Diaz, B. M. J. Engelbrecht, K. E. Harms, S. P. Hubbell, C. O. Marks, M. C. Ruiz-Jaen, C. M. Salvador, and A. E. Zanne. 2010. Functional traits and the growth-mortality trade-off in tropical trees. Ecology 91:3664–3674.

Yadav, R. K. P., K. Karamanoli, and D. Vokou. 2005. Bacterial colonization of the phyllosphere of mediterranean perennial species as influenced by leaf structural and chemical features. Microbial Ecology 50:185–196.

Yadav, R. K. P., E. M. Papatheodorou, K. Karamanoli, H. I. A. Constantinidou, and D. Vokou. 2008. Abundance and diversity of the phyllosphere bacterial communities of Mediterranean perennial plants that differ in leaf chemistry. Chemoecology 18:217–226.

Yang, C. H., D. E. Crowley, J. Borneman, and N. T. Keen. 2001. Microbial phyllosphere populations are more complex than previously realized. Proceedings of the National Academy of Sciences 98:3889–3894.

Yeoh, Y. K., P. G. Dennis, C. Paungfoo-Lonhienne, L. Weber, R. Brackin, M. A. Ragan, S. Schmidt, and P. Hugenholtz. 2017. Evolutionary conservation of a core root microbiome across plant phyla along a tropical soil chronosequence. Nature Communications 8:215.

Yilmaz, P., and E. Al. 2011. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nature Biotechnology 29:415–420.

Zangerl, A. A. R., and M. R. Berenbaum. 2003. Phenotype matching in wild parsnip and parsnip webworms : causes and consequences. Evolution 57:806–815.

Zanne, A. E., D. C. Tank, W. K. Cornwell, J. M. Eastman, S. A. Smith, R. G. FitzJohn, D. J. McGlinn, B. C. O'Meara, A. T. Moles, P. B. Reich, D. L. Royer, D. E. Soltis, P. F. Stevens, M. Westoby, I. J. Wright, L. Aarssen, R. I. Bertin, A. Calaminus, R. Govaerts, F. Hemmings, M. R. Leishman, J. Oleksyn, P. S. Soltis, N. G.

Swenson, L. Warman, and J. M. Beaulieu. 2014. Three keys to the radiation of angiosperms into freezing environments. Nature 506:89–92.

Zhong, F., M. Xu, P. Metz, P. Ghosh-Dastidar, and J. Zhu. 2018. A quantitative metabolomics study of bacterial metabolites in different domains. Analytica Chimica Acta 18:237–244.

Zhulin, I. B. 2015. Databases for microbiologists. Journal of Bacteriology 197:2458–2467.