THE UNIVERSITY OF CHICAGO

USING GENOMICALLY INFERRED CO-OCCURRENCE AND METABOLIC NETWORKS TO CHARACTERIZE THE BUILT ENVIRONMENT MICROBIOME

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

AND

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN BIOPHYSICAL SCIENCES

BY

CESAR EMILIO CARDONA URIBE

CHICAGO, ILLINOIS

AUGUST 2019

Dedicated to Sarah, for her unconditional support and infinite love.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

For many years, our scientific knowledge about microbes was limited to those few species that can be cultured in the lab. With the advent of high throughput sequencing methods, scientists started exploring the genomic information of microbial communities, obtaining a list of microbial identities and overall potential genes present in any ecosystem. Microbial identities are determined by sequencing a marker gene of interest (amplicon) and grouping sequences into Organizational Taxonomic Units (OTUs) where each OTU is believed to come from the same species or genera. The gene potential is established by sequencing all genomes (shotgun sequencing) present in the ecosystem. Naturally, identifying what microbial species and genes are present is just the beginning as scientists seek to understand the interactions among those species and how those genes are expressed in fulfilling biochemical exchanges.

Recently, microbial ecologists have begun using microbial association network inference to describe the non-random organization of microbial communities and the emerging ecosystem dynamics. Network inference uses the statistical and topological properties of a set of entities (nodes) and the interconnections between them (edges) to infer the nature and strength of interactions. Two approaches that use network interference at their core are microbial co-occurrence networks and metabolic network models. The first uses high-level data to examine shifts in microbial communities across environmental variables in order to understand the processes driving microbial community structure and dynamics. The latter uses a mechanistic understanding of biochemistry and cellular growth to generate predictive models and explore the potential metabolite exchanges among species.

However, integrating metabolic models into microbial ecology workflows is happening slowly primarily because the models require fully annotated genomes as input data, which is generally

only obtained from the most expensive sequencing technologies (shotgun). Here, I propose a new method (probabilistic OTU modeling) to generate metabolic models from inferred full genomes using amplicon sequences as data inputs. The methodology relies on mapping the amplicon sequence to a reference database and identifying a pangenome that is phylogenetically related to that sequence. Then, the superset of genes present in the pangenome is used to build the corresponding metabolic model. Community models can be created as a simple aggregation of individual metabolic models.

In this study, I created amplicon-based co-occurrence networks to characterize the dolphins and environmental microbiomes at the Shedd Aquarium's oceanarium habitat. I explored how each microbiome potentially influences other microbiomes and how the addition of probiotics to the dolphin's diet can affect their skin, chuff and gut microbiomes. In a different built environment study, I similarly applied amplicon and metabolite-based co-occurrence networks to study the microbe-metabolite succession on common construction woods subject to high relative humidity. I applied the new probabilistic OTU modeling method to generate metabolic models representing each type of wood and identified biochemical clues that connect microbial taxonomies and signature metabolites to the various types of woods. These findings provide evidence of the usefulness of amplicon co-occurrence analysis combined with community metabolic models to offer concrete evidence of how ecological relationships are established within the built environment.

CHAPTER 1

INTRODUCTION

Work Published at Current Opinions of Microbiology (Cardona and Weisenhorn *et al.*, 2016)

Microbes (or microorganisms) are minute living things: archaea, bacteria, fungi, protozoa, microscopic algae or virus that could not be seen with the naked eye and that for many years were only known in mainstream culture as disease agents for infections such as salmonella or smallpox. Today, we know that microbes inhabit almost every space of our planet and only a small portion of them are disease-causing. We have learned that microbes play a fundamental role in Earth's history; they gave rise to life itself billions of years ago and still today keep our planet alive. Ocean, air, and soil microbes recycle and regulate the chemical elements essential for healthy ecosystems and wildlife. They inhabit all plants and animals, including humans. With a similar number of host and bacterial cells in humans, about 10^{13}, (Sender *et al.*, 2016) microbes have shown to be essential for human health. They even help train our immune system and digest our food (Kim, 2018; LeBlanc *et al.*, 2011).

For many years, our scientific knowledge about microbes was limited to those few species that could be cultured in the lab. With the advent of high throughput sequencing methods, we have been able to explore the genomic information of microbial communities, obtaining a list of identities and potential genes present in any ecosystem. Naturally, identifying what microbial species and genes are present is just the beginning, as scientists seek to understand the interactions among those species, and how those genes are expressed to fulfill biochemical exchanges. These biochemical processes influence key aspects of microbial community dynamics and stability, as well as the emergent properties of these communities. Microbial

communities in nature tend to include tens of thousands of species coexisting and interacting, with millions of metabolic and signaling relationships (Curtis *et al.*, 2002). To deal with this complexity, algorithms have been developed to capture and describe microbial ecological dynamics, with the ultimate goal of computationally describing the structure and dynamics of microbial communities. Computational description of microbial communities would enable understanding of which factors drive community function, stability, and resilience in response to perturbations, leading to the development of approaches to engineer and manage these crucial communities.

As the data available to microbial ecologists grows, so does our ability to understand and predict the dynamics of microbial populations and communities. Two popular approaches to achieving this goal are microbial co-occurrence networks and metabolic network modeling. The first uses high-level data to examine patterns and shifts in microbial communities across environmental and other variables in order to understand the processes driving microbial community structure and dynamics. The latter uses a mechanistic understanding of cellular biochemistry and organism growth to generate predictive models of shared metabolites and explore the potential interactions among species. The ability to model and predict microbial metabolism and activity enables a range of significant scientific improvements, from developing novel biosynthetic materials to enhancing agricultural production, to producing more efficient water treatment plants or improving human health. At the most basic level, modeling of microbial systems has the potential to augment our understanding of ecology and geochemistry and how different microbial species chemically regulate life on Earth.

As sequencing and related technologies have developed, microbial ecologists have performed many surveys to better understand the distribution and diversity of microbial taxa. These studies have revealed a vast diversity of taxa unknown from culturing work and demonstrated that microorganisms display distinct biogeographical patterns that are apparently different from those shown by plant and animal taxa (Horner-Devine *et al.*, 2007). These inventory studies have used different sequencing methodologies, sequencing the full genome (shotgun) or only a marker gene of interest (amplicon), from which this second method gave rise to the concept of OTU (Organizational Taxonomic Unit). An OTU is a representative DNA sequence that is believed to come from the same microorganism and the identity is inferred by database matching. However, understanding how microbial community structure influences ecosystem level functions requires more than a superficial understanding of such patterns. Instead, it is necessary to look closer at the data to see patterns in which species tend to co-occur and what conditions apparently favor or disfavor such co-occurrence. These patterns in co-occurrence yield insight in two ways. Firstly, species co-occurrence patterns provide observations on the ecological relationships among species, allowing generation of testable hypotheses about the nature of these relationships. Secondly, they provide insight into the environmental conditions that may shape community structure.

Network analysis uses the statistical and topological properties of a set of entities (nodes) and the interconnections between them (edges) to infer the nature, extent, and strength of interactions. Network analysis can describe and explore the connections between entities, whether those entities are individuals in social networks (Barabasi *et al.*, 2002), computers on the internet (Albert *et al.*, 1999), species in ecological webs (Sander *et al.*, 2015), or gene networks (Meyer

et al., 2008). Recently, scientists have begun using microbial association network inference to describe the non-random organization of microbial communities and to model community assembly and ecosystem dynamics.

1.1 Microbial co-occurrence networks

Microbial co-occurrence network inference algorithms describe the frequency of taxa association patterns among samples, thereby identifying putative causal interactions among species from observed abundance data. In this way, they help characterize and visualize community structure and suggest possible ecological relationships (e.g. cooperation, pathogenesis) among community members. Frequently, network reconstruction uses taxa abundance data from multiple, individual samples collected at different sites, times, seasons, or environmental conditions to generate a pairwise similarity matrix, using a suitable similarity or distance measure. Identified patterns are compared against a randomized abundance data set to estimate the strength and significance of the predicted relationships. Taxon pairs with p-values below the threshold are visualized as a network, where nodes represent taxa and edges represent the significant relationships between them and the edge thickness can reflect the strength of the relationship. Predicted relationships can be generated in multiple ways, including correlation, local similarity analysis, Bayesian networks, and mutual information criteria, among others.

1.1.1 Correlation and regression-based network inference

A popular and straightforward approach is to use pairwise similarity in abundance (e.g. Pearson, Spearman, Bray Curtis), or incidence (e.g. Jaccard, hypergeometric distribution), across samples to infer a relationship between two species. Multiple regression analysis can expand on this,

detecting relationships involving more than two taxa. To reduce overfitting, sparse multiple regression is usually carried out; that is, a subset of source taxa that best predicts the target taxon abundance is selected. In this way taxon abundance can be explained as a linear combination of the abundance of a subset of the remaining taxa (Fauset and Raes, 2012). These approaches are particularly suited to the identification of species whose role is greater than their relative abundance would suggest, i.e. keystone species. In a study of grassland and savanna soil from Brazil, Lupatini and colleagues (2012) were able to identify *Proteobacteria* and *Actinobacteria* genera which played a role as keystone species through their role as connector species. These approaches can also be applied to look at the patterns in the distribution of species across habitat types and between generalist and specialist ecological categories (Barberán *et al.*, 2012).

While this approach is simple to understand and apply, it has a couple of drawbacks. First, it is generally underpowered as there are usually many more OTUs than there are samples. A second drawback to applying these correlation methods is that most OTU tables are normalized to the total sum of counts per sample; therefore, they report relative abundance, and taxa relative abundance are not independent of one another. Permutation and bootstrap methods, such as the ReBoot method (Faust *et al.*, 2012) and CCREPE (Schwager *et al.*, in prep) have been developed to account for this lack of independence. The CCREPE method also introduces a new dissimilarity measure, the N-dimensional checkerboard score (NC-score), which is a measure of the randomness of species co-occurrence. CCREPE was applied to the Human Microbiome Project dataset (Consortium, 2012) and found that dominant commensal taxa were frequent competitors while pathogens were more likely to co-occur in complementary niches. Another tool, SparCC, estimates the basis abundance, a theoretical measure of abundance that is not

relativized, in order to directly examine for correlations among taxa from compositional data (Friedman and Alm, 2012). Bakker and colleagues (2014) used SparCC to examine the diffuse effects of plant diversity in structuring plant-microbe and microbe-microbe interactions. A newer modification to this method, regularized estimation of the basis covariance based on compositional data (REBACCA), uses the basis abundance to find sparse solutions to a system with a deficient rank (Ban *et al.*, 2015). This method can be applied to transformed count or relative abundance data to examine pairwise correlations of OTUs across different experimental treatments (Ban *et al.*, 2015). Another recent correlation based method is correlation inference for compositional data through the least absolute shrinkage and selection operator, CCLasso (Fang *et al.*, 2015). By penalizing large regression coefficients, this method compensates for co-variation among predictors, thus enabling an automatic approach to appropriate variable selection and reducing the effects of 'too large' regression coefficients (Hastie *et al.*, 2001). However, many other approaches to examining microbial associations have been developed that avoid these issues altogether.

1.1.2 Graphical model inference

SPIEC-EASI is a statistical method for the inference of microbial ecological networks that addresses both of the weaknesses of classic correlation analysis: interdependence of taxon relative abundance values and the greater number of OTUs than samples in most experiments. SPIEC-EASI addresses interdependence through a center-log transformation of the relative abundance data and then estimates the interaction graph using neighborhood selection (Meinshausen and Bühlmann, 2006) or sparse inverse covariance selection (Friedman *et al.*,

2008) algorithms. An appropriate model sparseness is inferred through random subsampling and the final ecological network and regularized covariance matrix are produced. SPIEC-EASI inferred networks from the American Gut project (AGP) datasets predicted both known core subnetworks and previously unknown microbial associations (Kurtz *et al.*, 2015).

1.1.3 Local similarity analysis (LSA)

Local similarity analysis can be used to examine microbial community dynamics in response to environmental conditions. This method determines a partially directed association network by comparing shifts in OTU composition through time series data. It can be applied to identify shifts in the abundance of a target OTU in response to a change in the composition of another OTU (or set of OTUs) or an environmental condition. This approach has been applied to demonstrate the relationships among bacterioplankton species and environmental factors (Ruan *et al.*, 2006), within and between microorganisms from different domains (Steele *et al.*, 2011), the effects of seasonal variability in environmental conditions such as NO3, salinity, and chlorophyll (Steele *et al.*, 2011), and to identify the temperature and micronutrient composition that defines the boundaries of co-occurring *Synechococcus* ecotypes (Sohm *et al.*, 2016). Recently, methods have been developed to overcome two of LSA's biggest challenges: eLSA addresses the limitations of computational efficiency (Xia *et al.*, 2013) while FASTLSA reduces the normality assumptions underlying the original LSA algorithm (Durno *et al.*, 2013).

1.1.4 Bayesian networks

Bayesian inference algorithms are particularly useful at illuminating interactions within complex microbial communities as they capture the conditional interdependence between OTUs and

describe complex stochastic assembly processes. These models are able to capture diverse types of relationships: linear, non-linear, combinatorial, and others. This is possible because the approach compares the joint multivariate probability distributions of multiple OTUs concurrently. Bayesian network modeling has been widely used to study microbial gene networks (Friedman *et al.*, 2000; Schäfer and Korbinian, 2005) and ecological networks (Milns *et al.*, 2010; Aderhold *et al.*, 2012). Recently, Shafiei and colleagues (2015) presented a supervised Bayesian model (BioMiCo) to infer the microbial community structure from compositional data. Using BioMiCo, they could effectively predict both body site (hand, mouth, and gut) and host individual across multiple time points. When OTU data is paired with metagenomics data, a Bayesian and neural network approach called predicted relative metabolic turnover (PRMT) can be applied to relate both the genes and specific OTUs through time and environmental conditions, providing insight on changes in species interactions while linking these changes to ecosystem level processes, such as carbon dioxide production (Larsen *et al.*, 2012).

1.1.5 Mutual information

Mutual information approaches are useful for identifying non-random co-association patterns. CoNet is an ensemble based network reconstruction method that detects non-random patterns of microbial co-occurrences between OTUS by combining multiple association methodologies simultaneously and merging the results into an inferred network structure. CoNet has the option to use correlation, incidence, similarity or mutual information methods to define the associations between pairs of OTU. Mutual information methods can be implemented with the MINET R

package (Meyer *et al.*, 2008). Application of CoNet to the interactome in the marine photic zone showed the strength of interactions among grazers, primary producers, viruses, and symbionts relative to environmental factors (Lima-Mendez *et al.*, 2015). CoNet has also been used to show variation in the specificity of phage-bacteria interactions in coral disease, suggesting a potential role for phages in controlling stress-associated bacteria (Soffer *et al.*, 2015). Faust and colleagues (2015) used CoNet to compare the network structure of 20 different biomes: 7 environmental and 13 host-associated and observed that soil microbial networks contain proportionally fewer positive associations and are less densely interconnected than host-associated networks.

The maximal information coefficient (MIC) is a measure of the strength of linear or nonlinear associations between variables via mutual information. This approach is a nonparametric, exploratory statistical approach to identify novel interactions from a large dataset. It uses binning to apply mutual information to continuous distributions, and is roughly equal to the coefficient of determination (R2) produced by regression analysis methods. Recently, MIC methods were applied to community data from sediments of the Rhône River (Fagervold *et al.*, 2014). This study found that riverine organic matter inputs were the main drivers of community composition, although species interactions were also significant. A study by Wang and colleagues (2014) used MIC to identify potential interactions between bacteria, eukaryotes, and pathogens in drinking water that suggest manipulating the microbiome may lead to new avenues for controlling opportunistic pathogen growth.

1.1.6 Strengths and limitations of co-occurrence networks

Microbial interaction networks have proven useful for identifying potential interactions among species within complex communities and for providing insight into the key factors structuring a community within a set of samples. Consistency among samples can indicate the relative stability of a community and its interactions; although it is possible to go a step further and use microbial co-occurrence networks to test the effects of individual species loss. For example, the extent of network fragmentation after single node deletion can be used as a measure of robustness and has been examined in fluvial ecosystems (Widder *et al.*, 2014), lakes and oceans (Peura *et al.*, 2015), and the human gut (Coyte *et al.*, 2015). However, such extrapolation should be carefully interpreted without first identifying the mechanism underlying the identified interactions as pattern identification, like correlation does not equal causation.

Indeed, microbial co-occurrence networks can be misleading if other factors such as habitat filtering are causing non-random patterns in abundance of multiple taxa (Berry and Widder, 2014). In the case of habitat filtering, many species may be responding to different, co-varying gradients. Further, multiple species may be affected, independently, by a single 'hub' or keystone species leading to the identification of spurious relationships (Berry and Widder, 2014). Although microbial co-occurrence networks are useful for identifying patterns in community structure and providing some insight into the relative stability of communities, they require careful and conservative interpretation. Without an understanding of the underlying mechanism driving the interactions uncovered, microbial co-occurrence networks are of limited utility in understanding community dynamics.

1.2 Metabolic networks and metabolic models

Unlike microbial co-occurrence networks, metabolic networks can provide a detailed mechanistic understanding of the interactions between species. However, in order to do this, these models require more information than microbial co-occurrence networks and are dependent upon having a nearly complete, annotated genome of the species of interest. Because genome annotation is an arduous process, the construction of most metabolic network models relies on automated annotations, such as those provided by RAST (Aziz *et al.*, 2008). Such services typically provide a homology-based computational inference of the set of enzymatic genes present in a given genome. This process relies heavily on cross-species metabolic databases such as KEGG (Kanehisa *et al.*, 2006) and updated gene annotation databases (Huss *et al.*, 2008). Once a genome has been annotated, a metabolic reconstruction can be made which links annotated genes to metabolic-related enzymatic functions and the derived set of potential biochemical reactions. The topology of this reconstruction can be examined directly, or it can be used to generate a metabolic model, a mathematical representation of the organism metabolism, and run constraint-based optimizations to generate predictions of organismal behavior.

It has been observed that some nodes within a metabolic reconstruction have greater connectivity than others, indicating hubs or specialized subnetworks within metabolism that may be directly linked with the organismal phenotype (Jeong *et al.*, 2000). Indeed, *Escherichia coli* phenotypes have been predicted from topology-based measures of network transporters (Wunderlich and Mirny, 2006) and the genetic robustness of bacteria has been explored through the topological structure of metabolic reconstructions (e.g. Freilich *et al.*, 2010). These approaches can be

extended to examine the global topology of a community of interacting organisms, and to identify which organism is responsible for the production of key metabolites and identify which metabolites are exchanged among subsets of organisms in the community (Levy and Borenstein, 2012; Levy and Borenstein, 2013). Identification of these metabolic interactions and dependencies provides a mechanistic basis for understanding species' interactions and predicting shifts in species under novel conditions.

Beyond network topology, constraint-based metabolic modeling can be used to examine the biochemical activity within the cell as well as at the community level between individuals of different species. Constraint-based modeling is based on the assumption that metabolic transitions are more rapid than either growth or environmental changes; thus, cells are assumed to be in a quasi-steady state allowing for manageable mathematical solutions (Varma and Palsson, 1994). Constraint-based modeling takes into account the stoichiometry of reactions, and flux balance analysis can be applied to determine the flux (activity) through active metabolic pathways under a specified set of conditions. Flux variability analysis can be used to determine what metabolites can be produced, consumed or excreted as well as which reactions are essential for the production of biomass. Studies have shown that, given evolutionary pressure to optimize growth on the modeled substrate, constraint-based analysis can adequately predict microbial growth (Ibarra *et al.*, 2002; Edwards *et al.*, 2001). Individual metabolic models can be combined into compartmentalized community models to investigate metabolic cross-feeding. These models take into account the relative abundance of each microbial population to generate a model of the biochemical reactions occurring in that community. Application of dynamic flux balance analysis approaches, such as COMET, can identify mechanistic relationships among species and

also predict shifts in microbial populations in response to environmental perturbations and species interactions (Harcombe *et al.*, 2014).

1.3 Combining co-occurrence and metabolic networks

Microbial co-occurrence network representations allow the examination of the underlying organizational structure of a microbial community, the identification of keystone microbes (microbes that if removed, would destabilize the organizational structure of the network) in the network topology, and the identification of associations with environmental conditions that influence community assembly. Community metabolic networks can provide a mechanistic link between species through the identification of metabolite exchanges and species-specific resource requirements. When used together, co-occurrence and metabolic networks can provide a more in-depth view of the hidden rules that govern the stability and dynamics of microbial communities. The co-occurrence network could suggest interesting ecological relationships that you might want to investigate further with metabolic networks, and together curate the true ecological interconnection network of the microbial community.

CHAPTER 2

CO-OCCURRENCE NETWORKS TO CHARACTERIZE THE SHEDD AQUARIUM BUILT ENVIRONMENT MICROBIOME

Work Published at mSystems Journal (Cardona *et al.*, 2018)

2.1 Overview

Host-associated microbial dynamics are influenced by dietary and immune factors, but how exogenous microbial exposure shapes host-microbial dynamics remains poorly characterized. To investigate this phenomenon, we characterized the skin, rectum, and respiratory-associated microbiota in four aquarium-housed dolphins, daily over a period of 6 weeks, including administration of a probiotic during weeks 4-6. The environmental bacterial sources were also characterized, including the animals' human handlers, the aquarium air and water, and the dolphins' food supply.

Continuous microbial exposure occurred between all sites, yet each environment maintained a characteristic microbiota, suggesting that the majority of exposure events do not result in colonization. Small changes in water physicochemistry had a significant but weak correlation with the change in dolphin-associated bacterial richness and had no influence on phylogenetic diversity. Food and air microbiota were the richest and had the largest conditional influence on other microbiota in the absence of probiotics, but during probiotic administration food alone, had the largest influence on the stability of the dolphin microbiota.

Our results suggest that respiratory and gastrointestinal epithelial interaction with air and food-associated microbes have the biggest influence on host-microbial dynamics, while other

interactions, such as skin transmission, played only a minor role. Finally, direct oral stimulation with a foreign exogenous microbial source can have a profound effect on microbial stability.

These results provide valuable insights into the ecological influence of exogenous microbial exposure, as well as laying the foundation for improving aquarium management practices. By comparing dolphins between aquaria that use natural versus artificial seawater, we demonstrated the potential influence of aquarium water disinfection procedures on dolphin microbial dynamics.

2.2 Background

The relationship between a host and its resident microbiome has been implicated in health, with the microbiota providing benefits to the host through innate immunity, nutrition, and metabolism (Gilbert *et al.*, 2016). The microbiome of each individual host is significantly different, while the microbial composition of an individual host is remarkably stable over time (Costello *et al.*, 2009). However, despite compositional stability, the relative proportion of each microorganism (what we commonly refer to as community structure) in each host is dynamic. The factors that drive these changes have been identified as diet and disease (Bäckhed *et al.*, 2015; Brooks *et al.*, 2016; David *et al.*, 2014a; Dominguez-Bello *et al.*, 2016; Gilbert *et al.*, 2016; Raveh-Sadka *et al.*, 2015). However, the impact of microbial exposure on host-health has been well characterized with respect to disease-causing pathogens, and there is increasing evidence that microbial exposure influences host-health through direct immunological stimulation (Stein *et al.*, 2016). However, while studies have examined the interaction between human microbiota and environmental microbiota, it is virtually impossible to characterize all known sources of

exogenous microbes in a population (Lax *et al.*, 2014, 2017), and the impact of dietary changes on microbial dynamics is very difficult to control for (Caporaso *et al.*, 2011; David *et al.*, 2014a, 2014b). As such, the influence of environmental microbiota on the dynamics of host-associated microbiota remains largely unknown.

While the interaction between a human and their environmental and lifestyle-derived microbes is hard to control in longitudinal investigations, animals in managed systems have only a limited number of sources of exogenous microbes and are often provisioned with a highly stable diet (Hyde *et al.*, 2016). The sources for aquarium-housed marine mammals (such as dolphins) are essentially limited to water, food, human handlers, and air; additionally, they have a very stable diet. These animals, therefore, represent a useful model system in which to examine the influence of exogenous microbial exposure on the dynamics of host-associated microbial communities (Apprill, 2017). We hypothesized that the dolphin microbiota would exhibit an equilibrium with the regular environmental microbial exposure, and that if a foreign microbial exposure was administered, this would disturb host-associated microbial dynamics.

Dolphin-microbial interaction has been studied with respect to pathogen surveillance (Jaing *et al.*, 2015; Venn-Watson *et al.*, 2008), identification of potential probiotic strains (Diaz *et al.*, 2013), identification of novel taxa (Bik *et al.*, 2016; Harper *et al.*, 2002), and characterization of variation in the dolphin microbiota across body sites (Avalos-Téllez *et al.*; Bik *et al.*, 2016; Godoy-Vitorino *et al.*, 2017; Soverini *et al.*, 2016). In the wild, dolphins are exposed to a broad diversity of microbes in the water, their food, and physical interactions with other dolphins and animals. In an aquarium setting, this exposure is often markedly different. Importantly, aquarium

management practices include using stringent disinfection procedures with the intention of minimizing resident animal exposures to potential pathogenic microbes (Department of Agriculture). We aimed to use aquarium-housed dolphins to determine the influence of host-environment interactions on the stability of the dolphin-associated microbiota. We sampled and characterized the microbiota of four dolphins in an intensively managed aquarium setting daily over six weeks at different body sites, including skin, rectum, and respiratory exhalant (chuff). We also concurrently characterized the microbiota of the hands and noses of the animals' human handlers, the oceanarium air and water, as well as the dolphins' food supply. In addition, we examined the impact of a direct microbial administration intervention in week three, with dolphins randomized to receive either a single organism or multi-organism probiotic for three weeks. As with previous investigations (Lax *et al.*, 2014, 2017), we aimed to quantify the degree of interaction between the environment and the host-microbiome to understand whether this interaction correlated with the perceived stability of the microbial community of each animal. Understanding the sources of bacteria and archaea that can temporarily or permanently alter a host's microbiota could enable new strategies to improve healthcare for managed dolphin populations and further serve as a potential model for other animals housed in aquatic systems.

2.3 Results

Samples were collected over 42 consecutive days (09/24/2014 to 11/04/2014) in the Shedd Aquarium oceanarium, indoor temperature controlled exhibit that uses recirculating artificial seawater. The corpus of 4 Pacific White-sided dolphins (D1-4) were sampled, including skin (peri-umbilicus), rectum, and respiratory tract (forceful exhalation referred to as chuff). The

hands and noses of the animal's human handlers, oceanarium air, oceanarium water, and the dolphin's food (fish and squid blend) were also sampled. A total of 2,370 samples were processed (1,084 of these analyzed twice as technical replicates, producing a total of 1,286 pooled samples) using 16S rRNA V4 amplicon sequencing. After quality control (see Materials and methods), including rarefaction of the samples down to 5,000 reads, the final dataset comprised 1,214 samples; these 5,929,516 reads clustered with 97% similarity into 19,536 Operational Taxonomic Units (OTUs). The dominant taxonomic group in each environment sampled is shown in Figure 1 A) at the phylum level and B) family level.

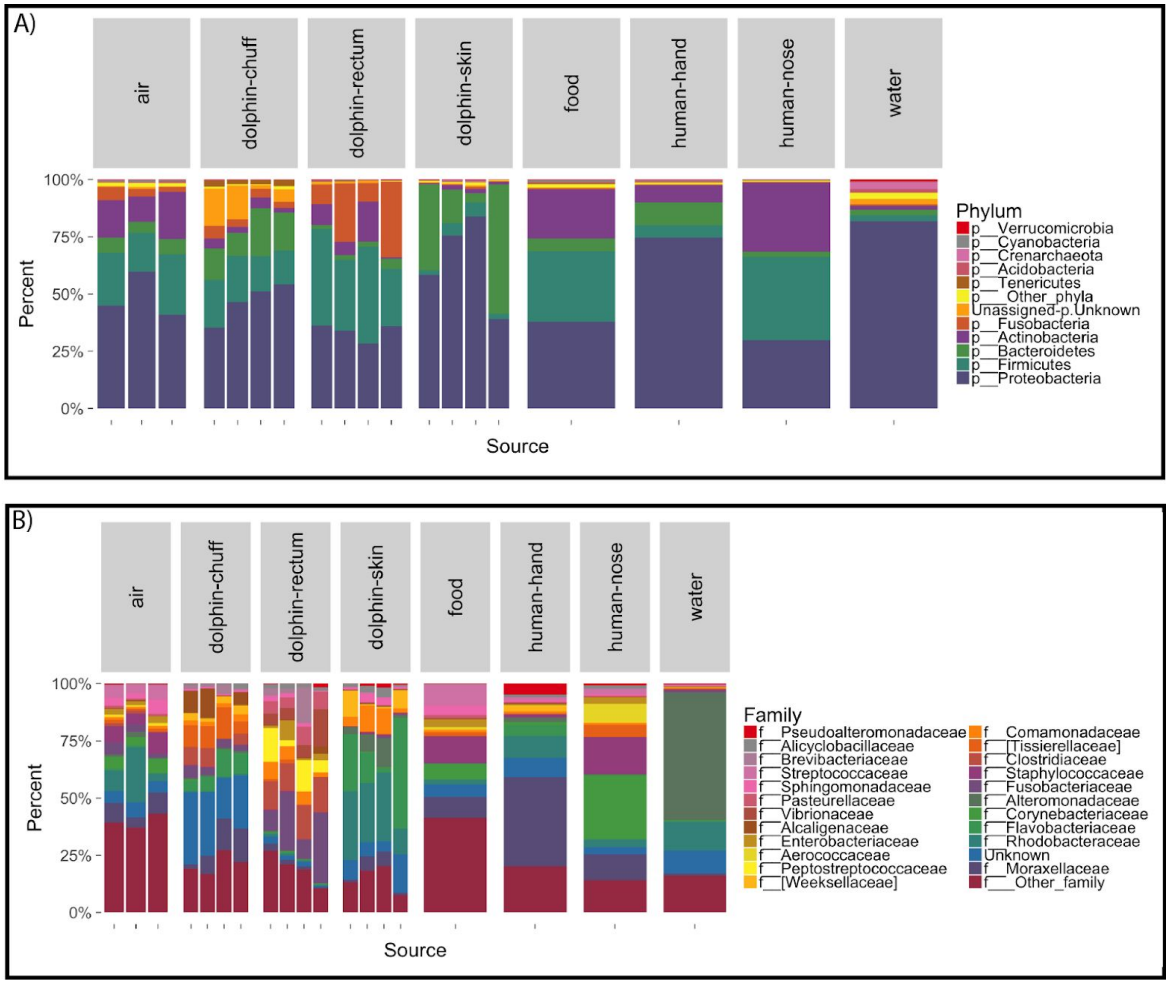

Figure 1: Taxonomical distribution of Shedd Aquarium oceanarium microbiome samples grouped by A) Phylum and B) Family level. A) Percent of reads for all samples with more than

Figure 1 (continued): 5000 reads (n=1241) grouped by sample location. Phyla are sorted by the abundance, starting with the most abundant phyla in the bottom of the stack. High microbial diversity was observed across all sources, 25 different phyla are identified with an abundance of at least 0.01% reads (592 reads). All sites contained between 21 to 25 different phyla: air 25, dolphins, food and water 23 and human trainers 21. Only the 11 more abundant phyla and 'Other phyla' categories are shown in the figure. Observations from this data breakdown are: 1) the dolphin-rectum, harbors nearly as much Firmicutes and Fusobacteria as Proteobacteria, 2) food and human-nose, had comparable amounts of Firmicutes, Actinobacteria and Proteobacteria 3) dolphin chuff harbored a significant number of Unknown and Tenericutes phyla greater than any other source 4) water contains greater abundance of Crenarchaeota archaea and Verrucomicrobia phyla than other sources. B) Percent of reads for the top 23 most abundant families and 'Other families' group found in samples with 5000 or more reads (n=1241). Families are sorted by the abundance, starting with the most abundant families in the bottom of the stack. A total of 297 families were identified where the most abundant 23 families covered only 50 to 75% of the total reads, reflecting that less abundant families are still accounting for a large portion of the different microbial communities.

2.3.1 Shedd aquarium alpha and beta diversity observations

Alpha diversity (Faith's PD) was variable across sampled sites (Figure 2A), with food and air having significantly greater alpha diversity (pairwise comparisons using Tukey and Kramer (Nemenyi) test, P<=0.002) than all other sites. Overall, as observed in other studies (Bik *et al.*, 2016; Costello *et al.*, 2009; Lax *et al.*, 2014) the within site beta-diversity distances were significantly smaller than the between-site beta-diversity distances (ANOSIM R=0.62, P=0.001), which is exemplified by a weighted UniFrac heat map (diagonal vs. non-diagonal elements, Figure 2B) and NMDS visualization (Figure 2C). All sites cluster within their environment based on weighted UniFrac distances, except dolphin skin, whereby the skin of dolphins D1 and D4 had a significantly different microbial composition to the skin of D2 and D3 (ANOSIM; R=0.45, P=0.001).

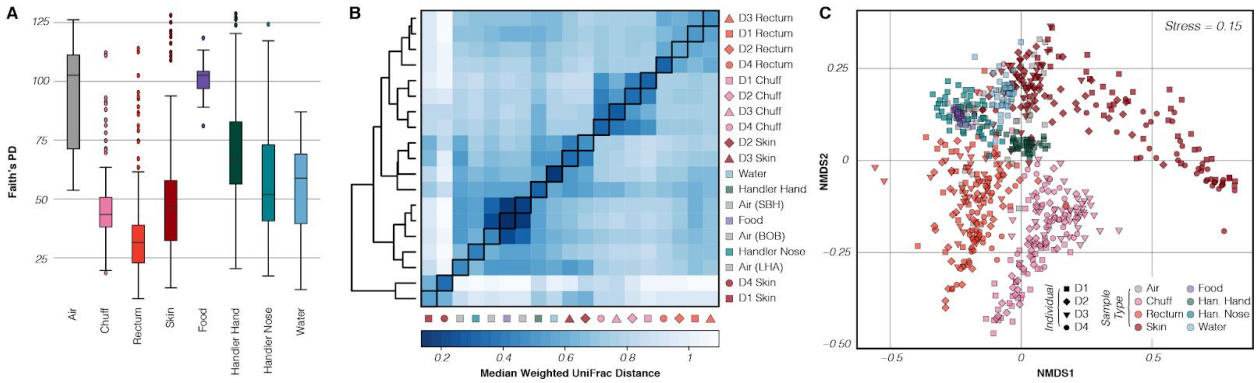

Figure 2: Overview of Shedd Aquarium microbial alpha and beta diversity data. A) variation of Faith's PD alpha diversity, food and air having significantly higher values than other sites and dolphin sites (chuff, rectum, and skin) having the lower diversity values. B) Weighted UniFrac Beta diversity hierarchical clustering heatmap shows how similar sites tend to cluster together. C) Weighted UniFrac NMDS plot of Shedd Aquarium samples, dolphin sites aggregate by sample site rather than per individual and they are clearly separated from each other. Non-dolphin samples, squared shape, are much more intertwined; with air samples showing high similarity to dolphin skin, human hand and nose, water, and food.

At three time points during the study (9/29, 10/14 and 11/04), more frequent sampling of the dolphin sites was performed (three to six times a day) to build a larger dataset focused on diurnal microbiome variations (Table S1 at Cardona *et al.*, 2018). This data showed that the rectum microbiome was the most dynamic microbiome across sampling intervals, with the largest difference for samples taken in the early morning around 8 am and the following sampling times from 10 am to 6 pm (Permutation T-test, $P<0.001$) with Pasteurellaceae, Peptostreptococcaceae, Fusobacteriaceae OTUs showing a sharp increase in abundance in the early morning, and a Brevibacteriaceae OTU showing a sharp decrease (DeSeq2 abs(fold2Change)>5 and $P<0.001$) in the same timeframe (Table S2 at Cardona *et al.*, 2018).

An inventory of all sampled sites: dolphin, food, air, water, and human-produced a total count of 15,581 (80.2%) OTUs shared between two sites or more, with 2,204 (11.3%) found in all sampled sites (Figure 2B). Dolphin-associated sites (skin, rectum, chuff) maintained the greatest

proportion of unique OTUs, with 2,543 (13.1%) found only in dolphin samples. In contrast, and perhaps surprisingly, food (which was a stable composite of fish and invertebrates) only had a single OTU that was not shared with another environment. This suggests that while each environment maintains a core microbiota, likely driven by niche selection where the environment selects for specific taxa, there is a large number of OTUs overlapping between sites. Common OTUs could be the result of independent selection or OTU transfer between sites. Due to the well-mixed aquatic medium where all dolphins cohabitate, there is indeed ample opportunity for bacterial transfer.

To determine the potential influence of probiotics on the dolphin microbiome, the dolphin population was split into two groups of two, and a different probiotic was administered daily to each group beginning on day 19 of the study. Dolphins D1 and D2 (Group A) received a multi-species *Lactobacillus reuteri* dominated consortium, while dolphins D3 and D4 (Group B) received *Lactobacillus salivarius*. Probiotic A (*L. reuteri* combination) comprised 9 different OTUs, while Probiotic B (L. salivarius) was represented by a single organism. It is important to note that food samples never included the probiotic bacteria supplemented. To track the probiotic organisms in the dolphin population, the probiotic 16S rRNA OTUs annotated as *Lactobacillus* or *Bifidobacterium* were subject to oligotyping analysis (Eren *et al.*, 2013). Oligotyping uses entropy to identify unique 16S rRNA V4 marker sequences at sub-OTU resolution. Probiotic A comprised 11 oligotypes, four *Lactobacillus*, and seven *Bifidobacterium*; only three and four of these oligotypes, respectively, were present at ≥5% relative abundance. Probiotic B comprised three *Lactobacillus* oligotypes, one of which accounted for 97% of the reads. The abundance of these oligotypes was quantified in the dolphin rectum samples. The most abundant oligotypes in

Probiotic A (abundances of ≥5%) were absent or rare (<0.1% reads) prior to probiotic administration, and became significantly more abundant (reaching up to 1.2% or 4.4% of total reads for *Lactobacillus* and *Bifidobacterium* respectively) during administration only in dolphins D1 and D2 (Permutation T-test, P<=0.01; Figure 3A and Figure 3C). Similarly, the most abundant oligotype in Probiotic B was significantly more abundant following administration only in the dolphins to which it was administered, D3 and D4 (Permutation T-test, P<0.003; Figure 3B).

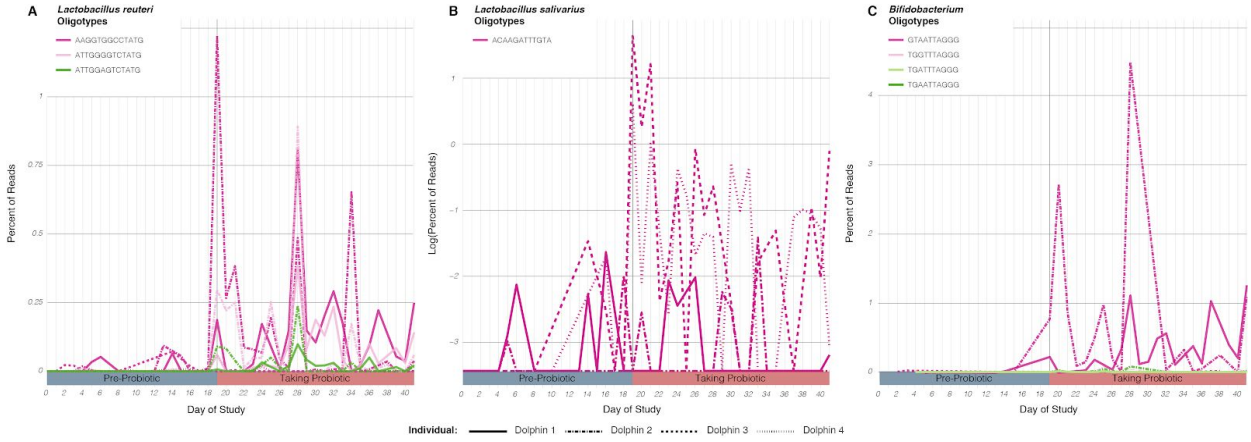

Figure 3: Daily percent of probiotics oligotypes reads found in the dolphin rectum samples. Probiotic A oligotypes significantly increases for D1 and D2 A) *Lactobacillus reuteri* and C) *Bifidobacterium*. Probiotic B oligotypes B) *Lactobacillus salivarius* (ratio in log scale) significantly increases for D3 and D4 while taking probiotics.

The probiotic administration was associated with a significant decrease in alpha diversity (Faith's PD) in the rectal-associated microbiome of Group A dolphins (before vs. while taking probiotics comparison; Permutation T-test, P=0.049), but not in other body sites nor in Group B dolphins (before vs. while taking probiotics comparison). However, the variance in alpha diversity (Faith's PD) between samples significantly decreased in both dolphin groups following

probiotic administration. This suggests that the microbial diversity of dolphin-associated sites became more stable following probiotic therapy (Table 1).

Location tested	Samples tested	PD Permutation T-Test, *P* value	PD F-Test, *P* value
Dolphin's chuff	Dolphins Group A only	0.168	0.000
Dolphin's rectum	Dolphins Group A only	0.049	0.000
Dolphin's skin	Dolphins Group A only	0.352	0.047
Dolphin's chuff	Dolphins Group B only	0.111	0.000
Dolphin's rectum	Dolphins Group B only	0.653	0.019
Dolphin's skin	Dolphins Group B only	0.279	0.047

Table 1: Comparison of dolphin samples alpha diversity before and while taking probiotics. Permutation T-test and variance F-test comparisons for changes in phylogenetic diversity (PD) on different locations and samples subgroups. Significant values (P values less than 0.05) highlighted with red color.

The alpha diversity for food and water was not significantly affected by whether the dolphins were being administered probiotics, suggesting that the differences observed in the dolphin rectal samples were an influence of dietary probiotics rather than environmental changes. In sites that have a reduced likelihood of being able to influence the dolphin microbiota, for example, air and human-hand samples, there was a significant change in alpha diversity and variance during probiotic administration. However, there is no indication that this was associated with probiotic administration to the dolphin diet. In fact, air-filters in the oceanarium were changed around the same time that probiotics treatment started, which might explain the sudden drop of air-associated microbial diversity (Table S3 at Cardona *et al.*, 2018)

While alpha diversity only changed for one animal group (Group A) at a single site (rectum) upon administration of the probiotic, microbial composition for the dolphin sites was significantly different prior to and during probiotic administration, and between the two dolphin probiotic groups (Groups A and B).

We used ANOSIM to compare whether distances between samples of the same date period (before and while taking probiotics) were significantly lower than the distance between samples of different date periods. Using ANOSIM with beta diversity Unweighted UniFrac distances the microbial composition of both periods across all dolphin sites significantly differed from each other with the probiotic administration. In contrast, using weighted UniFrac distances produced significant differences only for dolphin rectum and skin locations. Also, weighted UniFrac produced significant ANOSIM R discriminant values (0.081-0.122) were smaller than those produced by the unweighted UniFrac. The varying results of the two UniFrac distances suggest that observed shifts are predominantly due to changes in the proportion of less abundant taxa (~<0.01%) in the dolphin microbiota (Table 2). Also, we tested for microbial compositional differences between the dolphin groups (Group A vs. Group B) for both time ranges, before and while taking probiotics, revealing that during both periods the dolphin groups had significant differences in their microbial composition (ANOSIM R>0.03 P<0.028) with the exception of dolphin skin prior to probiotic administration. This variance in the dolphin microbiota underlies great individual differences per animal and suggests that our statistical power for investigating community structural shifts related to the probiotic administration were not sufficient with only two dolphins per group (Table S4 at Cardona *et al.*, 2018).

Location tested	Distance	Samples tested	Anosim R	Anosim *P* value
Dolphin's chuff	WU	Dolphins Group A only	0.007	0.275
Dolphin's rectum	WU	Dolphins Group A only	0.122	0.001
Dolphin's skin	WU	Dolphins Group A only	0.109	0.001
Dolphin's chuff	WU	Dolphins Group B only	0.010	0.271
Dolphin's rectum	WU	Dolphins Group B only	0.081	0.011
Dolphin's skin	WU	Dolphins Group B only	0.082	0.004
Dolphin's chuff	UWU	Dolphins Group A only	0.135	0.001
Dolphin's rectum	UWU	Dolphins Group A only	0.268	0.001
Dolphin's skin	UWU	Dolphins Group A only	0.273	0.001
Dolphin's chuff	UWU	Dolphins Group B only	0.215	0.001
Dolphin's rectum	UWU	Dolphins Group B only	0.185	0.001
Dolphin's skin	UWU	Dolphins Group B only	0.161	0.001

Table 2: ANOSIM statistics on differences between before and while taking probiotics for different dolphin sample location and different groups. Using weighted UniFrac distance Distance (WU) and using unweighted UniFrac distance (UWU). Significant values (P values less than 0.05) highlighted with red color.

NMDS ordination plots visualized changes in unweighted UniFrac beta diversity distances before and while probiotics were administered (Figure 4). The skin microbiome showed a significant separation of the samples before and after probiotics (Group A ANOSIM R=0.27 P=0.001, Group B ANOSIM R=0.16 P=0.001). The rectum microbiome also showed a separation of the samples before and after probiotics (Group A ANOSIM R=0.12 P=0.001, Group B ANOSIM R=0.27 P=0.001). Interestingly, Group A dolphin rectum also showed a

significant variation in dispersion (Permutation test for homogeneity, P=0.001), which could be driving the significant differences found with the ANOSIM test.

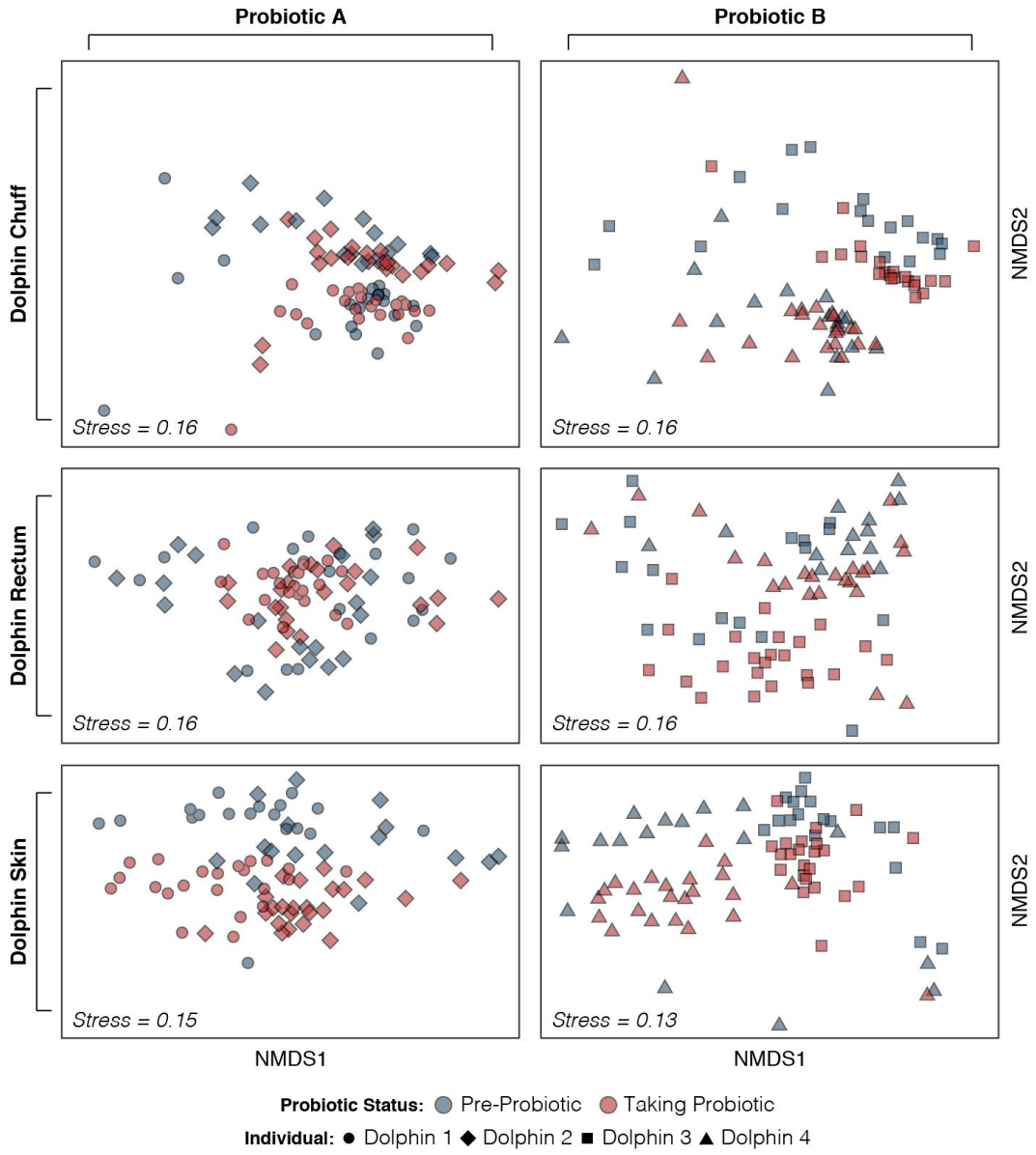

Figure 4: NMDS ordination presents changes in unweighted UniFrac beta diversity distances before and while taking probiotics for both dolphin groups and all dolphin sites.

To explore the stability in community composition over time within each site (dolphins, water, air, humans) we calculated pairwise UniFrac distances for each consecutive time pair belonging

to the same sample across the entire time series. These distances were used later to determine whether the day-to-day variance in beta diversity was significantly influenced by probiotic administration (Figure 5). Strikingly, all dolphin sites showed significant differences in average pairwise beta diversity between pre and post probiotic administration (except for dolphin rectum and skin in Group B), while all non-dolphin samples were not significantly different (except human nose $P<=0.007$). The pairwise unweighted UniFrac distances within most dolphin sites during probiotic administration were significantly smaller than prior to probiotic administration. This suggests that all dolphin microbial communities for Group A and chuff microbial communities for Group B became significantly more similar day-to-day during probiotic administration, suggesting that the probiotics stabilized community dynamics. There were no significant differences in any site when comparing the weighted UniFrac distance metrics, again suggesting that any stabilizing effect may come from changes in the composition of relatively low abundance bacterial taxa.

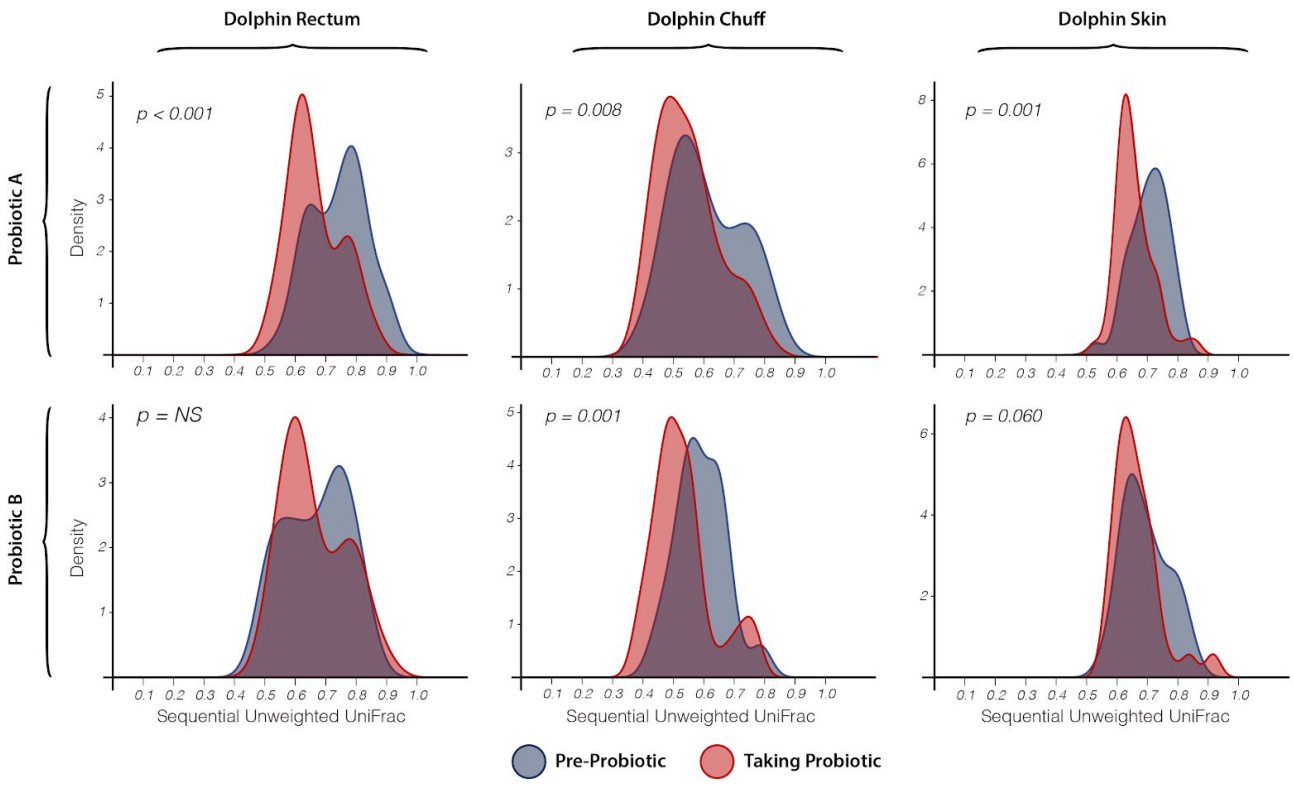

Figure 5: Pairwise UniFrac distances for each consecutive time pair from the same sample aggregated and smoothed to distribution functions before and while taking probiotics for different dolphin sites and two probiotic groups. Permutation T-tests confirm that dolphin chuff and skin are significantly different before and while taking probiotics, while dolphin rectum is significantly different only for Probiotic A.

2.3.2 Dolphin's rectum co-occurrence networks

Potential keystone OTUs and dense OTU modules that may correspond to distinct subsets of communities within the rectum microbiome were inferred by examining the topology of co-occurrence networks (Figure 4). A keystone node in a microbial co-occurrence network has been defined as one with 1) high degree (number of connections per node), 2) low betweenness centrality (number of shortest paths between any two nodes in the graph passing through that node), 3) high closeness centrality (average distance from this node to any other one), and 4) high transitivity (probability that adjacent nodes are connected) (Berry and Widder, 2014). By

this definition, two OTUs assigned to the genera *Kineococcus* and *Brevibacterium* in the phylum Actinobacteria (OTU ids 543684 and 206826, respectively) were identified as potential keystone taxa, with *Brevibacterium* interestingly being the genus that already showed a significant fluctuation of abundance between early morning samples and the rest of the day. These two OTUs were in the top 30% greatest values for the degree, closeness centrality, and transitivity values, as well as in the bottom 30% of values for betweenness centrality. Studies in the human gut environment have also suggested that Actinobacteria fit the definition of keystone taxa, since they are relatively rare, have a high degree of ecological connectedness, and are positively correlated with diversity both within and between different individuals (Trosvik and de Muinck, 2015). One potential explanation for this similarity is that certain dolphin gut-associated Actinobacteria may hold a similar niche to their comparative role in the human gut microbiota.

The dolphin rectal co-occurrence network produced a set of twelve modules. The OTUs from Probiotic A co-aggregate with 330 other OTUs to make Module 1, while the OTU associated with Probiotic B aggregated with only a single other OTU to make Module 11. Of the twelve inferred modules, five of them composed of 20 or more OTUs (modules 1, 2, 3, 6, and 7). Only Modules 6 and 7 had significantly differentially abundant OTU counts between before and during probiotic administration for both dolphin groups with Module 7 also including the two keystone OTU candidates identified earlier (OTU ids 543684 and 206826). Following probiotic administration, the abundance of Module 6 significantly increased, while Module 7 significantly decreased (permutation T-test, $P<=0.04$). The abundance of Module 1 also decreased but was only significantly different in group B dolphins (Permutation T-test, $P<0.001$). Module 11 was very sparse, with detectable values for only a few days in the study. Each module had its own

microbial taxonomic signature, for example, Module 1, in which Probiotic A clustered comprised 50% Gammaproteobacteria, while Module 11 with Probiotic B comprised 100% Bacilli (Figure 6).

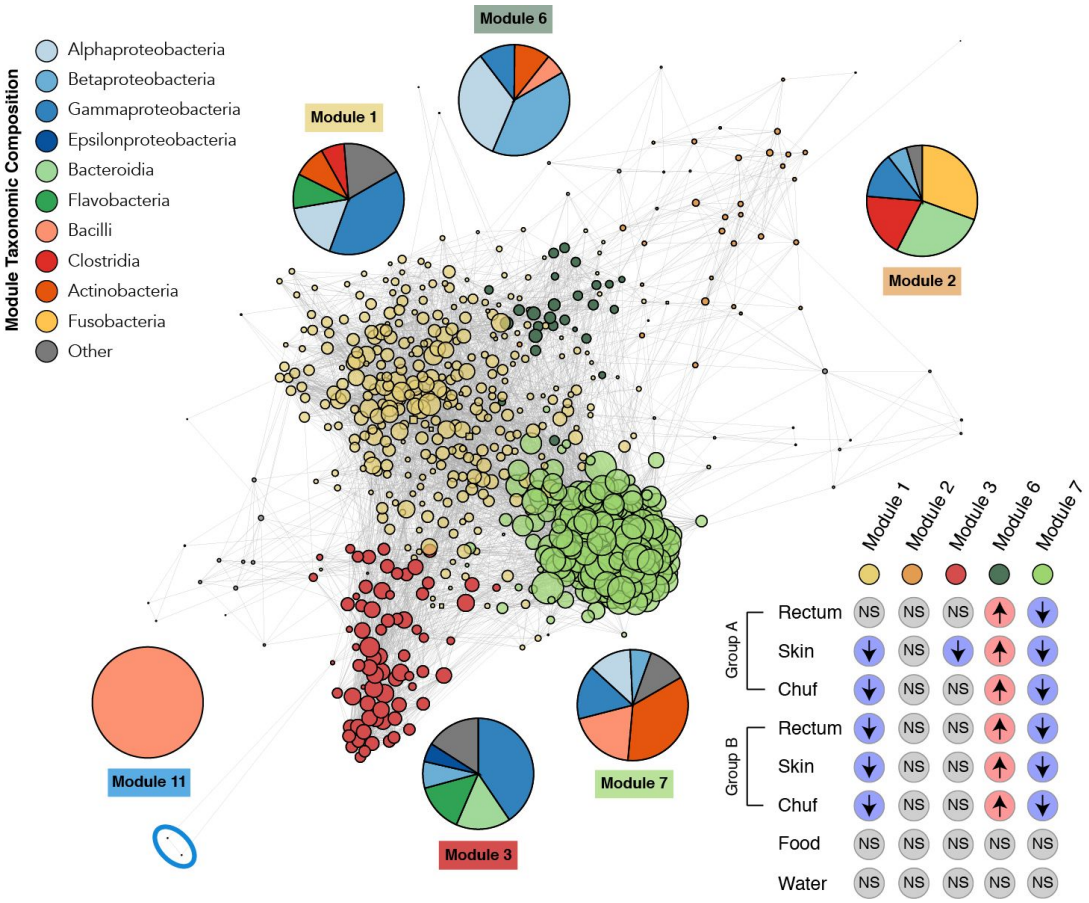

Figure 6: Co-occurrence network for dolphin rectum microbiome. OTUs are the solid colored circles (nodes), which are connected by lines (edges) if significantly co-occurring in the rectum samples. Size of the node is proportional to the number of connected edges. Microbial communities (modules) were identified following the walktrap algorithm, based on random walks that produce most likely subnetworks. From a total of 12 modules identified, six are further described: Modules 1, 2, 3, 6 and 7 are five the largest ones (more than 20 OTUs), and Module 11 is the module where the single OTU found in Probiotic B plus one more OTU are found. All OTUs contained in probiotic A (and 330 additional OTUs) are found in Module 1. The pie charts display the taxonomy break down of each module at the taxonomic class level. The inset heat map shows a statistical enrichment test after probiotic administration for OTUs present on the five largest modules across the different dolphin, water, and food microbiomes.

For water and food samples (which have the most direct possibility of influencing the dolphin rectum microbiota) we tested if their microbial communities were enriched by the OTUs present in the five largest dolphin rectum modules (n>=20; modules 1, 2, 3, 6, and 7), yet no significant difference was observed between before and during probiotic administration (Figure 6 inset). As before, this suggests that the differences in the rectal samples are not correlated with water or food OTU abundance changes, but likely due to probiotic administration. A full comparison with all the sample types and the modules enrichment is available in Table S5 at Cardona *et al.*, 2018.

2.3.3 Oceanarium water, chemistry, and microbiome correlations

The oceanarium water temperature and chemistry were essentially stable with low variance in parameter values over the course of our study. Despite this stability, even minor variation in temperature and ammonia concentration significantly correlated with changes in the number of observed species of the water-associated microbiota (water temperature correlated positively, and ammonia correlated negatively; Kendall FDR corrected, $P<0.04$). Interestingly, the change in water temperature also positively correlated with changes in the number of observed species for dolphin skin, chuff and rectum (Kendall FDR corrected, $P<0.001$); however, it is not possible to determine whether this association suggests any mechanistic interaction. In addition, ammonia, pH and alkalinity showed significant correlations with the alpha diversity of dolphin skin-associated microbiota based on total OTU counts (ammonia and pH correlated positively, and alkalinity negatively; Kendall FDR corrected $P<0.001$); however, there were no significant correlations between the water chemistry and the Faith's PD of the dolphin microbiota. This suggests that even though the number of species in a community might be changing, the

phylogenetic diversity of these communities remains conserved, and that observed changes may be due to fluctuations in rare taxa.

Multiple non-parametric tests of mean similarities were calculated, contrasting samples before and during probiotic administration for each one of the measured water properties to rule out any interconnections between the measured water properties and the dates when the probiotic treatment was administered. The correlation of water temperature, alkalinity, and salinity showed a very small but significant reduction in the period during which the dolphins received probiotics (permutation T-test FDR corrected, P<=0.01). Mean values of temperature changed from 16.3 to 15.2 °C, alkalinity from 283 to 274 parts per million (ppm), and salinity from 30.8 to 30.7 parts per thousand (ppt) during the probiotic treatment phase. Nitrate and pH increased significantly, between these two periods from a mean of 285 to 381 ppm and 7.8 to 7.9, respectively (Permutation T-test FDR corrected, P<=0.01). Meanwhile, chlorine, ammonia, and nitrite were not significantly different between the two periods of the study.

2.3.4 Site-specific microbiomes and their influence network

Potential temporospatial correlation between OTUs from different sites (dolphin, water, food, air, human) was assessed using a Dynamic Bayesian Network (DBN) analysis. DBNs relate OTU counts over adjacent time steps providing a metric for the influence of specific OTU abundances at time *t* to various other OTU abundances observed at time *t+1*. For visualization simplicity, the network was summarized at the site level, showing a multidirectional exchange among all sites (Figure 7A). Site-specific OTUs at time *t* that significantly correlated with OTUs from other sites at time *t+1* were identified with binomial tests, and food-associated OTUs had

the greatest significant association with OTUs in other sites, and their influence was significantly greater than random ($P<0.001$). In addition, the OTUs associated with dolphin rectum were also significantly associated with the microbiota of dolphin skin ($P<=0.002$), which could be seen either as a mechanistic effect (Campbell and Koch, 2017) or a direct physical exposure, as dolphin feces are readily mixed into the water and could, therefore, influence the dolphin's umbilicus region, which was the site sampled for skin microbiota.

To determine the intersection between environmental and probiotic influence, we constructed two additional DBNs, one before and one during probiotic administration (Figure 7B and C). The 'before-probiotics' network inferred that abundance of air ($P<0.001$) and food ($P=0.002$) OTUs conditionally predicted abundances of OTUs on other surfaces. Therefore, changes in the abundance of air- and food-associated taxa correlate more frequently with the changing abundance of the microbiota of other surfaces, which suggests that air and food have the biggest impact on the dolphin-associated microbiota. For the 'during probiotic administration' network, the number of OTU abundances conditionally inferred from food OTUs, nearly doubled, making food the single site in the network with significant influence on taxa abundances on other surfaces ($P<0.001$), possibly because the probiotics were interacting with the food microbiome.

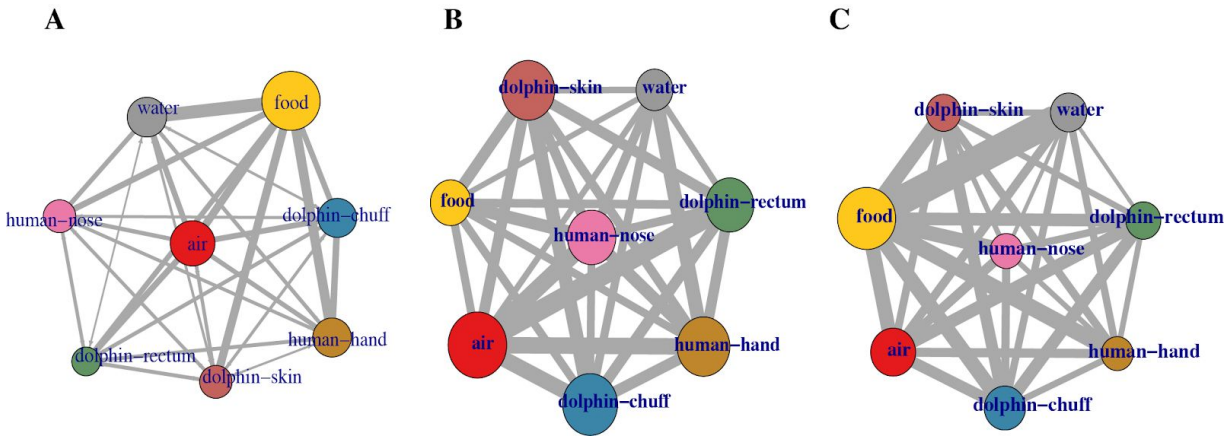

Figure 7**:** Dynamic Bayesian inference network, nodes represent the sampled sites and edges thickness the number of OTUs that are potentially influencing each other across different sites. A) Overall site level summary of the Dynamic Bayesian Network for the Shedd Aquarium microbiome, where food is the statistically most influential site in the network (Binomial, P<0.001). B) network before and C) while taking probiotics.

2.3.5 Comparison with other related studies

A recent study (Bik *et al.*, 2016), examined the microbiota of dolphins and sea lions at San Diego Bay in California, and a second undisclosed location, both part of the Navy Marine Mammal Program (MMP) (Houser *et al.*; Venn-Watson *et al.*, 2011), as well as wild dolphins from Sarasota Bay in Florida, part of a catch and release conservation program (Wells, 2009; Wells *et al.*, 2013). We combined these datasets with data from the current study (Table S6 at Cardona *et al.*, 2018) and created a bipartite network (which includes both samples and OTUs) displaying sample similarity as a function of how close samples are to each other (Figure 8). Statistical calculations determined that the greatest difference among locations was in the unweighted UniFrac beta diversity of the water (ANOSIM, R=0.72, P=0.001). Seawater samples from the MMP sites and Sarasota had greater alpha diversity (observed phylum-level bacterial diversity) than the corresponding animal and food samples from these sites; in contrast, the artificial

seawater at the Shedd Aquarium had lower phylum-level diversity than the corresponding dolphin and food microbiota from this current study (Figure 1A).

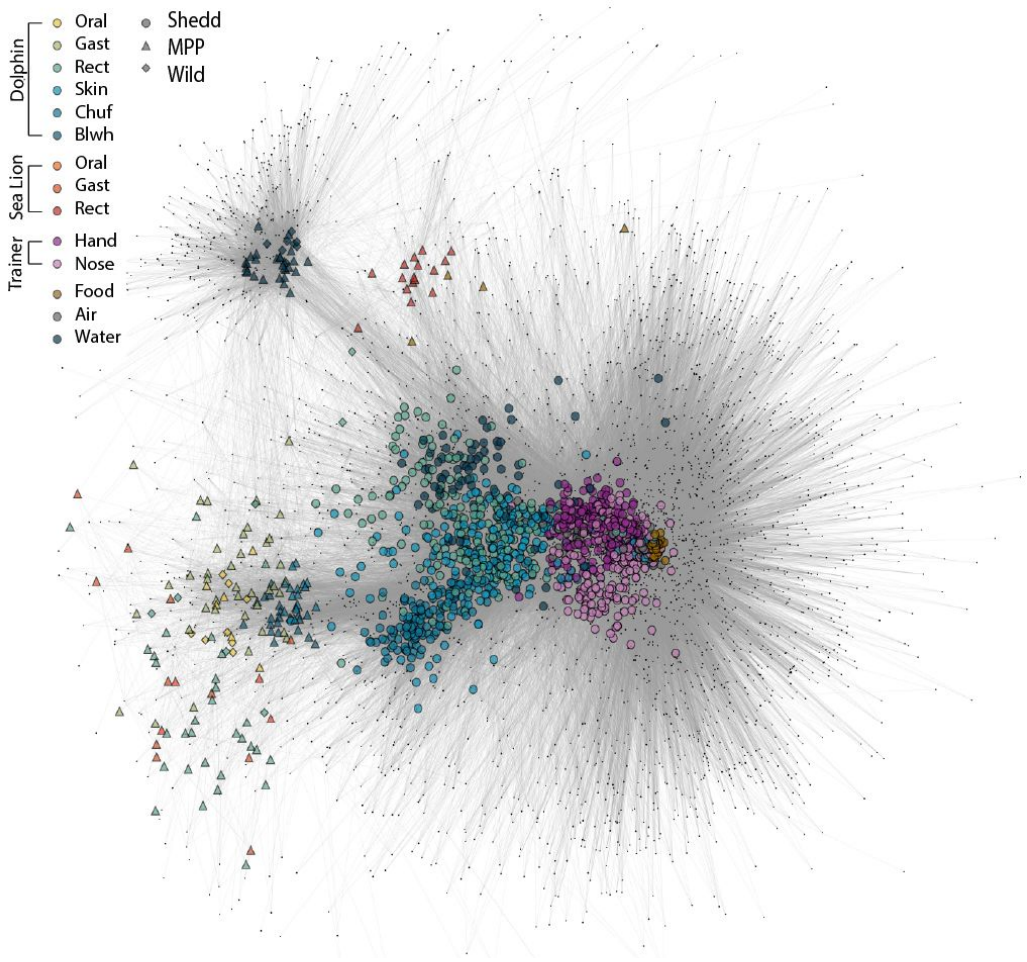

Figure 8: Bipartite network, from unweighted UniFrac distances, features Shedd (circle), MMP (triangle) and Sarasota (diamond) microbiome comparisons. Human samples were only taken in the Shedd Aquarium. Sea lion, dolphin blowhole, and dolphin gastric samples were only taken in the MMP.

Beta diversity (unweighted UniFrac) within rectum samples was significantly smaller within the location (MMP, Sarasota, and Shedd) than between location (ANOSIM R=0.79, P=0.001). MMP and Sarasota rectum microbiomes were more similar to each other than those from the Shedd Aquarium, (MMP vs. Sarasota ANOSIM R=0.25, P=0.02, MMP vs. Shedd ANOSIM R=0.84,

P=0.001 and Sarasota vs. Shedd R=0.69, P=0.001); which could have resulted from batch effect as these samples were generated in a different lab, but could also be an effect of exposure to natural seawater versus artificial seawater. While clustering the unweighted UniFrac distances for water, food, air, human handlers, and animal samples for the three locations (MMP, Sarasota and Shedd), only dolphin chuff samples were aggregated together regardless of the geographical location (Figure 9). This points to a very distinctive microbiome in the chuff, which, as shown previously, is also the site that harbored the greater number of bacteria without a known phylogeny (Figure 1A) (Harper *et al.*, 2002; Johnson *et al.*, 2009).

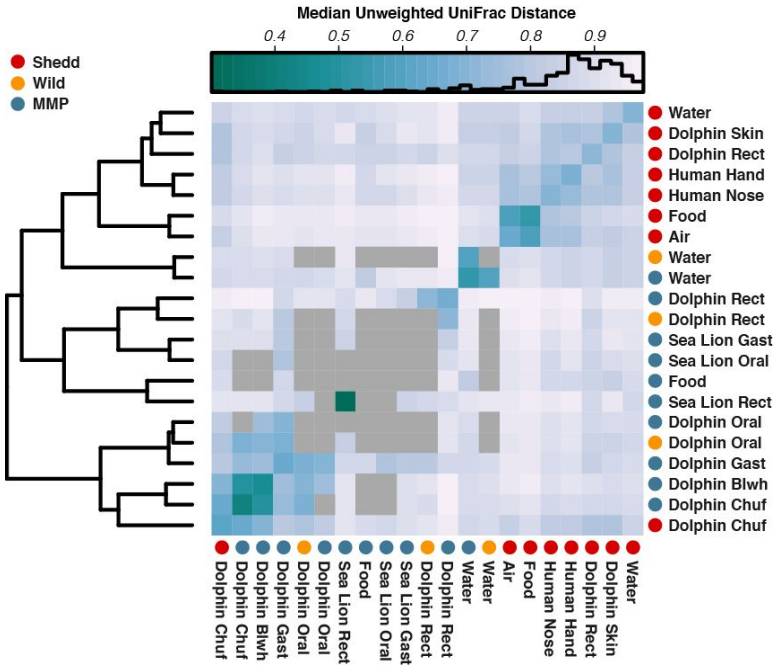

Figure 9: Median unweighted UniFrac distances between three of the dolphin's habitats: Shedd, MMP, and Sarasota. The rows and columns have been grouped with hierarchical clustering to present more similar patterns next to each other. Darker blue squares represent the higher similarity between samples. Gray squares represent comparisons with less than the minimum 100 UniFrac distances to confidently compute the statistics.

2.4. Discussion

In this study, we explored the microbiota of Pacific White-sided dolphins over a six week period, determined the association between their microbiome and that of the environments they came into contact with and observed the impact of probiotic administration on the host-microbiota. In the human microbiome, disturbances due to medications such as antibiotics, potential immune system activation due to pathogenic exposure, or diet changes can lead to sudden and dramatic changes in the structure of the microbiome (David *et al.*, 2014b). However, the degree to which the environmental microbiota interacts with the host microbiota and host environment remains unknown.

The Shedd Aquarium comprises a closed ecosystem where each environment (human, water, air, animal) shows characteristic microbiota that is predicted to strongly influence each other, with more than 80% of OTUs shared between two or more sites. This implies that there is a continuous exposure of microbes between sites, which is important to understand when implementing any microbial manipulation of the ecosystem, in particular, with water management practices. Previous studies have shown similar patterns. For example, individuals and pets under the same roof still maintained their unique signature microbiome, despite constant changes in the structure of their microbiota (Lax *et al.*, 2014). Also, despite different environmental sources associated with seawater at different locations, tunicates maintain unique microbial signatures with some degree of overlap between nearby tunicates populations (Cahill *et al.*, 2016). While the current study has not performed direct manipulation of the aquarium microbiota through controlled intervention, these data still demonstrate that air- and

food-associated microbial exposure has the largest potential influence on the host-associated microbial dynamics. When probiotics are added to the food, the microbiota in the mixture becomes the dominant influence on dolphin-associated microbial dynamics.

Despite minimal differences in the temperature, ammonia, salinity, and pH of the dolphin habitat water day-to-day, the small variations that did occur correlated with changes in the microbial diversity of the water and dolphin-sites. However, this was likely associated with the dynamics of low abundance of bacterial taxa. This still suggests that even tight control of environmental variables can elicit shifts in the structure of the microbiota, but that such changes do not have dramatic impacts on the phylogenetic diversity of microbial composition. Food and air microbiota maintained the greatest alpha diversity, and in the absence of probiotics, they also hosted the greatest number of OTUs that conditionally influenced abundances on other microbiomes (binomial, $P<0.002$). OTUs associated with dolphin-rectum significantly influenced the microbiota of the dolphin-skin either by a direct transfer (from the rectum through water to skin) or through immune modulation (Campbell and Koch, 2017). The rectum microbiome co-occurrence network also suggests that OTUs associated with the genera *Kineococcus* and *Brevibacterium* could be keystone taxa in the dolphin gut. *Kineococcus* has been associated with human oral microbiota, where it co-aggregates to support biofilm formation (Benítez-Páez *et al.*, 2014). It is possible that *Kineococcus* is also playing a role in community aggregation for the dolphin rectum microbiome. *Brevibacterium* has been associated with human skin, where it is involved in sulfur metabolism (Wood and Kelly, 2010); it is possible that it could be playing a similar role in dolphin rectum. Its presence also supports the supposition of continual exposure of microbiota between skin and rectum in this well-mixed aquatic

environment. *Brevibacterium* also showed diurnal oscillations, a phenomenon which has proven likely to increase metabolic homeostasis (Thaiss *et al.*, 2014) and could potentially be connected to the water oxidation reduction potential (ORP) water cycles (Materials and methods). Comparison between the Shedd, MMP and Sarasota environments demonstrated that the dolphin, food and water microbiota were all unique to the specific locations, although MMP and Sarasota were more similar to each other than either were to Shedd. This suggests that either difference in water and aquarium management, or a substantial batch effect in how and where samples were processed, influenced the community composition and structure substantially. However, the beta diversity for dolphin chuff was not significantly different between sites, suggesting that this is an extremely conserved environment, and also suggesting that a batch effect may not explain the other dissimilarities. The chuff-associated microbiota also harbored the most bacterial taxa that could not be reliably identified to at least a phylum (Figure 1A); indeed previous studies have characterized dolphin and whale chuff microbiota, and found similar results (Harper *et al.*, 2002; Johnson *et al.*, 2009).

Moreover, the artificial seawater used by the Shedd Aquarium oceanarium had significantly lower diversity than the seawater at both MMP and Sarasota, and while seawater generally had a greater diversity than host-associated environments, the artificial seawater was less diverse than the dolphin sites in our study. Water management practices can have a profound impact on the diversity of aquarium water (Van Bonn *et al.*, 2015), which suggests that if found to be health-promoting it might be possible to promote an increase in the microbial diversity of artificial seawater, so as to more closely resemble that of the animal's native seawater. While some authors (Attramadal *et al.*, 2016) describe how organic load and maturation of water

increases the microbial carrying capacity of the aquatic habitats, making them more stable and less open for opportunistic proliferation, other scientists present contrasting results (Fair *et al.*, 2017) suggesting that increasing microbial exposure will also increase pathogen burden. It is essential that research be done to inform recommendations, and such research must carefully examine the interaction between water chemistry, the microbiome and animal health indicators. We suggest optimal management practices will result in environmental microbiomes somewhere in between those realized due to contemporary disinfection practices and those in which no disinfection exists, as in native environments. Probiotic supplementation of mice and farm animal diets have shown how certain microbial strains are able to provide higher resistance to pathogens, possibly by competitive exclusion or stimulating the host immune system responses (Liang *et al.*, 2014; Ma *et al.*, 2008; Zhang *et al.*, 2007, 2011). In the case of the current study, we had two probiotic formulations, one with a single organism and one with more than 10 organisms. Both probiotics used in this study appeared to associate with an increase in microbial stability, but the *L. reuteri* dominated multi-member consortium led to a greater stability increase when compared to the single *L. Salivarius* probiotic. Also, the *L. reuteri* dominated formulation increased the stability of the microbiota in all three dolphin sites (chuff, rectum and skin), while the single *L. Salivarius* probiotic only increased stability for the dolphin's chuff. This evidence supports the findings from prior studies of immune modulation (Round and Mazmanian, 2009) specific to *Lactobacillus* and *Bifidobacterium* (Feighery *et al.*, 2008; Saez-Lara *et al.*, 2015; Smelt *et al.*, 2012; Vlasova *et al.*, 2016), which have already been identified as likely beneficial for the hosts.

The topology of the co-occurrence network for OTUs in the rectum samples provided an insight into the microbial ecology of the rectum and therefore, into the dolphin gut microbiota. Strains associated with the *L. reuteri* dominated formulation (including *Bifidobacterium*), formed a large cluster with more than 300 other taxa, which suggests that these probiotics show changes in abundance that match this community more than any other members of the gut microbiota. Meanwhile, changes in the abundance of the *L. Salivarius* strains only correlated with a single host-associated OTU. As the *L. reuteri* formulation is a multi-species probiotic that showed co-occurrence clustering with the largest number of host-associated taxa and was associated with the greatest increase in host-microbiota stability, this might suggest that the stabilization effect of probiotics may be more likely if the formulation can establish interconnections with the existing host microbial community; although of course, other explanations could also be relevant. For example, the *L. reuteri* dominated formulation may have exerted the biggest influence on the immune system, which led to the largest number of changes in the abundance of the gut microbiota, and hence the association density was related to indirect influence. The addition of probiotics to the diet also seems to change the topology of the microbial network of interactions, it almost doubled the number of OTU abundances conditionally associated with food OTUs, making food the single site in the network with significant influence on taxa abundances on other surfaces (Binomial, $P<0.001$).

In summary, the current study demonstrates that probiotic administration is associated with an increase in the stability of host-associated microbiota. The treatment was also associated with changes in the network structure of correlations in microbial abundance, resulting in food microorganisms having a dominant influence on the OTUs associated with dolphin and

non-dolphin sites. The study suggests that while environmentally derived exogenous bacteria can exert some influence on the dynamics of host microbiota, these differences are not as great as direct stimulation with a completely foreign exogenous microbial source. It is important to state that changes in influence and stability statistics were observed using the unweighted UniFrac metrics, suggesting that many of the stability effects are driven by changes in rare species only. This suggests that in host-associated systems, equilibrium is achieved in the presence of common microbial exposures, for example, those in the immediate usual environment. It also suggests that food and air, and hence oral gastrointestinal and respiratory interactions, have the largest effect overall. Meanwhile, un-common microbial exposures can have a profound impact on the stability and structure of microbial associations. Demonstrating that direct probiotic administration influences host microbial dynamics has major implications for animal health and aquarium management practices.

2.5 Materials and methods

2.5.1 Animals included in the study

Individual animals sampled included four Pacific White-sided Dolphins (*Lagenhorynchus obliquidens*), three females that originated from the North Pacific Ocean and had been housed at Shedd for over 20 years and one aquarium born male. The approximate ages of the females during the study period ranged from 27 to 29 years and the male was of known age 2 years 4 months to 2 years 5 months during the study period. For the purpose of this paper, the labels D1 to D4 uniquely identify each one of the dolphins.

2.5.2 Animal-trainer interactions

All dolphins included in the study were housed together in an indoor, closed recirculating synthetic seawater habitat of approximately 11.3 million liters total volume. The habitat is subdivided into several enclosures by a gate system however water circulates freely between all. In addition to the four dolphins studied, four California sea lions (*Zalophus californianu*s) and seven Beluga whales (*Delphinapterus leucas*) occupied adjacent enclosures. Each animal was fed individually by a dedicated trainer during simultaneous sessions but physically separated from each other to ensure that all food items, including those containing probiotic capsules, were consumed only by the target animal. During the study period, all trainers worked with all animals and rotated between them at different sessions. Trainers were not assigned to specific animals. During the study period, a total of 27 individual trainers were sampled.

The twenty-four-hour activity cycle of the oceanarium habitat was regular during the study period. Initial feeding sessions with the animals were conducted between 0800 and 0900 daily. During weekdays follow on scheduled sessions occurred with the animals at 1030, 1230 and 1430. On weekend days, the additional scheduled sessions were conducted at 1030, 1230, 1430, 1600. Unscheduled sessions were conducted between the scheduled sessions so that each animal was interacted with and fed up to a total of eight sessions daily.

2.5.3 Water oxidation reduction potential (ORP)

In many of the aquarium systems that use ozone contact disinfection, which is the only oxidant used in the study system, the oxidation-reduction potential (ORP) demonstrates a daily cycle, gradually increasing overnight and dropping during the morning hours when daily activities

begin. This is presumed to be a result of the increased bioload on the system during feeding and cleaning activities. System ORP was not measured during the study period but when it has been measured in the past has shown this pattern.

2.5.4 Samples collected

Dolphin's sites sampled included skin (peri-umbilicus), rectum and respiratory tract (forceful exhalation referred to as chuff). In addition, hands and noses of the animal's human handlers, oceanarium air, oceanarium water, and the dolphin's food (fish and squid blend). A total of 2,370 samples were processed. For details on the sample collection methodology and the number of samples collected on each date access the Supplementary information (SI) at Table S1 and Materials and methods, Cardona *et al.*, 2018.

2.5.5 Amplicon library preparation

Genomic DNA was extracted from environmental samples using the PowerSoil DNA Isolation Kit (MO BIO) and genomic DNA was amplified using the Earth Microbiome Project (EMP) protocols; www.earthmicrobiome.org (Gilbert *et al.*, 2014). For more details see SI Materials and Methods, Cardona *et al.*, 2018.

2.5.6 Sequence processing and statistical analysis

A total of 2,370 samples were sequenced on the Illumina MiSeq. After 1,084 technical replicates were identified the data was reduced to 1,286 pooled samples. The reads were quality filtered using Quantitative Insights Into Microbial Ecology (QIIME) (Caporaso *et al.*, 2010) and

downstream processing of sequence data utilized QIIME, R and Oligotyping (Eren *et al.*, 2013). For more details see SI Materials and methods, Cardona *et al.*, 2018.

2.5.7 Co-occurrence and inference networks

The co-occurrence network was calculated for rectum samples only, with a resulting network of 717 nodes and 68,515 edges. In preparation for network creation we removed OTUs with abundances less than 0.01% of the total number of OTUs, for a subset of 717 OTUs. Co-occurrence of OTUs used the WGCNA package (Langfelder and Horvath, 2008). OTUs from all surfaces were prefiltered with DESeq2 (Love *et al.*, 2014) to select only the 41 OTUs with a statistically differential abundance in at least one sampled site (P=0.001 and abs(log2(abundance))>1) and DBNs were created via Banjo (Hartemink *et al.*, 2005). Network properties and visualization was done with the CAVNet R package (Cardona, 2017). For more details see SI Materials and methods, Cardona *et al.*, 2018.

CHAPTER 3

CO-OCCURRENCE NETWORKS TO CHARACTERIZE HIGH RELATIVE HUMIDITY BUILT ENVIRONMENT MICROBIOMES

Work Published at Nature Communications (Lax and Cardona *et al.*, 2019)

3.1 Overview

Despite considerable efforts to characterize the ecology of bacteria and fungi in the built environment (BE), the metabolic mechanisms underpinning their colonization and successional dynamics remain unclear. Here, we applied bacterial/viral particle counting, qPCR, amplicon sequencing of the genes encoding 16S and ITS rRNA, and metabolomics to longitudinally characterize the ecological dynamics of four commonly used building materials (oriented strand board, medium density fiberboard, gypsum board, and mold-free gypsum (MF) board) maintained at constant high humidity conditions (~94% RH). We varied the natural inoculum provided to each material by placing them in different occupied spaces, and we wet the surface of half of the samples of each material to simulate a potable water leak. As expected, different materials showed different bacterial and viral particle abundance, with wet materials having higher growth rates and lower alpha diversity compared to non-wetted materials. Wetting described the majority of the variance in bacterial, fungal and metabolite structure, while the location of inoculation was weakly associated with bacterial and fungal beta diversity. Material type only influenced bacterial and metabolic diversity. Metabolites indicative of microbial activity were identified, as were those that were native to the surface material. Glucose-phosphate was abundant on all materials (except mold-resistant gypsum) and was correlated with *Enterobacteriaceae*, which could indicate a potential bacterial nutrient source. A compound consistent with scopoletin, a plant metabolite with antimicrobial activity, was

significantly negatively correlated with *Bacillus* and positively correlated with *Pseudomonas* and enriched in medium density fiberboard materials. In wet samples, the alkaloids nigragillin and fumigaclavine C, both with antimicrobial properties, were significantly positively correlated with the fungal phylum Ascomycota. Nigragillin was also negatively correlated with *Bacillus* and *Pseudomonas* abundance. Thiabendazole and azoxystrobin (antifungal compounds) were highly abundant on mold-resistant gypsum board and likely directly influenced the decreased fungal growth observed on this material. The mold-resistant gypsum board also showed a significant increase in bacterial alpha diversity, and bacterial and viral particle abundance, as well as a decrease in metabolite diversity, likely a result of reduced fungal growth. *Penicillium* taxa were positively correlated with thiabendazole, which suggested the persistence of resistant strains. Also, specific to the wet samples, *Bacillus* abundance was positively correlated with the azoxystrobin, suggesting bi-directional competitive adaptation, and positively correlated with metabolites known to interfere with *Pseudomonas* biofilm formation, which could explain the anti-correlation between these taxa. As expected, high moisture conditions enabled faster growth of microorganisms, whose composition, chemistry, and competition were shaped by material and inoculation location, suggesting that both fungal and bacterial growth need to be considered when determining the impact of dampness in built environments.

3.2 Background

The microbiology of the built environment comprises bacteria, archaea, fungi, viruses and protists, all of which maintain growth potential under varying physicochemical regimes. Many recent studies of this ecosystem have applied molecular sequencing techniques to characterize

microbial community relationships and dynamics under varying occupant density, building type and location, environmental conditions, and material type (Lax *et al.*, 2017; Adams *et al.*, 2016; Chase *et al.*, 2016; Stephens, 2016; Lax *et al.*, 2014). However, most of these studies have investigated communities sampled from relatively dry materials on which microbes are likely biologically inactive unless they experience liquid water or high relative humidity (RH) (Chase *et al.*, 2016). It is widely accepted that fungal growth can occur at RH >75-80% and material decay can occur at RH >95%, depending on the material (Viitanen *et al.*, 2010, Johansson *et al.*, 2012).

Dampness is a fairly common occurrence in buildings, with approximately half of all homes in the U.S. having experienced dampness or mold (LBNL Indoor Environment Group, 2019a). Building material dampness occurs for different reasons, including: bulk liquid entry from floods, extreme weather events, and plumbing system problems; rain or snow entry through leaks in building envelopes and roofing systems; and high water vapor content resulting from moisture migration through building materials or condensation of warm humid air on cold surfaces (LBNL Indoor Environment Group, 2019b). Dampness and the presence of visible mold have been consistently associated with adverse human health outcomes, including respiratory and allergic effects (Mendell *et al.,* 2011, Quansah *et al.*, 2012, Fisk *et al.*, 2010, Jaakkola *et al.*, 2013). Hypotheses to potentially explain these associations include a combination of exposure to specific microbial agents (Institute of Medicine, 2004), varied gene expression and metabolism (Hegarty *et al.*, 2018), and the release of fungal metabolites including mycotoxins (Miller *et al.*, 2014) and microbial volatile organic compounds (mVOCs) (Roze *et al.*, 2013).

Although fungal growth on building materials has been studied for decades (Hyvärinen *et al.*, 2002, Gravesen *et al.*, 1999, Hoang *et al.*, 2010; Pasanen *et al.*, 1992), only a limited number of studies have used molecular techniques to investigate bacterial and fungal growth, microbial community dynamics, and/or metabolic activity on common buildings materials exposed to liquid water and/or high humidity conditions (Coombs *et al.*, 2017). Therefore, we characterized the bacterial and fungal concentration and diversity, as well as the production of microbial metabolites, on samples of four common building materials incubated at constant ~94% relative humidity: oriented strand board (OSB), medium density fiberboard (MDF), gypsum wallboard, and mold-free gypsum wallboard (MF). The materials were selected as a sample of convenience to represent a relatively wide variety of common building and furniture material types that were also likely to experience high variability in microbial growth. We varied the BE source of inoculation and purposely wet half of the samples to assess how indoor microbial sources and the presence of liquid water influence community structure and metabolite profiles of these materials over multiple time points. We used several techniques to quantify microbial growth and microbial community composition and functional metabolism including: bacterial and viral like particle counts, image processing of visible mold growth, qPCR, amplicon sequencing of 16S and ITS rRNA marker genes, and metabolomics. Results from these different methods were integrated via co-occurrence network approaches, which provided insights into microbial community organization and environmental interaction. Improved understanding of how bacterial and fungal metabolism is shaped by environmental properties (e.g., the presence of water, surface material composition) and inoculating source (e.g., building location, occupancy patterns) could have important implications for architectural design, construction, building

management, and occupant health. Therefore, determining the microbial metabolic dynamics in these high RH environments should be an important research priority.

3.3 Results

Different sampling strategies were tested, including repeated sampling of the same coupons at each time point and sampling new coupons at each time point, producing four sampling groups: regular (REG), duplicate (DUP), resampled (RE) and resampled duplicate (RE.DUP) for both fungal and bacterial samples (Figure 10). After Mantel test results verified the agreement of these approaches, all samples of the same type were combined as technical replicates (Table 3 for fungal and Table 4 for bacterial samples**)**. Microbial datasets were later rarefied to the same sequencing depth: 1,000 reads for bacteria and 10,000 reads for fungi. Unfortunately, rarefaction removed all bacterial samples from MDF materials, which had very low read counts. After rarefying the data, a comparison of the control (lab-inoculated) and non-control (residence-inoculated) samples reflected that control samples looked very similar in bacterial and fungi diversity to non-control samples, (mantel >= 0.49 and >= 0.43 for location 1 and location 2 respectively, all with a $p < 1E\text{-}05$), perhaps because air could still transmit through the non-hermetic foil cover and microbes from the interior of the wood (not killed with the sterilization) could have found their way to the surface. It is also possible that the coupon itself could remain an important reservoir of microbial communities that contribute to the microbial diversity of the samples. Based on these results, the covered laboratory location was treated indistinctly than the other two locations. For more details see Materials and methods section.

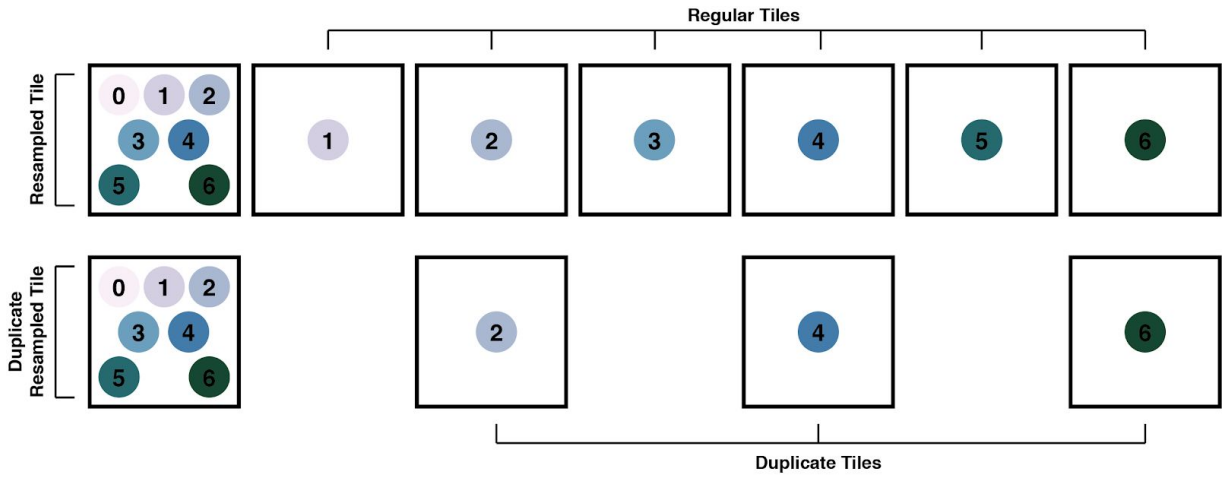

Figure 10: Experimental setup of the project; illustration of the experimental setup and coupon sampling procedures, regular, duplicate, resampled and duplicate resampled tiles.

	REG			RE			RE.DUP		
	mantel	significance	n	mantel	significance	n	mantel	significance	n
RE	0.67	1E-05	74	<NA>			<NA>		
RE.DUP	0.71	1E-05	79	0.85	1E-05	77	<NA>		
DUP	0.81	1E-05	44	0.75	1E-05	40	0.79	1E-05	39

Table 3: Triangular matrix of Mantel test results calculating the correlation among fungal samples across different sampling strategies.

	REG			RE			RE.DUP		
	mantel	significance	n	mantel	significance	n	mantel	significance	n
RE	0.62	1E-05	75	<NA>			<NA>		
RE.DUP	0.56	1E-05	75	0.72	1E-05	80	<NA>		
DUP	0.63	1E-05	56	0.53	1E-05	36	0.50	1E-05	35

Table 4: Triangular matrix of Mantel test results calculating the correlation among bacterial samples across different sampling strategies.

3.3.1 Visible growth, particulate counts and qPCR

Visible microbial growth occurred much faster and covered a far greater percentage of the surface area on wet coupons than on non-wetted coupons (Figure 11A). OSB and MDF had the greatest coverage and fastest growth: all wet OSB and MDF coupons reached at least 50% visible microbial coverage by day 20, while non-wetted coupons of these types reached < 25% coverage. No growth was ever visible on the mold-resistant gypsum coupons. Epifluorescence microscopy revealed that counts of bacterial like particles (BLP) and viral-like particles (VLPs) calculated on samples TP0 to TP3 (15 days post incubation) were strongly correlated ($R^2 = 0.65$, $p = 2.8e^{-23}$) (Figure 11B), with VLP counts statistically lower than bacterial counts in all samples (ANOVA $<=10^{-4}$) and in both wet and non-wet conditions (two-sided non-parametric t-test $p <= 0.035$) (Figure 11C). This is in keeping with previous research that found very low VLP:bacteria ratios on built surfaces (Gibbons *et al.*, 2015) and indoor aerosol samples (Prussin *et al.*, 2015). In our dataset, the mean VLP-bacteria ratio was 0.86 ± 0.07, with a minimum of 0.61 and a maximum of 1.02 across all samples.

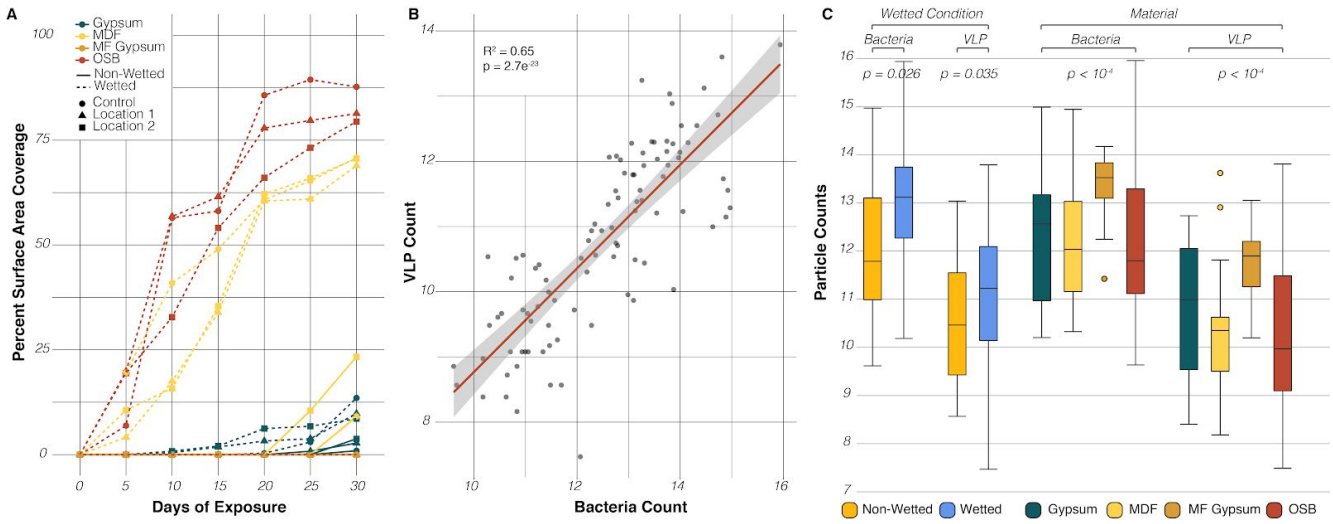

Figure 11: Microbial growth rates vary across sample types. (A) Percent of surface area covered by visible microbial growth through time (n=168, 4 materials, 3 locations, 2 wetting conditions, 7-time points). The color indicates coupon material, point shape indicates inoculating location, and line type indicate whether the coupon was wet before incubation. (B) Correlation in the counts of bacteria-like (BLP) and viral-like particles (VLP) across all coupons (n=96 samples, 4 materials, 3 locations, 2 wetting conditions, 4-time points). (C) Boxplots of BLP and VLP counts by wetting condition and by material (n=96 samples).

While BLP (only estimated for TP0 to TP3 samples) and bacterial qPCR agree that wetted samples had higher counts than non-wetted samples, with a 4 and 99-fold median cell count increase respectively, cell counts inferred from these two methods drastically differ for different material types and over time. Most notably, the mold-free gypsum (MF) had the greatest BLP counts but also the lowest 16S rRNA qPCR median cell counts (44 to 97-fold lower than other materials). Moreover, the BLP cell counts were essentially constant over time, while qPCR counts steadily increased, with TP6 (30 days post incubation) being 209-fold greater than the 394 cells median count per µl at TP0. To further confirm the differences, we calculated the overall correlation between paired bacterial qPCR and BLP counts and the results were not significant, emphasizing different biases for each method.

For fungal qPCR we observed MF had the lowest abundance, while all other materials had a range of 37 to 239-fold increase over MF. Wetted samples revealed a 72-fold increase in qPCR median read abundance over non-wetted samples. Also, the qPCR read abundance increased steadily over time, in such a way that TP6 was 750-fold greater than the 5 cells median count per µl at TP0.

3.3.2 Bacterial, fungal and metabolite diversity

The bacterial and fungal communities in our study tended to decrease in diversity over time, as measured by the Shannon Index (Shannon H'), which incorporates both the richness and evenness of the community. Given that our data was rarified to an even depth before analysis, this decrease in diversity is indicative of the increased relative abundance of certain community members, and suggests the preferential proliferation of certain taxa in the inoculating community. In our bacterial dataset, wet samples experienced faster declines in diversity than non-wetted samples, and were significantly lower in diversity at the end of the study than non-wetted samples (Figure 12A), suggesting that certain bacterial taxa grew quickly in the wet environment and became dominant within the community. In our ITS dataset, we also observed a faster decline in diversity in wet samples, although wet samples were significantly more diverse than non-wetted samples by the end of the study (Figure 12B). The decrease in fungal diversity in wetted samples was not monotonic, with an initially steep decline and a subsequent increase. This may reflect fast growth by a small number of taxa that quickly dominated the community, followed by the growth of other taxa with slower growth rates. Similar patterns are observed

when looking at the diversity changes for each individual material (Figure 13) with the exception of a lack of bacterial growth for wet MDF samples and reduced bacterial growth on dry OSB after the study was halfway completed. In contrast, we observed no significant changes in metabolic diversity over time for either wet or non-wetted samples (Figure 12C).

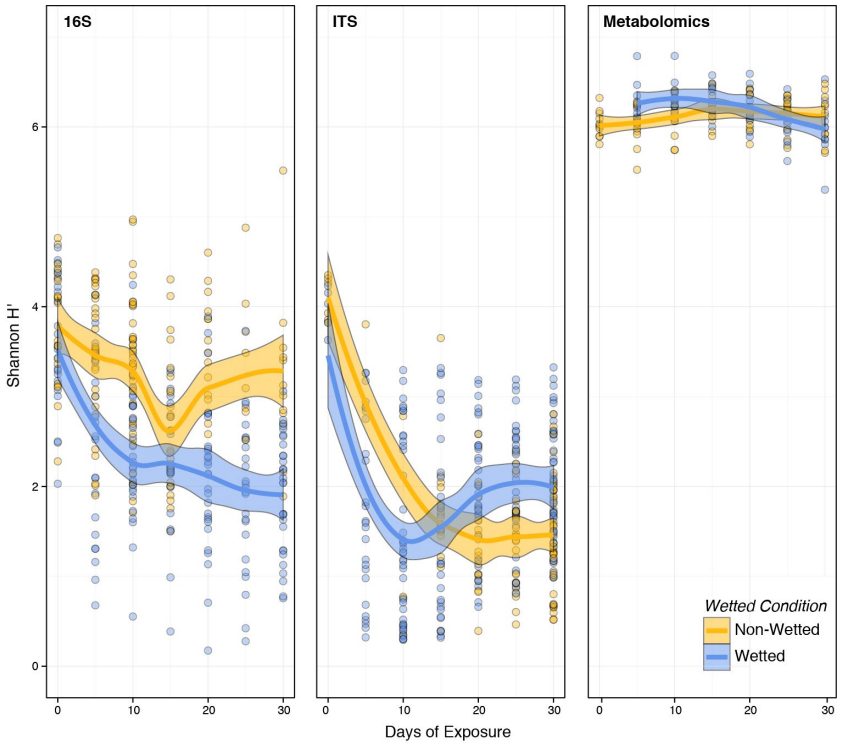

Figure 12: Change in the Shannon Index of samples over time. Points represent individual samples and the trend lines are a smoothed moving-average of the mean and shaded regions indicate the standard error (n=338, 330 and 144 samples for 16S, ITS and Metabolomics, respectively).

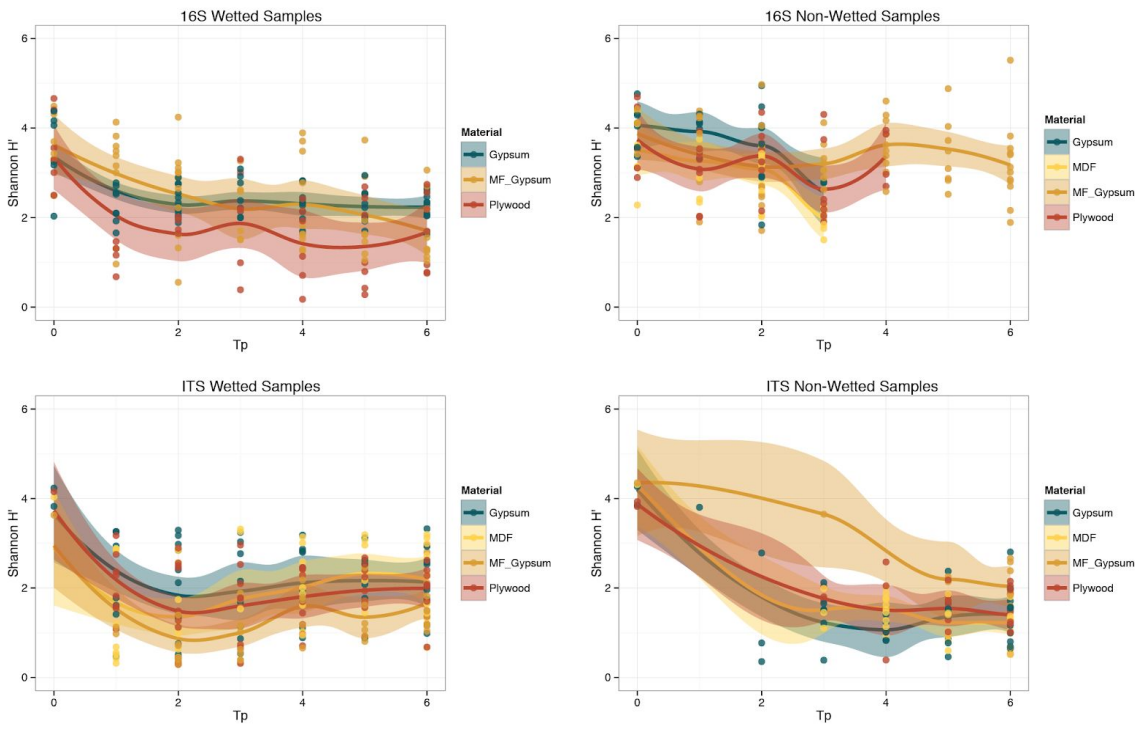

Figure 13: Change in the Shannon Index by material over time. Points represent individual samples (n=338, 330 samples for 16S and ITS, respectively) and the trend lines are a smoothed moving-average of the mean and shaded regions indicate the standard error.

3.3.3 Microbial compositional changes

Across all samples, the diversity of bacteria within the community was significantly correlated to the diversity of fungi (Corr = 0.28, p=0.0003) (Figure 14A). Interestingly, neither bacterial nor fungal diversity was significantly correlated to the metabolite diversity, perhaps because of a narrower range of observed metabolite diversity compared to the taxonomic datasets. We observed striking changes in the relative abundance of certain bacterial (Figure 14B) and fungal (Figure 14C) genera over time, which were largely dependent on wetting condition. In the bacterial dataset, *Bacillus* almost immediately came to dominate wet samples, with an average relative abundance as high as 50% after the 2nd time point, even though it represented a

negligible part of the community at the start of sampling. *Bacillus* abundance also increased in non-wetted samples, although to a much smaller extent. A similar pattern was observed for the genera *Pseudomonas* and *Erwinia*, which also represented a very small fraction of community diversity at the start of sampling but quickly increased in abundance in wet (but not non-wetted) samples. Interestingly, a very large percentage of reads from early time point samples, both wet and non-wet, were of chloroplast origin. In wet samples, the number of chloroplast reads quickly declined as the bacterial genera proliferated. In non-wetted samples, chloroplast read abundance remained high, and dominated the sequencing effort to such an extent that discarding those reads would have dropped the majority of non-wetted samples below the rarefaction depth. While these likely represent residual DNA signatures from the plant material used to construct each coupon, we have chosen to keep them in the analysis. Figure 15A shows how this dynamic slightly vary for each different material type.

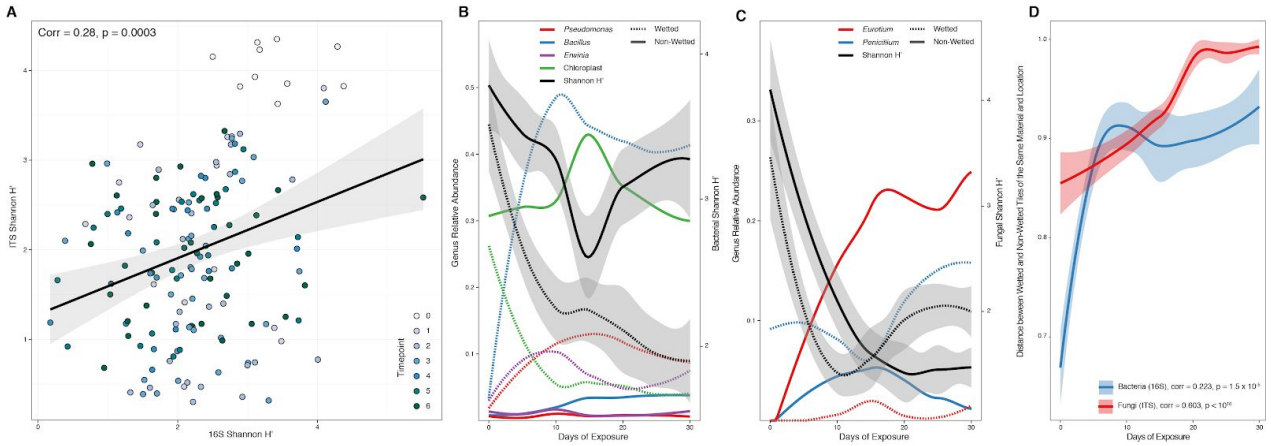

Figure 14: Overview of microbial community succession in the built environment. (A) Fungal diversity and bacterial diversity are significantly correlated across communities (n=153 samples). Points represent individual samples, colored by the time point at which the sample was taken. (B) Changes in the relative abundance of selected bacterial genera over the course of succession (n=338 samples). Lines represent a smoothed moving average of the mean. Genus is indicated by color and wetting condition is indicated by line style. Average community diversity (Shannon H' at OTU level, as in Figure 3) is indicated by black lines with standard error indicated by the gray shaded region. Genus abundance is indicated on the left y-axis and Shannon H' is indicated on

Figure 14 (continued): the right y-axis. (C) Changes in the relative abundance of selected fungal genera over the course of succession (n=330 samples). Formatting is as in (B). (D) Wet vs. non-wetted replicates of coupons of the same material and inoculating location become increasingly dissimilar over the course of community succession (n=338, 330 samples for 16S and ITS, respectively). The y-axis is the Bray-Curtis distance between replicates. Spearman correlation between community dissimilarity and time is indicated in the legend error.

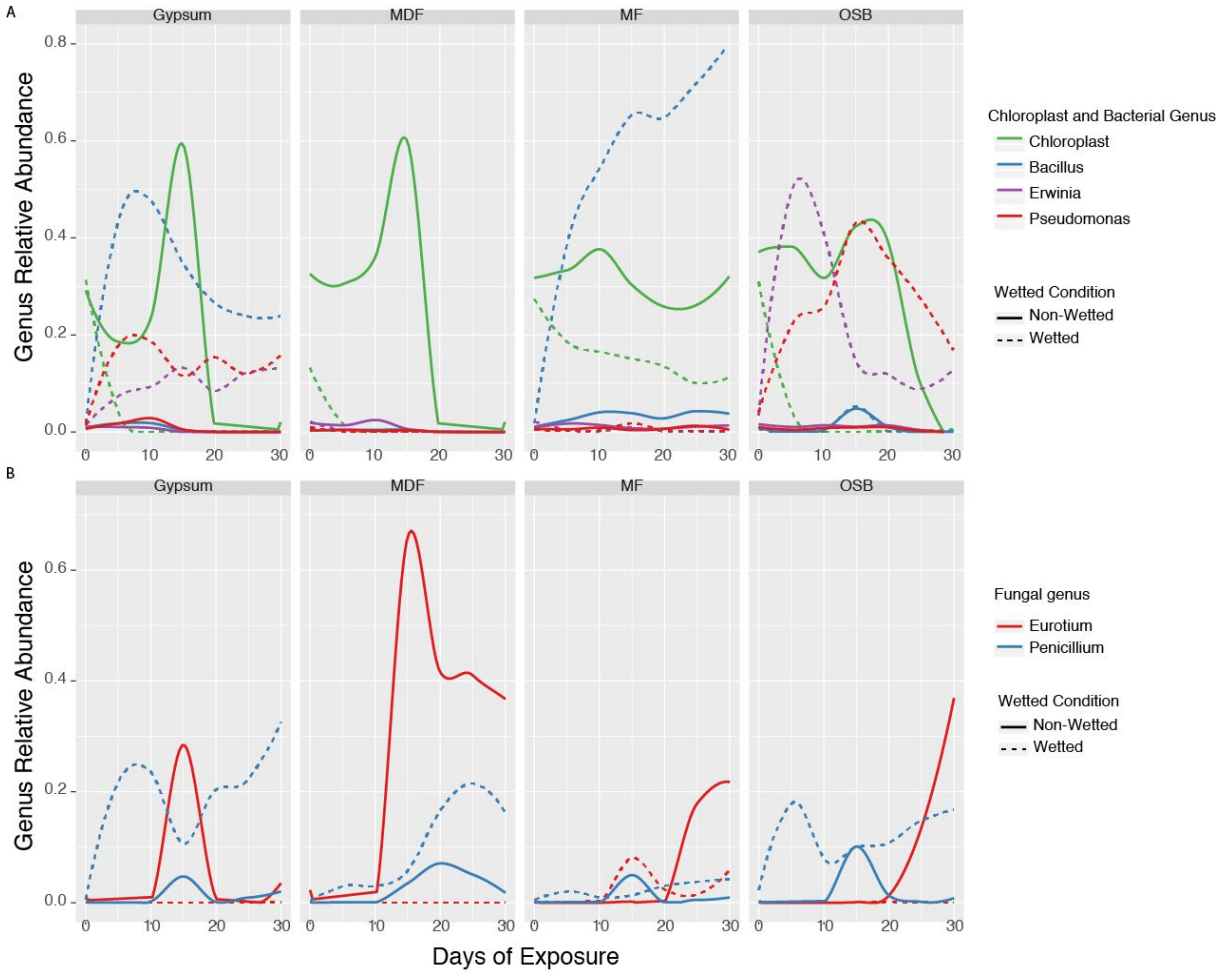

Figure 15: Microbial succession by material over time Changes in the relative abundance of selected microbial genera for each material over the course of succession. (n=338, 330 samples for 16S and ITS, respectively). (A) Lines represent a smoothed moving average of the mean. Genus is indicated by color and wetting condition is indicated by line style. (B) Changes in the relative abundance of selected fungal genera over the course of succession. Formatting is as in (A).

The majority of reads in the ITS dataset that could be taxonomically assigned to a genus belonged to one of two genera: *Eurotium* and *Penicillium*. *Eurotium* abundance was negligible at

the beginning of community succession but quickly flourished in non-wetted samples, becoming the most abundant known genus in those samples by time point 2 (Figure 14C). By contrast, *Eurotium* did not become abundant across wet samples. *Penicillium* abundance was, on average, consistently higher in wet samples than in non-wetted samples, and its abundance was significantly anti-correlated to *Eurotium* relative abundance (corr = -0.12, p = 0.033). These taxa-specific changes were mirrored by community level differentiation, where wet vs. non-wetted coupons of the same material and inoculating location became significantly more dissimilar (Bray Curtis, Spearman's Correlation, $p<0.01$) in both their bacterial and fungal community structure over time (Figure 14D). Figure 15B shows how this dynamic slightly vary for each different material type.

3.3.4 Environmental factors associated with microbial and metabolite diversity

We used ANOSIM to calculate the factors significantly correlated with differences in the microbial communities across our three datasets. Bray-Curtis dissimilarity was calculated for the bacterial, fungal, and metabolite datasets, and ANOSIM was used to determine whether distances between samples of the same metadata factor (i.e. wetting condition, inoculating location, and material) were significantly lower than distances between samples of different types (Figure 16). In our bacterial rRNA dataset, wetting condition, location, and material each had a significant impact on bacterial community structure (all $p < 0.0001$ based on 10^5 randomized permutations), with wetting having the most pronounced effect ($R = 0.418$). In general, non-wetted samples tended to be more similar to each other than wet samples were to each other, which is likely due to the dominance of a single chloroplast OTU. The material had a

less pronounced effect (R=0.247) and location had the least evident effect on bacterial community structure (R=0.133). The time that occupants spent in close proximity of the samples was similar in each residential location (0.16% and 0.18% of total time, respectively) and was not included in the environmental factor analysis.

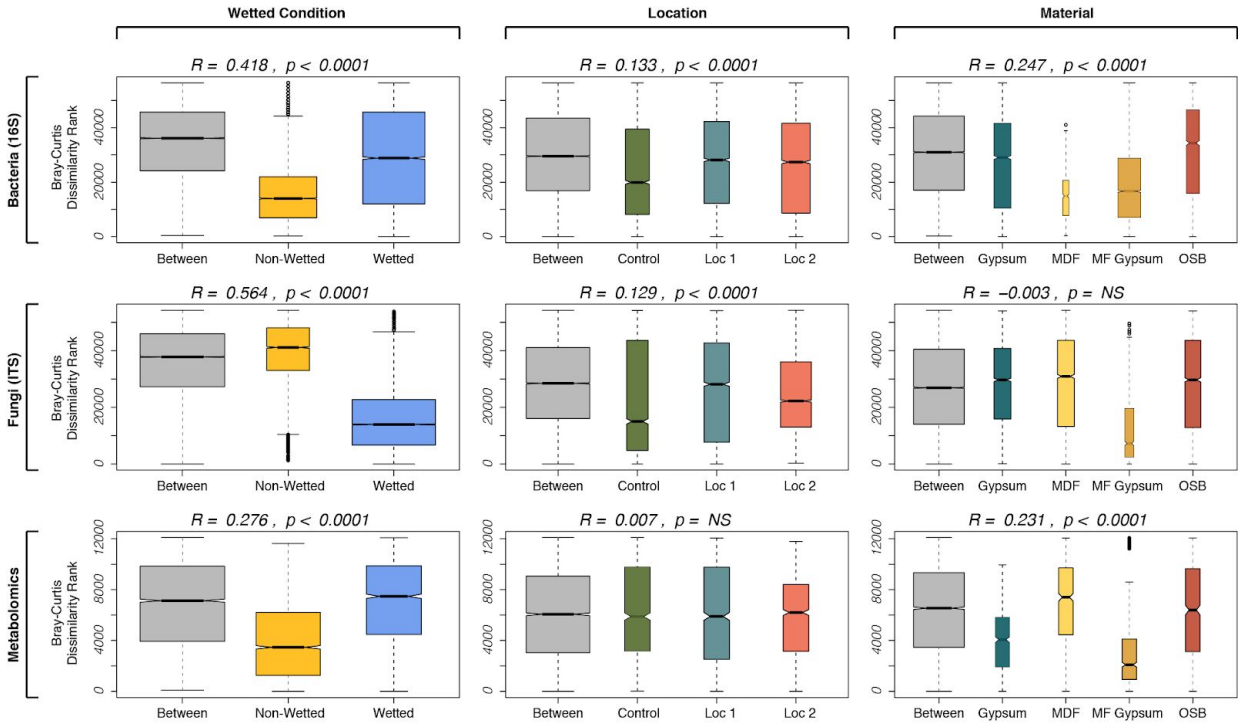

Figure 16: ANOSIM quantifies the influence of metadata factors on the dissimilarity between samples. Columns represent different metadata factors and rows represent the three datasets in this study (n=338, 330 samples for 16S and ITS, respectively). Boxplots depict the range of ranked Bray-Curtis dissimilarities within and between factors (lower rank = lower dissimilarity). Boxplot width indicates the number of samples represented by the boxplot.

Interestingly, fungal community structure was not significantly described by variance in material, while location had a relatively weak (R = 0.129) though highly significant ($p < 0.0001$) association, suggesting that variations in fungal communities that settle on materials (which have been shown to be driven largely by outdoor fungal communities (Adams *et al.*, 2013) influence community structure upon experiencing wetting and high RH conditions. The wetting condition

was by far the most influential factor influencing fungal community structure ($R = 0.564$, $p < 0.0001$), and in contrast to the bacterial data, wet samples were much more similar to each other than were non-wetted samples. Metabolite diversity within the community was also affected by the wetting condition ($R = 0.276$, $p < 0.0001$), with non-wetted samples more similar to each other than wet samples. The material also played a significant role in metabolite diversity ($R = 0.231$, $p < 0.0001$), and mold-free gypsum samples were particularly metabolically similar, likely due to the lack of fungal growth and the underlying chemical composition of the material. The inoculating location had no significant effect on the diversity of metabolites despite having a significant effect on both the bacterial and fungal community membership. We visualized sample similarity using non-metric multidimensional scaling (NMDS) ordination based on Bray-Curtis dissimilarity (Figure 17). We converted material, location, and wetting condition into binary variables (1 = yes, 0 = no), which were fitted onto the ordination, keeping only the significant vectors (R^2 values for each vector and their significance is presented in Supplementary Table 1 at Lax and Cardona *et al.*, 2019). Visually, both bacterial and fungal beta diversity were more differentiated by wetting condition, while metabolites were visually differentiated by both wetting condition and surface material, likely due to the underlying chemistry of the material and then the subsequent metabolic activity of the microbes when coupons were wetted.

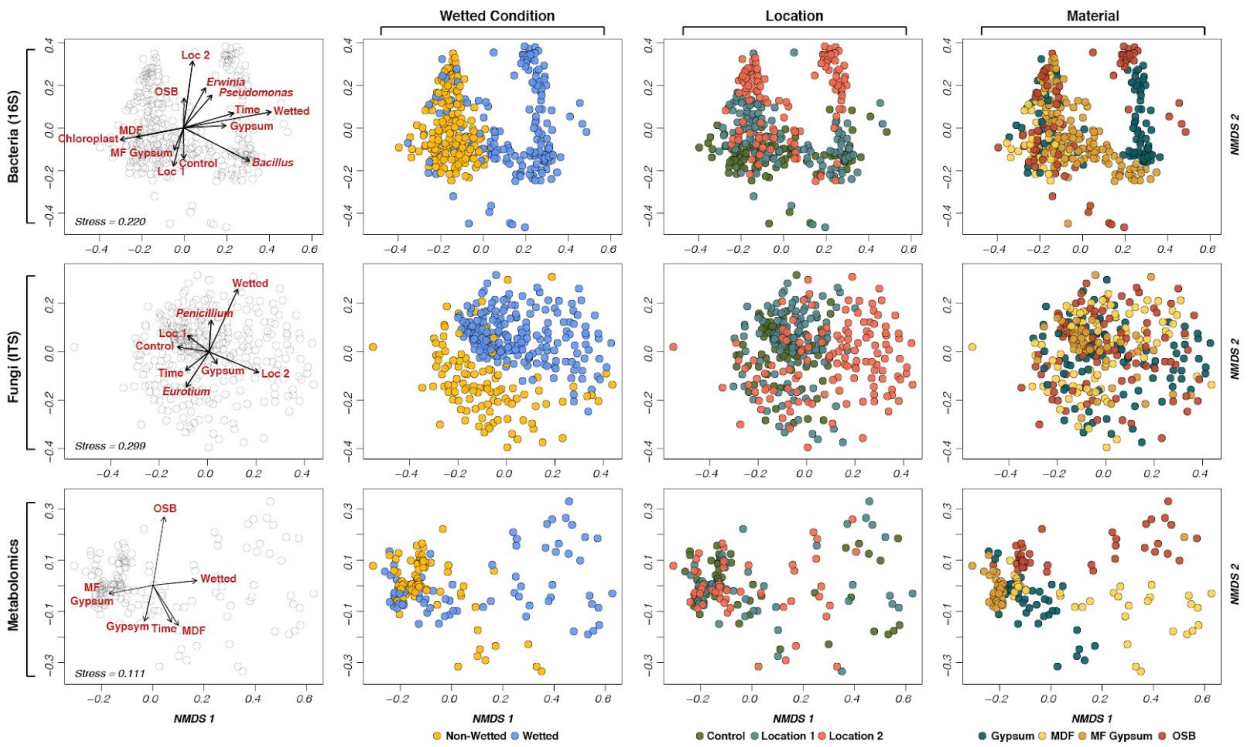

Figure 17: NMDS plots illustrate the clustering of sample diversity by sample type. Each row comprises four identical NMDS plots (n=338, 330 and 144 samples for 16S, ITS and Metabolomics, respectively). The leftmost plot illustrates the ordination's association with environmental variables and the remaining plots color sample points by various metadata factors. The stress on the NMDS plot is indicated in the rightmost plot in each row error.

3.3.5 Bacterial and fungal network co-occurrence

SparCC (Friedman & Alm, 2012), an algorithm developed to quantify correlations on microbial compositional data (data that has been subject to rarefaction), and a correlation threshold >0.4, uncovered co-occurrence patterns between taxa from each kingdom. In the bacterial network (Figure 18) three co-occurrence clusters were identified, the *Bacillus* cluster, *Pseudomonas* cluster, and a cluster comprising chloroplasts and mitochondria. As expected, these groups correspond with the most abundant taxa. On all wet materials and on all samples of gypsum (both wet and non-wet), 95% of associations between *Bacillus* and *Pseudomonas* were negative

correlations (Figure 19 and Figure 20). On non-wetted OSB, MDF, and MF there were no negative correlations between *Pseudomonas* and *Bacillus.* Interestingly, there is a dramatic increase in the absolute number of significant co-occurrence relationships between bacterial OTUs in wet (74) versus non-wetted samples (48), which is a 54% increase in the number of edges, suggesting that the wetting event has made the environment more suitable for microbial growth interactions. In the fungal correlation network, *Penicillium* OTUs co-occurred with many unknown fungal genera, while OTUs corresponding to *Aspergillus* and its subset, *Eurotium,* maintained monophyletic clusters (Figure 21). As with the bacterial co-occurrence networks, fungal OTUs associated with wet coupons had negative correlations among each other, although the number was much smaller than for bacteria. Only 7 fungal OTUs were negatively correlated on wet materials, mainly between unknown genera and an abundant *Penicillium* OTU (Figure 22). Strikingly, unlike bacteria, the absolute number of significant co-occurrence relationships between fungal OTUs declined in wet (555) versus non-wetted samples (1,133), which was a 104% decrease in the number of edges, suggesting an inverse co-abundance response between bacteria and fungi during growth.

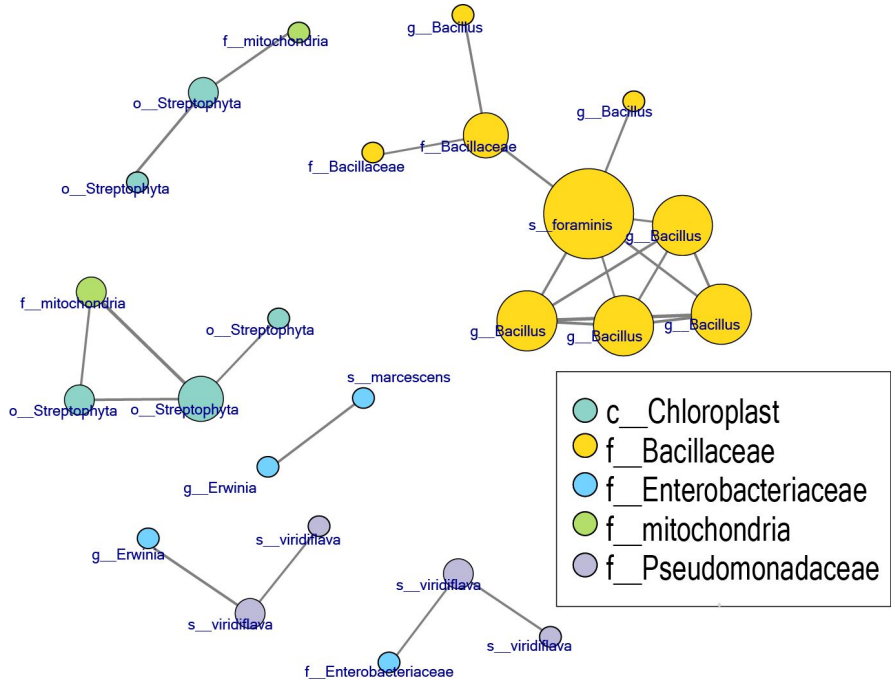

Figure 18: Bacteria-Bacteria co-occurrence network. Co-occurrence network (from n=83 bacteria samples) shows highly correlated bacteria form monophyletic clusters for samples containing more abundant taxa.

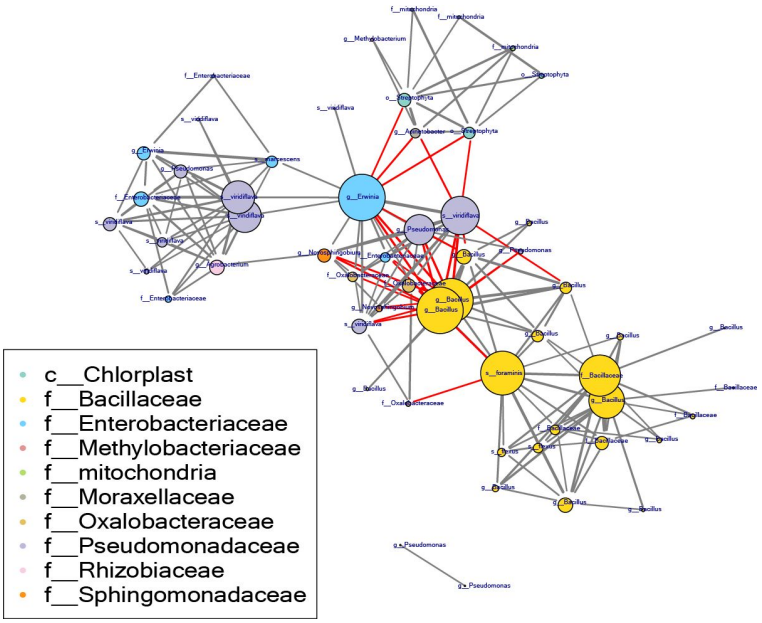

Figure 19: Bacteria-Bacteria co-occurrence network on wet samples. Co-occurrence network (from n=39 wet samples) shows how *Pseudomonas* and *Bacillus* are anticorrelated on wet samples for samples containing more abundant taxa.

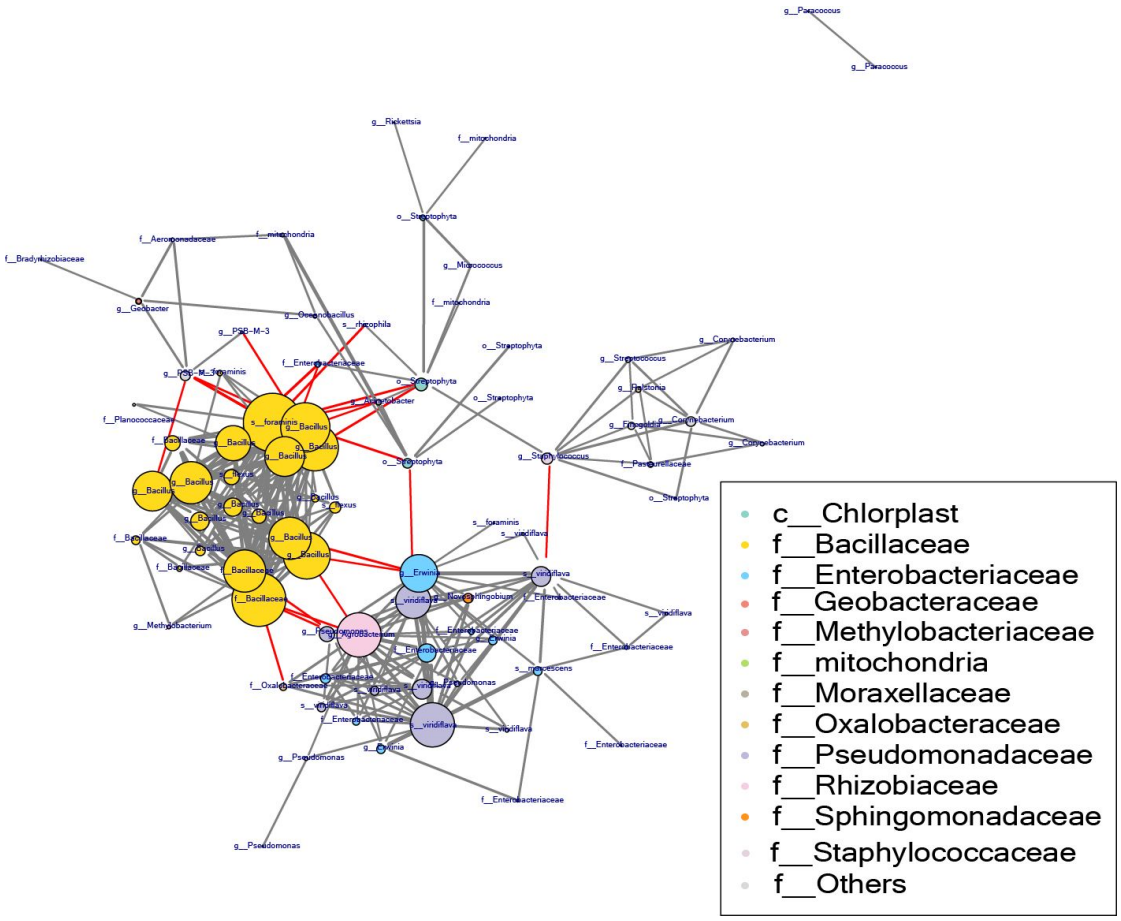

Figure 20: Bacteria-Bacteria co-occurrence network on gypsum samples. Co-occurrence network (from n=24 gypsum samples) shows how *Pseudomonas* and *Bacillus* are anticorrelated on all gypsum samples.

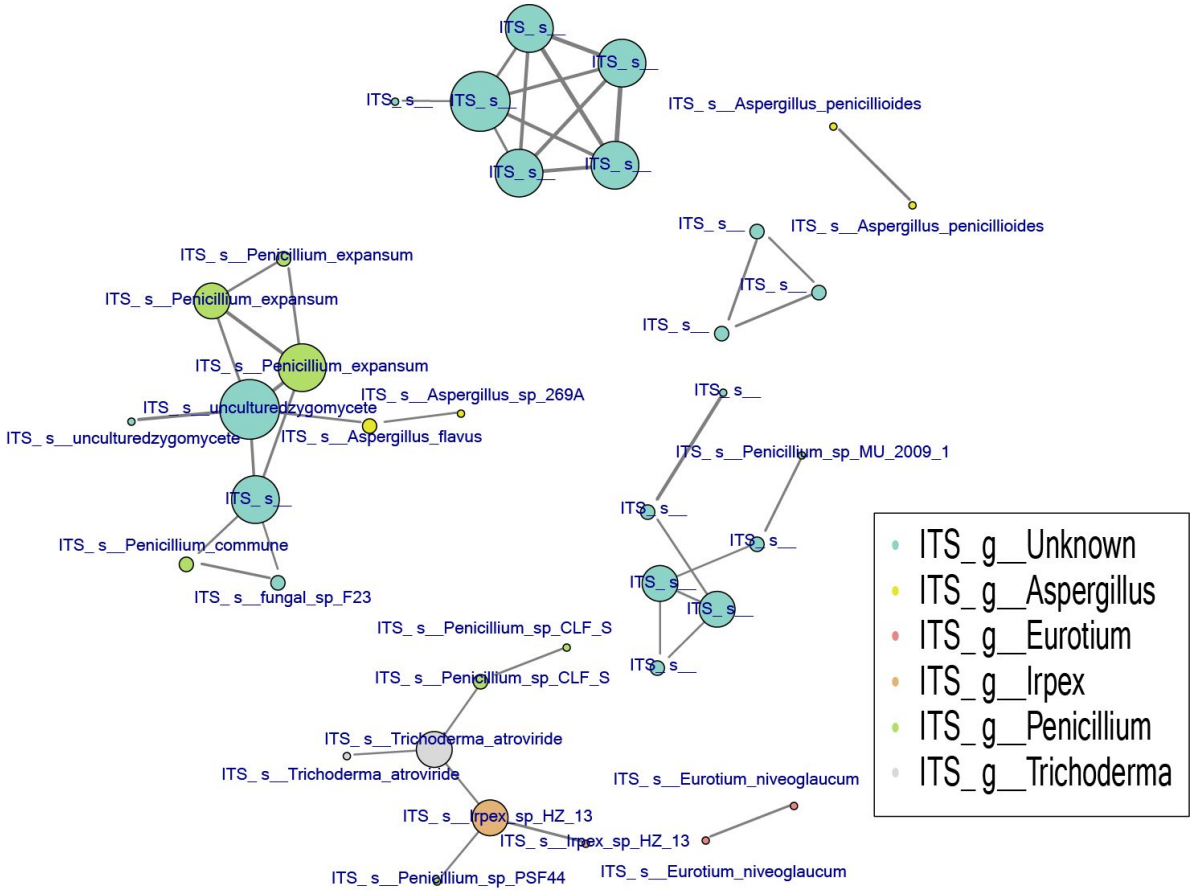

Figure 21: Fungi-Fungi co-occurrence network. Co-occurrence network (from n=91 fungi samples) shows highly correlated fungi forms mostly monophyletic clusters.

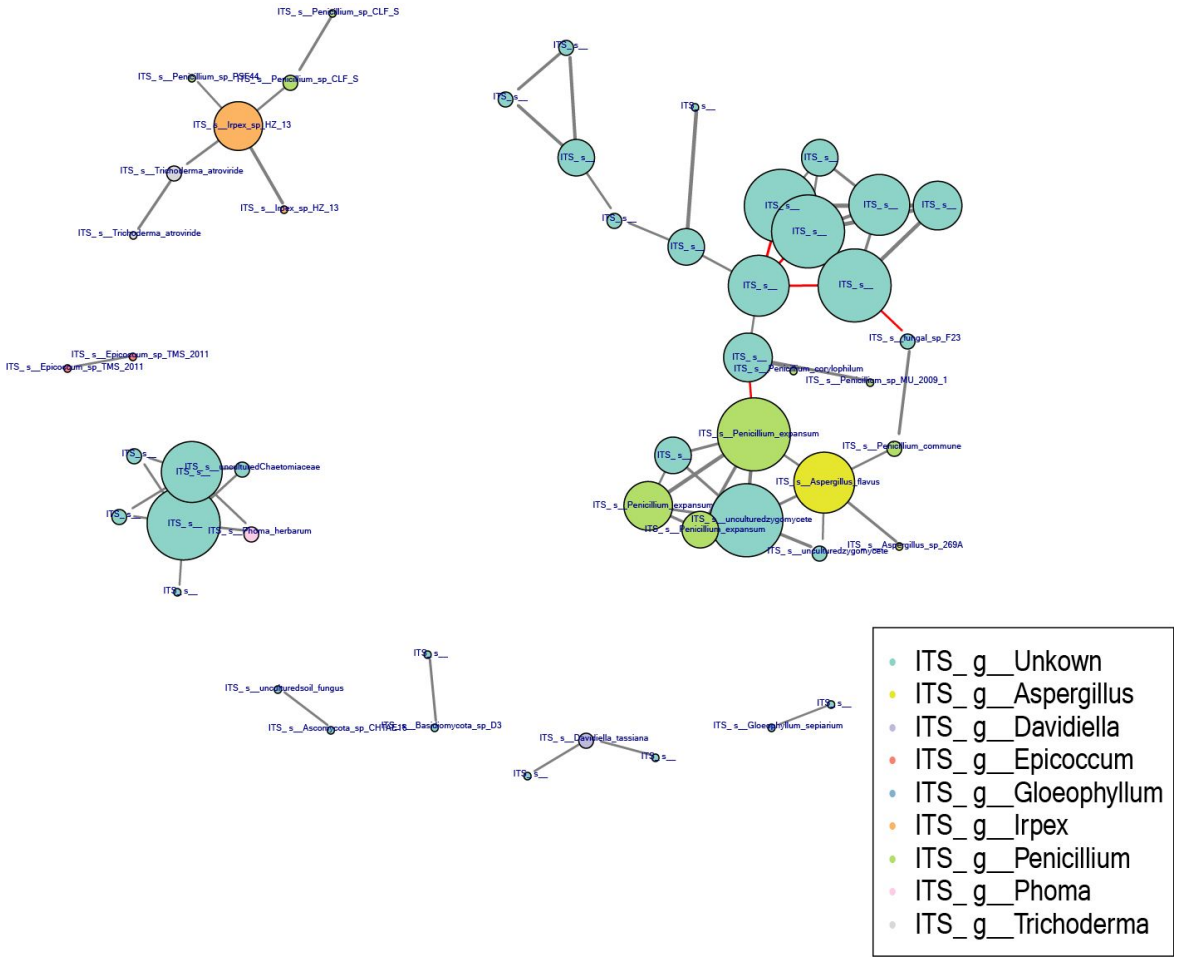

Figure 22: Fungi-Fungi co-occurrence network on wet samples. Co-occurrence network (from n=58 wet samples) shows how certain Fungi OTUs are anticorrelated on wet samples.

To better understand the co-associations between bacteria and fungi, 16S and ITS OTUs were co-correlated in a single network. A random walk-based method uncovered four distinct modules within the network, with modularity of 0.45 (Figure 23A). In general, the taxa present in each sample tended to cluster within an individual network module (median sample association to module = 0.88). We correlated various metadata factors to module membership (Figure 23B) and observed that wetting condition had a significant impact on which samples dominated each module: modules 1 and 3 were associated with wet samples, while modules 2 and 4 were

associated with non-wetted samples (Figure 23C). Location 1 samples dominated module 3, while Location 2 samples dominated module 1 (Figure 23D). Overall, the wetting condition appears to be the most important factor driving community succession, resulting in two different community structures even when the source community is identical. We also visualized the nodes that were assigned to the genera previously discussed in Figure 3. Nodes in the bacterial genera *Bacillus, Pseudomonas*, and *Erwinia*, as well as the fungal genus *Penicillium*, were nearly exclusively enriched in the two wet-associated modules (1 and 3), while chloroplast reads and *Eurotium* nodes all clustered within the non-wetted modules (2 and 4; Figure 23E**)**.

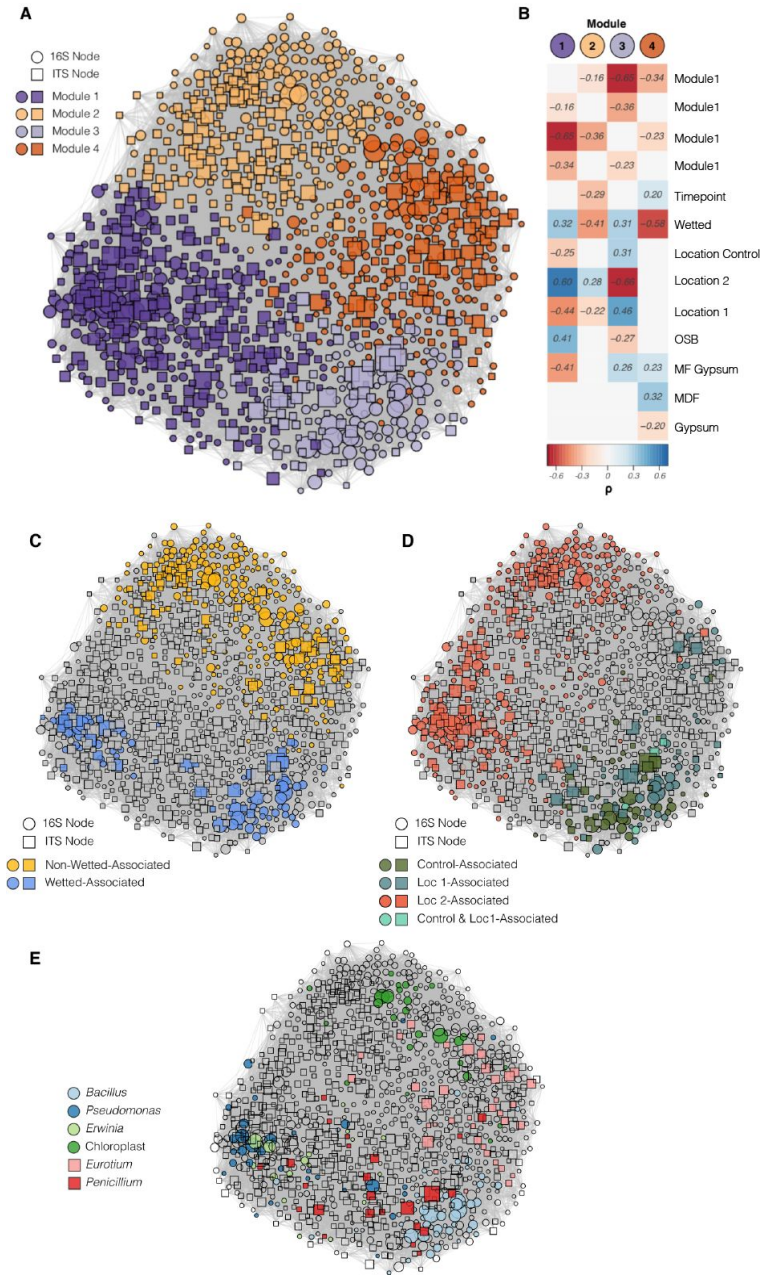

Figure 23: Network of SparCC OTU correlations. (A) Edge-weighted, spring-embedded network ordination, with nodes colored by module membership. Node shape represents node type (16S or ITS) and node size is based on the log-transformed abundance of each node (n=153 with both 16S and ITS, respectively). (B) Correlations between metadata factors (treated as dummy variables where true = 1 and false = 0) and the percent of reads in network modules. Non-significant correlations are not shown. (C) Taxa enriched in wet or non-wetted samples, as determined through a two-sided non-parametric t-test with 10^5 permutations. (D) Taxa enriched in samples originating from an individual inoculating location, with statistical methods as in (A). (E) Taxonomy of nodes in the genera included in Figure 17.

3.3.6 Metabolite network co-occurrence

A co-occurrence network correlation was calculated for the sample metabolite profiles (Figure 24**)**. As these data are not compositional, we built this network using significantly positive Spearman correlations between nodes and included only the 1,000 most abundant metabolites in the dataset. This resulted in a network with 149,316 edges (density = 0.30) when the significance threshold (alpha) was set to 0.001. Using the same module discovery method described above, we uncovered 7 distinct modules (modularity = 0.32), excluding 12 metabolites around the periphery of the network that clustered into modules of <5 nodes. Three modules (3, 4, and 7) were significantly correlated with wet samples, while modules 1, 2, 5 and 6 were associated with non-wetted samples. There was almost no correlation between network modules and inoculating locations, further suggesting that while location (and hence the primary inoculating microbes) may influence community taxonomic diversity, it does not appear to strongly affect metabolic diversity during growth. The abundance of metabolites in module 7 were anti-correlated with all other modules, but specifically with module 2 (corr =-0.87, p<0.001). Module 7 is dominated by wet samples at later time points, suggesting that community succession in wet environments may converge to a common metabolic profile, which is wholly distinct from the non-wetted samples in module 2.

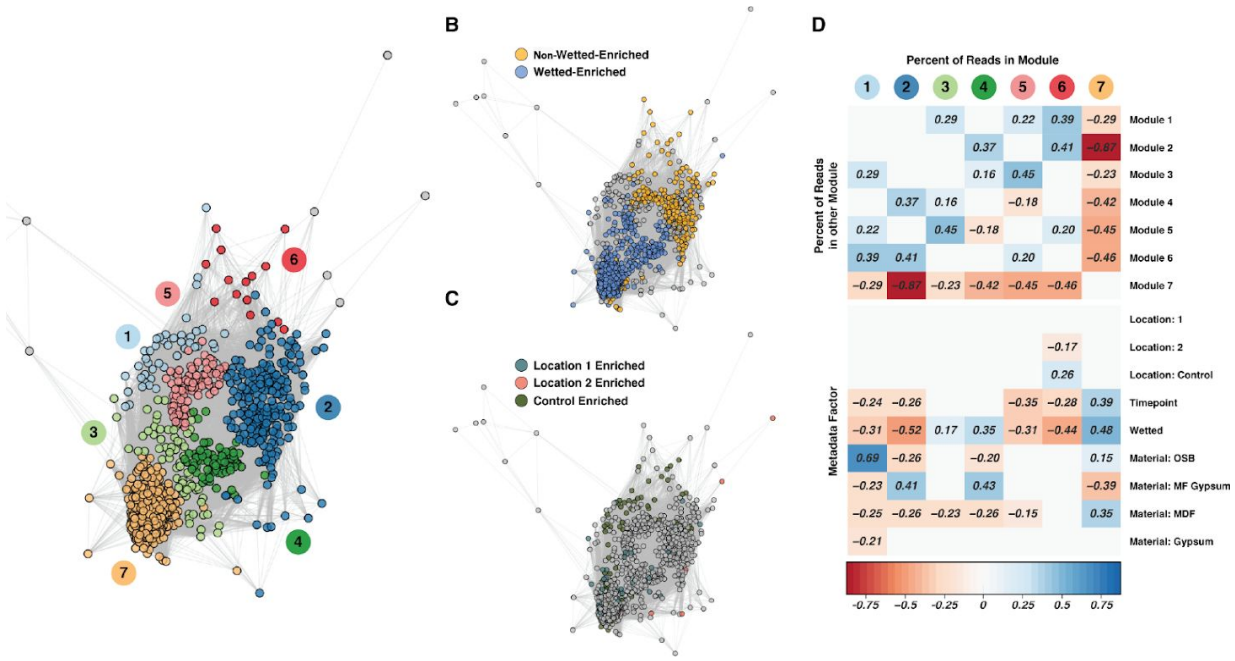

Figure 24: Metabolite network co-occurrence. (A) Network of significantly positive Spearman correlations between metabolites, with the network module indicated by color (n=144 samples). (B) Metabolites enriched in wet or non-wetted samples, as determined through a two-sided non-parametric t-test with 10^5 permutations. (B) Metabolites enriched in samples originating from an individual inoculating location, with statistical methods as in (B). (C) Correlations between metadata factors (treated as dummy variables where true = 1 and false = 0) and the percent of metabolites in network modules. Non-significant correlations are not shown.

3.3.7 Metabolite features can predict sample type

Random Forest analysis was employed to determine the metabolites associated with various sample types. Models classifying whether a sample was wet had an average accuracy of 98% (error ratio = 25, with expected random error 0.5), and wet-samples were never misclassified as a non-wetted sample in any of the 10 model iterations. Models classifying samples based on the material were similarly successful, with an average accuracy of 97% (error ratio = 25, with expected random error 0.75). Metabolomics models were much less successful at predicting the inoculating location, with a mean success of 72% (error ratio = 2.36 with expected random error 0.67). We sought to gain insight into the chemical composition of metabolites that comprise the

signatures observed in these models. Feature importance scores were assigned to compounds based on their relative contributions to predicting sample type. For both the wetting condition dependent and material dependent groups, we selected the 100 highest-scoring metabolite features for further examination and chemical identification (Figure 25). Wet samples were enriched with 98 of the 100 top-scoring metabolites that differentiated wet and non-wetted. None of these compounds were automatically identified by mzCloud, so the metabolites were analyzed via external database searches, and compound classes were designated based on fragmentation spectra. A diverse set of compound families was observed, including compounds likely to be carbohydrates and glycoconjugates, fatty acids, prenol lipids, sterol lipids, polyketides, and glycerolipids, as well as several pyridine derivatives including a form of vitamin B6, indicative of microbial activity and growth and compounds associated with the surface materials.

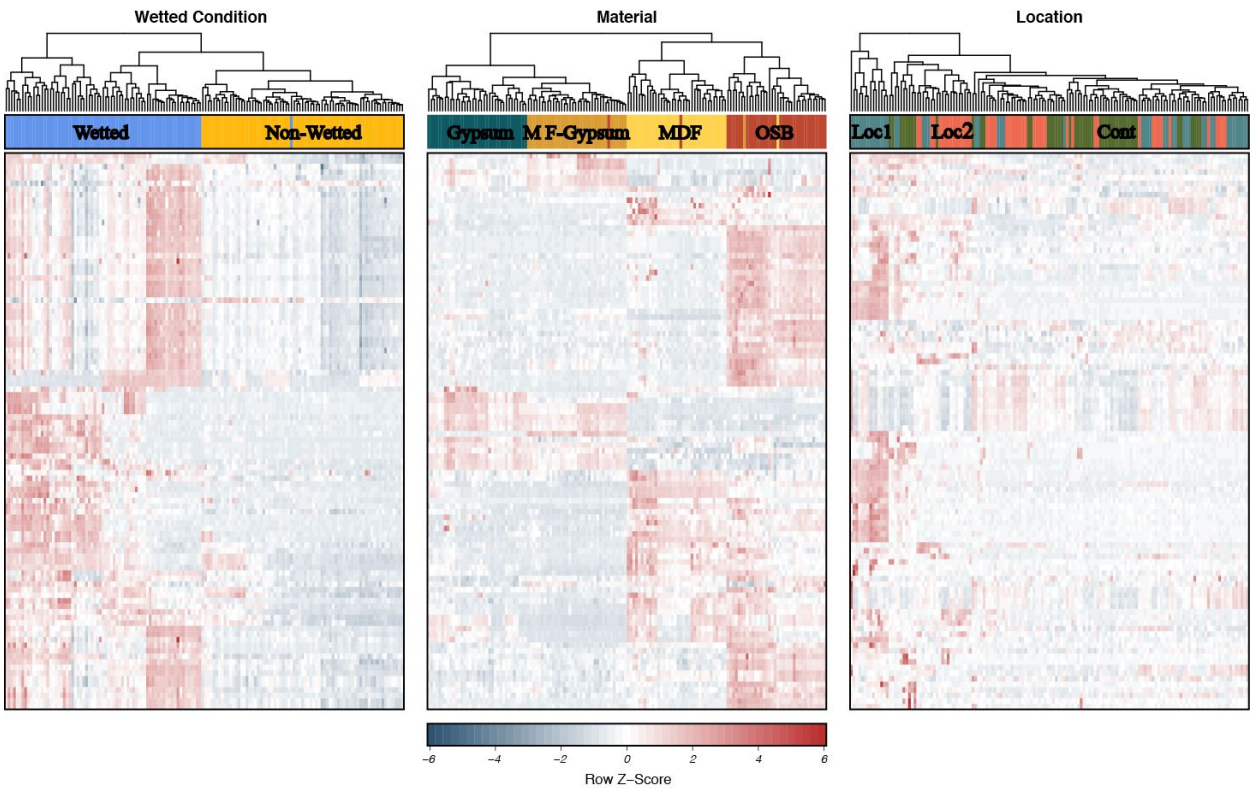

Figure 25: Random forest metabolite selection heatmap. Random forest learning was used to select the metabolites that most distinctly identify each environmental condition, wetted or non-wetted, wood material type and inoculation location (n=144 samples and 3187 metabolites), the 100 highest-scoring metabolite features for each condition were selected for further examination.

Metabolites that were highly enriched in wet vs non-wetted conditions underwent additional manual analysis for confident structural identification. One of these metabolites was identified as nigragillin ($C_{13}H_{22}N_2O$, accurate mass = 222.1723), which is a fungal alkaloid first identified in *Aspergillus niger* (Isogai *et al.*, 1975). Nigragillin abundance was significantly enriched in wet MDF and OSB samples (505- and 280-fold, respectively) compared to non-wetted samples. However, no significant differences in nigragillin were observed for gypsum or MF. In both wet MDF and OSB the nigragillin concentration increases over time (Figure 26). Another high-scoring metabolite showed MS/MS fragmentation consistent with fumigaclavine C

($C_{23}H_{30}N_2O_2$, accurate mass = 366.2291), which is a fungal alkaloid first identified in *Aspergillus fumigatus* (Cole *et al.*, 1977). Fumigaclavine C was enriched in wet samples of gypsum, MDF, and OSB (23-, 26-, and 13-fold increase in comparison to non-wetted samples, respectively), with equivalent abundance in mold-free gypsum regardless of wetting. While the concentration of fumigaclavine C remained flat or increased slowly in most materials, wet gypsum showed a dramatic increase in abundance at TP3 and 4 (15 and 20 days post incubation, respectively) (Figure 26).

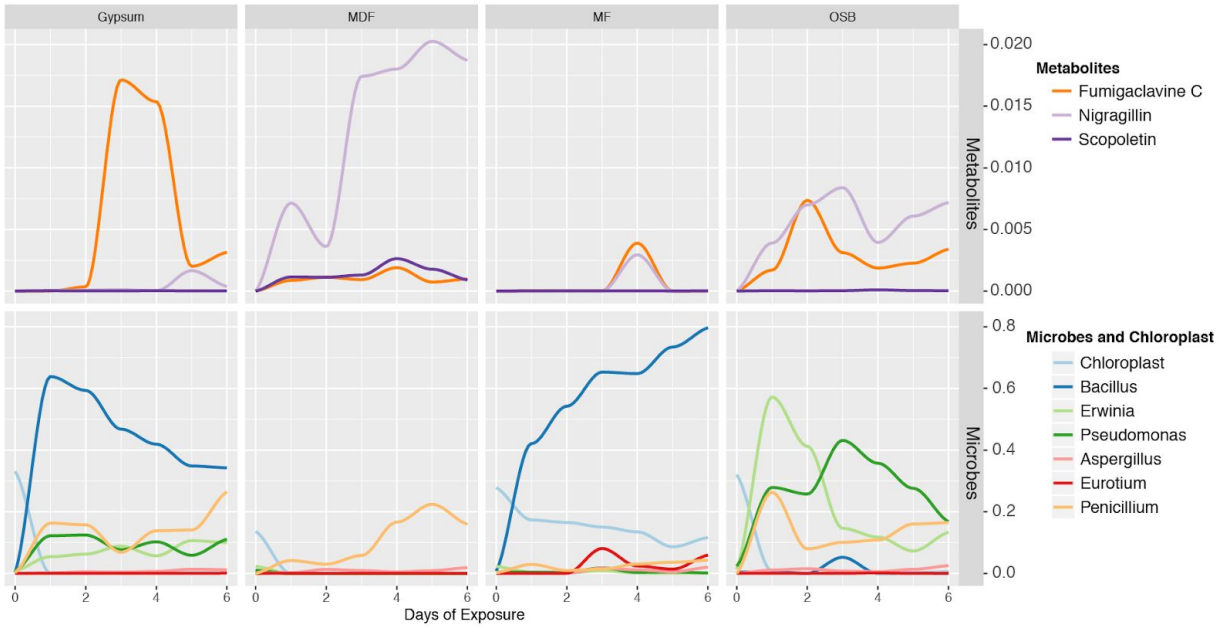

Figure 26: Metabolite and microbial succession on wet samples by material over time. Changes in the relative abundance of selected bacterial genera for each material over the course of succession (n=176, 218 and 72 wet samples for bacteria, fungi, and metabolites, respectively). Lines represent a smoothed moving average of the mean. Genus and metabolites are indicated by different colors.

Metabolites that were predictive of material type (OSB, MDF, gypsum, and MF) were also further analyzed to determine how these materials influence the chemical composition of metabolites. Of these metabolites, 80% eluted with a retention time of >7 minutes, indicating a

skew toward more hydrophobic compounds. This suggests that hydrophobic compounds are more diverse between the materials and therefore could have a greater influence on microbial metabolism than the ubiquitous hydrophilic components. Two of these metabolites were identified by MzCloud search: glucose-phosphate, which was about 10-fold less abundant in MF compared to all other materials, and scopoletin, a metabolite produced by plants that has antimicrobial activity (Lerat *et al.*, 2009, Gnonlonfin *et al.*, 2012, Nascimento *et al.*, 2013) which was about 60-fold more abundant in MDF samples than in other materials and could be influencing the reduced bacterial growth on this material (Figure 26**)**. Thiabendazole and azoxystrobin, known antifungal compounds (Clausen & Yang, 2007, Balba, 2007), were highly overrepresented on MF, 333 and 595-fold respectively more abundant than the average content for the other three materials, and as such are likely some of the active compounds in MF.

3.3.8 Microbe-metabolite co-occurrences

The abundances of nigragillin and fumigaclavine C were each significantly positively correlated with a fungal OTU annotated to the phylum Ascomycota (corr = 0.66, FDR p = 0.0004), which contains species known to produce these two alkaloids. Both nigragillin and fumigaclavine C have been reported to display antibacterial activity (Magdy *et al.*, 2017 & Pinheiro *et al.*, 2013). Interestingly, nigragillin was negatively correlated with the abundance of *Bacillus* and *Pseudomonas* OTUs; this could suggest fungal competition for space and resources (Mille-Lindblom *et al.*, 2006) against bacteria, and in the specific case of MDF, when nigragillin abundance was greatest no bacterial growth was detected (Figure 13 and Figure 26**).** The abundance of glucose-phosphate was significantly correlated to the proportion of a dominant

Enterobacteriaceae OTU, a genus which is known to synthesize it (Herter *et al.*, 2006, corr = 0.72, FDR p = 0.000002). Thiabendazole was positively correlated with *Penicillium* abundance (corr = 0.80, FDR $p < 10^{-9}$). As thiabendazole is prevalent and persistent in the natural environment, this correlation may indicate the presence of thiabendazole-resistant *Penicillium* strains colonizing the material from the built environment (Holmes & Eckert, 1999).

Co-occurrence networks were constructed between the bacterial OTUs and metabolites (SparCC correlation of >0.4; Figure 27) to explore further specific microbe-metabolite associations and possible mechanistic interactions. Significant correlations were observed between *Bacillus* OTUs and a number of different lipid classes which have been previously implicated in either formation or disruption of biofilms (Diomande *et al.*, 2015, Dubois-Brissonnet *et al.*, 2016, Kong *et al.*, 2014, Jensen *et al.*, 2014). In addition, azoxystrobin as well as several lipids were positively correlated with *Bacillus* and negatively correlated with *Pseudomonas*, whereas scopoletin was positively correlated with *Pseudomonas* and negatively correlated with *Bacillus* (Figure 28**).** These additional antagonistic compound interactions between *Bacillus* and *Pseudomonas* could represent either competitive interactions between these organisms or different adaptation to the different materials and wetting conditions.

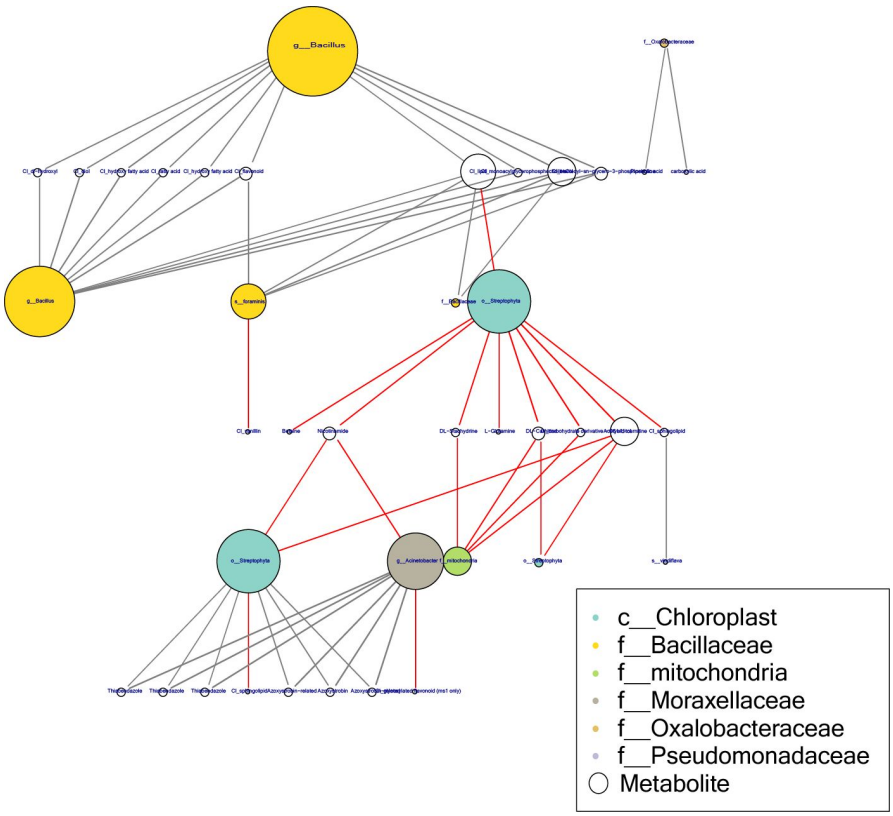

Figure 27: Bacteria-Metabolite co-occurrence network. Bacteria and Metabolite paired co-occurrences suggesting biochemical exchanges (from n=83, 144 samples, respectively). Lipid and hydroxyl compounds are strongly connected to Bacillaceae groups. Some specific lipids correlate positively with bacteria and negatively with wood material (plants). Vitamins and small carbon compounds negatively correlated with wood material (plants).

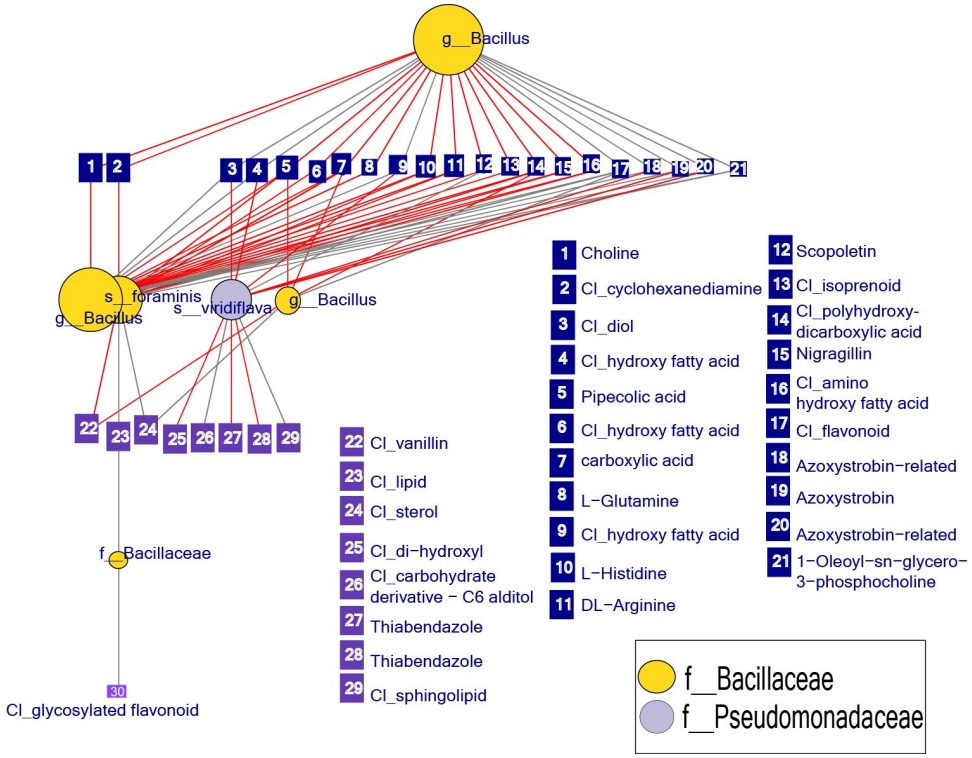

Figure 28: Bacteria-Metabolite co-occurrence network for Bacillaceae and Pseudomonadaceae interactions only. Bacteria and Metabolite paired co-occurrences suggesting biochemical exchanges (from n=83, 144 samples, respectively). Nigragillin is negatively correlated with both *Pseudomonas* and *Bacillus.* Azoxystrobin correlates negatively with *Pseudomonas*, but positively with *Bacillus*. Hydroxyl compounds correlate negatively with *Pseudomonas* but positively with *Bacillus*.

3.4 Discussion

As expected, wetted materials had higher bacterial and fungal growth rates and were dominated by a few particular microbes, most notably the bacterial genera *Bacillus*, *Erwinia*, and *Pseudomonas* and the fungal genera *Eurotium* and *Penicillium*. This dominance led to an overall lower alpha diversity compared to non-wetted coupons. Wetting condition and material type described the majority of the variance in bacterial, fungal and metabolite structure. Interestingly, each wetted material showed its own unique microbe-metabolite dynamics.

Gypsum and MF were mostly colonized by *Bacillus*, with gypsum being a less selective environment, which allowed for several bacterial species to thrive on the same coupon simultaneously, each of them with high relative abundance and apparently sharing both the physical space and resources. In contrast, MF prevented most fungal growth and allowed *Bacillus* to dominate with little competition. MDF selected for fungal growth primarily, which allowed for the rapid accumulation of the antibacterial chemical, nigragillin, which is known to be made by the *Aspergillus* fungi. On OSB material, nigragillin and fumigaclavine C, a second fungal-synthesized antibacterial metabolite, may play important roles in microbial growth dynamics. Nigragillin, fumigaclavine C, and *Aspergillus* relative abundance each gradually increases over time, whereas the abundance of *Pseudomonas* declines after the antibacterial metabolites reach peak abundance (Supplementary Figure 11). These observations bolster our hypothesis that production of antibacterial metabolites by *Aspergillus* may inhibit the proliferation of surrounding bacteria. Also, there is a human health risk associated with the proliferation of the *Aspergillus* fungi in the BE. While the most common species identified in our data was *Aspergillus penicillioides*, a common indoor fungus in damp buildings with known associations to allergies and asthma (Edwards *et al.*, 2012, Hay *et al.*, 1992), other *Aspergillus* species are known to be able to produce mycotoxins (including aflatoxins), molecules that have been associated with cancer and immunosuppression on humans (Roze *et al.,* 2013). Of course, our correlation-based analyses do not definitively establish interactions between taxa or the origins of individual metabolites. Still, we believe these insights will be useful in generating testable hypotheses for future, more specifically-targeted studies.

3.4.1 Mold-resistant gypsum

Traditional wood-based building materials contain natural polymers such as cellulose and lignin that are susceptible to degradation by fungal colonization (Gravesen *et al.*, 1999, Pasanen *et al.*, 1992). With some fungi having been shown to produce mycotoxins including aflatoxins that could affect human health (Roze *et al.*, 2013, Rand *et al.*, 2017), building materials such as mold-resistant gypsum have been developed, which contain antifungal compounds intended to discourage fungal growth. We were particularly interested to examine the microbial communities on these surfaces and as expected, found that fungal growth was diminished on MF compared to other materials. However, it appeared that the scarcity of fungal colonies made way for bacterial species to flourish with less competition; on non-wetted materials we observed MF bacterial particle counts greater than on the other three materials, and on wetted materials while the MF bacterial counts were second to MDF, the abundance level between non-wetted and wetted coupons, unlike MDF, remained minimally changed. This raises the potential that pathogenic bacteria colonization could occur on MF and if wetted could grow and lead to negative health outcomes. In terms of metabolite production, thiabendazole and azoxystrobin were some of the antifungal compounds found in high abundance and overall a similar subset of compounds accounted for most of the metabolite abundance on this material, indicating lower metabolic diversity when the colonization is dominated by bacterial growth. We also detected a correlation between thiabendazole and *Penicillium*, which suggested the persistence of thiabendazole-resistant fungal strains.

3.4.2 Diversity and interaction between microbes and environment

Additionally, certain lipid metabolites (indicative of biofilm formation) showed significant positive correlation with both *Bacillus* and *Pseudomonas* OTUs, and these lipids were negatively correlated with the abundance of chloroplast OTUs, indicating that when bacteria and metabolites indicative of biofilm formation are detected in greater abundance, we see a proportional reduction in plant-associated signal. Similar to the lipids, metabolites annotated to organic molecules and vitamins were also negatively correlated with the chloroplast OTUs, which suggests that bacterial growth, indicated by increased proportion of 16S, cellular counts and associated metabolites, tends to swamp out the background material-chloroplast signal. We hypothesize that this may be because these molecules are being produced by bacteria colonizing and forming biofilms on the woody material. When the relative abundance of the bacteria increases, it reduces our ability to detect chloroplast sequences (based on a given sequencing depth); as such this negative correlation is likely due to the increased abundance of the microbes that mediate the production of these metabolites, reducing the detection frequency of specific chloroplast OTUs, and not due to some mechanistic relationship between the wood and these molecules.

3.4.3 Microbial-metabolite interactions

Pseudomonads and *Bacillus* are often the main contributors to biofilm formation on material surfaces in the built environment (Ronan *et al.*, 2013; Powers *et al.*, 2015; Raaijmakers *et al.*, 2010). Biofilms are complex extracellular matrices formed by bacteria through the excretion of lipopeptide biosurfactants, to provide attachment to a surface to support colocalization with a

nutrient source and protection from dehydration and chemical activity. Some of these lipopeptide biosurfactants produced by *Pseudomonas* and *Bacillus* species have been shown to have lytic or growth-inhibitory activity against many microorganisms such as bacteria, viruses, mycoplasmas, and fungi (Raaijmakers *et al.*, 2010). Powers *et al.* (2015) demonstrated that *Pseudomonas* protegens produces antibiotics that inhibit biofilm formation and sporulation in *Bacillus subtilis*. They also found that *Pseudomonas* putida secretes an unknown inhibitory compound that prevented biofilm-associated gene expression. In our study we demonstrate a number of compounds known to have potential biofilm inhibitory qualities that also co-correlate with either *Pseudomonas* or *Bacillus* abundance, suggesting potential competitive activity between these organisms. While *Pseudomonas–Bacillus* interactions have been shown to be competitive, interspecies interactions within the genus *Bacillus* are also important in the formation of biofilms, lipids like hydroxy fatty acids and mono-acyl-glycerophosphocholines could be building blocks or residual products of the biofilm creation (Shank *et al.*, 2011, Diomande *et al.*, 2015, Dubois-Brissonnet *et al.*, 2016).

3.4.4 Closing remarks

The simultaneous collection of environmental, metabolomic and microbial profiles reveals insights into the chemical signals that may govern BE microbial communities under high humidity conditions, as well as providing evidence that these the membership compete for space and resources. Here we show that wetting condition can profoundly alter both fungal and bacterial community succession, and that the taxa which dominate samples after wetting or exposure to high humidity are not abundant in non-wetted materials and have little relation to the

skin-associated taxa which dominate samples of indoor environments. After wetting, the microbial community undergoes a successional trajectory that can result in convergence of metabolic diversity even when taxonomic diversity remains variable. We further show that while material choice significantly influences bacterial diversity, the same is not true of fungal diversity. In summary, BE microbial ecology once seen as a wasteland (Gibbons, 2016) could rather be seen as a desert environment mostly formed with smaller assemblages that can rapidly become an active ecologically dynamic community if water, in liquid or vapor form, is added. When a material experiences high moisture conditions, both fungal and bacterial growth rapidly accelerate and the metabolites associated with their adaptation to different surface materials and competition for resources demonstrate ready-made eco-evolutionary adaptation to this sporadic availability of a crucial resource; this phenomenon is very similar to what has been observed in real desert soil microbiomes (Neilson *et al.*, 2017), as well as in very different ecosystems, such as sediments exposed to oil pollution (Handley *et al.*, 2017).

3.5 Materials and methods

3.5.1 Test materials

Four building materials were used in this study: oriented strand board (OSB), medium density fiberboard (MDF), regular gypsum wallboard, and mold-resistant (i.e., mold-free, or 'MF') gypsum wallboard. All samples were purchased new from a home improvement store in Chicago, IL. The building materials were cut into 5 cm × 5 cm coupons for testing. The coupons were sterilized by UV irradiation for 20 minutes followed by surface cleaning with a 70%

ethanol solution. This approach likely did not render the coupons DNA free but certainly non-viable.

3.5.2 Inoculation

The material coupons were naturally inoculated by passive settling for about 50 days each at one of three locations: two private residences ("location 1" and "location 2") and in a laboratory where they were subsequently incubated ("control"). The "control" coupons were kept covered by aluminum foil and kept in the laboratory for the same duration to minimize natural inoculation by deposition. These control samples were initially treated as a unique location in our analyses and can be thought of as "lab-inoculated" (albeit with minimal environmental influence) rather than "residence-inoculated." Each set of test coupons included 44 coupons for each type of building material (i.e., 176 coupons in total) to allow for multiple subsequent sampling strategies. One set of test coupons was placed inside a 6th floor apartment unit with two adult occupants and a medium sized dog located in downtown Chicago, IL (Location 1). The other set of materials was placed inside a 2-story single-family residence without any pets near the main campus of Illinois Institute of Technology, approximately 8 km south of the downtown residence (Location 2). During the inoculation periods, built environment metadata (Ramos & Stephens 2014) were collected in each residence, including temperature (T) and relative humidity (RH) using Onset HOBO U12 data loggers and occupant presence within the immediate vicinity of the samples using Onset UX90 data loggers. The UX90 occupancy sensors were placed on the floor next to the samples, facing up, to record movements within the sensor's field of vision as an indicator of how often occupants were in close proximity to the surfaces, which may have

affected natural inoculation through direct human shedding. Coupons at a third location (the Built Environment Research Laboratory at the of Illinois Institute of Technology) were covered with aluminum foil to minimize natural inoculation, serving as a control group.

3.5.3 Wetting and incubation

After inoculation, half of each set of materials (i.e., 22 coupons each) from each location, as well as 22 coupons from the control group, were submerged in sterilized tap water in separate pans for ~12 hours to simulate the process of wetting of building materials by potable water. The other half of each set of materials (i.e., the other 22 coupons each) from each location and the other 22 coupons from the control group were not submerged in water. Next, just about 10 minutes after the submersion, to encourage fungal growth on all of the building materials, all of the coupons were placed in trays (each tray contained all 22 coupons of one type of material from one location or control group) and were incubated at room temperature (24 ± 2.7°C) inside a static airtight chamber (0.9 m × 1.2 m × 0.4 m). Salt solutions (potassium nitrate) were used to maintain constant high RH near ~94% for the duration of the experiment. Constant high RH may not accurately reflect realistic building conditions but was used to encourage growth and has been used in many prior investigates (e.g. Johansson *et al.*, 2012; Hoang *et al.*, 2010). Temperature and RH in the chamber were also recorded using Onset HOBO U12 data loggers.

3.5.4 Sampling procedures

The building material coupons were sampled for off-line biological and chemical analysis every 5 days at 7 different sampling time points, referred to here as: TP0 (immediately prior to incubation), TP1 (5 days after incubation), TP2 (10 days), TP3 (15 days), TP4 (20 days), TP5

(25 days), and TP6 (30 days). The initial samples (TP0) were taken just after retrieving the samples from the field inoculation and represent non-wetted, naturally inoculated samples previously held at normal indoor environmental conditions. The remaining sampling time points occurred every ~5 days. At each time point, a new coupon of each material from each condition that had never been swabbed before was swabbed, while duplicates of previously un-swabbed samples were also swabbed periodically (at TP0, TP2, TP4, and TP6) for comparison. Two samples ('TP0' and 'TP0 duplicated') were also swabbed at every time point to investigate whether repeatedly swabbing the surfaces impacted the results. Duplicates of both previously swabbed and previously un-swabbed samples were also included to investigate whether or not natural inoculation and subsequent growth was evenly distributed across multiple coupons. Figure 10 illustrates the experimental setup and Figure 29 shows coupons' photographs at TP5 and TP6 for each one of the three locations. Details of swabbing procedure at each time point are described below.

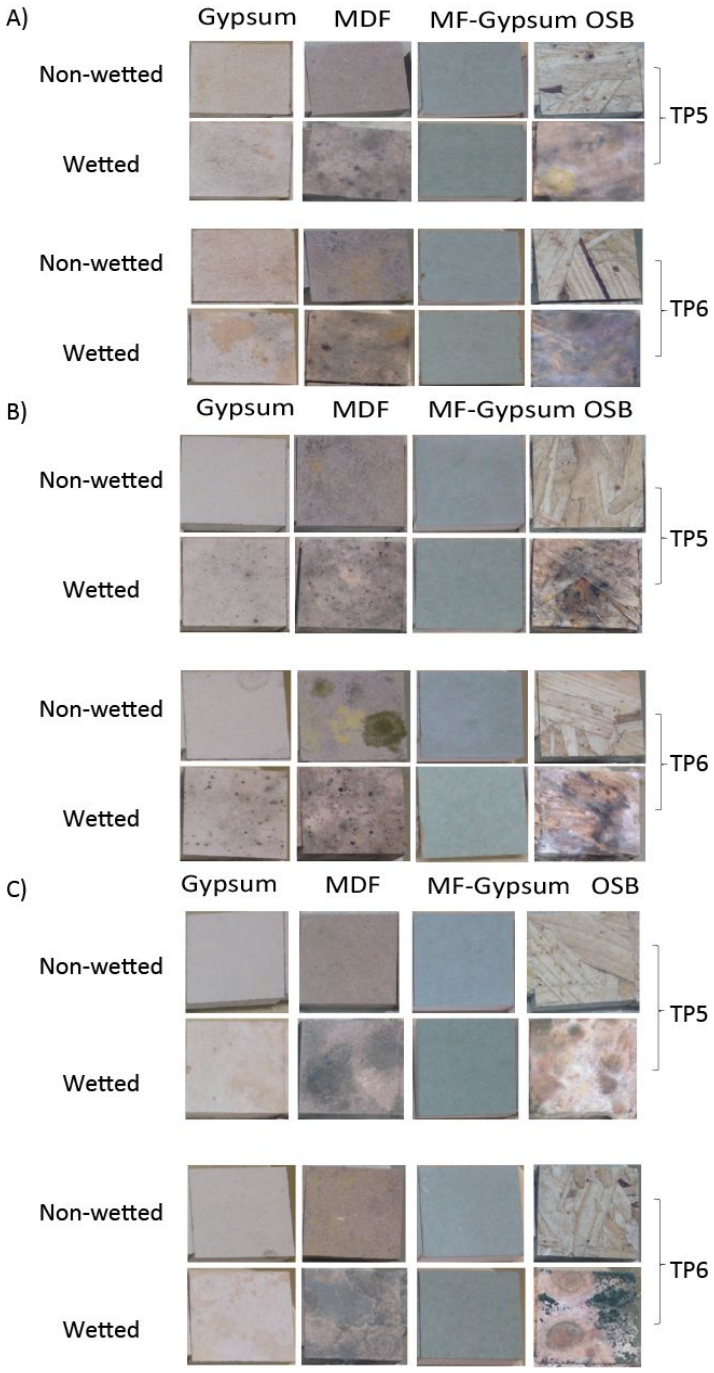

Figure 29: Photographs of wood coupons from different materials and wetting conditions at TP5 and TP6. Bacterial and Fungal growth on coupons surface photographs for (A) location 1 (B) location 2 and (C) control location.

First, sampling reagents were prepared as follows. Phosphate-buffered saline (PBS) was used for microbial samples that were to be analyzed for DNA and formaldehyde was used for microbial samples that were to be analyzed by microscopy. For PBS, 500 µl 1X PBS was added to 1.7 ml microtubes for each sample to be collected. For formaldehyde, 100 µl 4% paraformaldehyde was added to 1.7 ml microtubes for each sample. Microcentrifuge tubes were filled with ethanol solution (200 µl 50% EtOH) to preserve samples for surface chemistry/metabolomics analysis.

For subsequent DNA sequencing and analysis, the entire surface of the test coupons was swabbed using two BD Screw Cap SWUBETM Polyester swabs for 20 sec. The same researcher swabbed every time, intentionally applying approximately the same pressure and swabbing in the same pattern to keep the swabbing process consistent. Although polyester swabs have been shown to be less efficient in recovering microbial biomass from surfaces than some other materials such as nylon (Rose *et al.*, 2004), they were used in a consistent manner and have been shown to perform well versus roller samplers and electrostatic wipes when controlling for the actual area sampled (Lutz *et al.*, 2013). Swabs also allowed for a minimal disturbance on our small surface area coupons compared to other sampling approaches. One of the double swabs was placed into the tube with PBS and frozen for shipping for subsequent sequencing. The tip of the other of the double swabs was placed into microtubes and the swab tips were vortexed for 10 seconds. 100 µL of sample buffer was removed added to the tube containing 100 µL 4% paraformaldehyde for fixation. These fixed samples were stored in a refrigerator at 4°C and then sent to the San Diego State University team for running numerical counts of cells and virus particles using microscopy.

For surface and microbial chemistry analysis (i.e., metabolomics), another test coupon was swabbed using a cotton-tipped applicator that is dipped in ethanol (Petras *et al.*, 2016). The end of the swabs were cut directly into pre-prepared collection tubes, stored at 4˚C for 2-3 hours, and then stored at -20˚C overnight. Swabs were then removed with clean forceps the next morning, then re-sealed into microcentrifuge tubes and sent to the Northwestern University team on ice at -20˚C or lower. Overhead photos of each tray of coupons were also taken at each time point for image analysis using ImageJ to calculate the percentage of visible microbial growth coverage (Hoang *et al.* 2010).

3.5.5 Viral-like particle and bacterial microscopy counts

Epifluorescence microscopy was used to ensure that all samples contained bacteria and virus-like particles and to estimate their abundance. 100 µL of the paraformaldehyde-fixed samples were resuspended into 5 mL of sterile 0.02 µm filtered water. Each suspended sample was then filtered onto a 0.02 µm Whatman Anodisc filter membrane (Thurber *et al.*, 2009). The filters were stained with 1X SYBR Gold and incubated for 10 minutes in the dark. Each filter was washed and mounted onto slides to be observed. Visualization was performed using a QImaging Retiga EXi Fast Cooled Mono 12-bit microscope and Image-Pro Plus software was used to collect digital images and estimate VLP and bacterial abundances.

3.5.6 Metabolomics analysis

Samples were analyzed by High-Performance Liquid Chromatography and High-Resolution Mass Spectrometry and Tandem Mass Spectrometry (HPLC-MS/MS). Specifically, the system consisted of a Thermo Q-Exactive in line with an electrospray source and an Agilent 1200 series

HPLC stack including a binary pump, degasser, and autosampler, outfitted with a column (Waters XBridge BEH Shield RP18, 100x2.1 mm, 5µm particle size with matching guard). The mobile phase A was H_2O with 0.1% Formic Acid; B was Acetonitrile with 0.1% Formic Acid. The gradient was as follows: 0-0.5 min, 98% A; 5 min, 80% A; 10-10.5 min, 5% A; 10.6-15 min, 98% A, with a flow rate of 400 µL/min. The capillary of the ESI source was set to 275 °C, with sheath gas at 40 arbitrary units and the spray voltage at 4.0 kV. In positive polarity mode, MS1 data were collected at a resolution of 35,000. The precursor ions were subsequently fragmented using the higher energy collisional dissociation (HCD) cell set to 30% normalized collision energy in MS2 at a resolution power of 17,500. Data were processed with Compound Discoverer 2.0 (Thermo Fisher) with MS/MS metabolite identifications made by comparing experimental MS/MS spectra with library spectra from MZCloud (lower cutoff score of 90% match).

For the metabolites that were selected for more in-depth characterization, classification of structure or substructure was performed by searching databases such as the Dictionary of Natural Products, the LIPID MAPS Structure Database, and GNPS (Global Natural Products Social Molecular Networking). Predicted structures resulting from a matched intact mass (≤10 ppm error) were subsequently validated through manual analysis of fragmentation mass spectra. Metabolite differential abundances (fold calculations) were calculated from Compound Discoverer median peak areas for each compound including all three sampled locations.

3.5.7 DNA extraction and sequencing

To perform DNA extraction, the Qiagen DNeasy Powersoil HTP kit was used with a modified protocol optimized for low-biomass samples. Swab tips were inserted into each well of the bead plate, and then cut off using a sterilized wire cutter. The manufacturer's protocol was then followed, with the following modifications: before cell lysis, the bead plates (containing beads, bead solution, swabs, and the C1 solution) were heated for 20 minutes at 60°C in a water bath. Additionally, the protocol steps using solutions C2 and C3 were combined into a single step, by adding 150 µl each of C2 and C3 together to the lysed sample in the 1 ml plate.

The DNA obtained from the DNA extraction was used for both high-throughput 16S/ITS sequencing, and qPCR. The 16S sequencing targeted the V4 region of the bacterial 16S rRNA gene, using the primer pairs 515F/806R. The ITS sequencing targeted the highly variable fungal internal transcribed spacer region 1 located between the 5.8S and 18S rRNA genes, using the ITS1f and ITS2 primer pairs (Walters, 2015). Both primer sets used the same reaction mix and thermocycler instructions: Reaction mix: 9.5 µL of molecular biology grade H2O, 12.5 µL of Accustart II PCR Toughmix, 1 µL each of forward and reverse primers at 5 µM, and 1 µL of sample DNA for a total reaction volume of 25 µL.

To make both the 16S and ITS amplicons, the following PCR program was used: Initial denaturation step at 94°C for 3 minutes, followed by 35 cycles of: 94°C for 45 seconds, 50°C for 60 seconds, and 72°C for 90 seconds, followed by a final extension step of 72°C for 10 minutes. The resulting amplicons were quantified using the Picogreen dsDNA binding fluorescent dye on

a Tecan Infinity M200 Pro plate reader and pooled to 70 ng DNA per sample using the Eppendorf epMotion 5075 liquid handling robot. Primers and PCR reagents were removed using Agencourt AMPure beads, and then the clean amplicon pool was sequenced at Argonne National Laboratory's Environmental Sample Preparation and Sequencing Facility, following the Earth Microbiome Protocol (Caporaso *et al.,* 2011). Sequencing was performed on an Illumina Miseq using V3 chemistry, generating 2x150nt reads. For the fungal sequencing, 2x300nt reads were generated by using additional cycles.

qPCR was performed using a Roche LightCycler 480 II. The 515F/806R primer pair was used again for amplification, using a mix of 10 µL Light Cycler 480 SYBR Green I Master mix, 6 µL of molecular biology grade H2O, 1 µL of 515F primer (10 µM), 1 µL of 806R primer (10 µM), and 2 µL of template DNA for a total of 20 µL per reaction. The following thermocycler conditions were used: (1) 95°C for 5 minutes, (2) 95°C for 10 seconds, (3) 45°C for 45 seconds, (4) Measure fluorescence, with steps 2 through 4 repeated 50 times. To determine the copy number of the 16S gene (and therefore the number of organisms per swab), a standard curve was generated using a serial dilution of a plasmid containing the *E. coli* 16S rRNA gene.

For fungal qPCR, each 20ul reaction contained 1uL each of the forward ITS1f (CTTGGTCATTTAGAGGAAGTAA) and reverse ITS2 (GCTGCGTTCTTCATCGATGC) primers at 10uM, 10ul of the mastermix, 6 ul of DNA-free water, and 2ul of the target DNA extraction. The target was amplified using the following conditions: Initial denaturation step for 15 minutes at 95C, followed by 45 cycles of a 95C denaturing step for 1 minute, an annealing step at 53C for thirty seconds, and an extension step of 72C for 1 minute. The amount of

amplified DNA was quantified after each extension step. A standard curve was constructed using the Zymo Femto Kit, which contains serially diluted DNA from Saccharomyces cerevisiae TMY18. Using this curve, the copy number per ul for each extraction was determined.

3.5.8 Treatment of technical replicates

We used the Mantel test to determine whether the bacterial communities on replicate coupons (coupons on the same tray sampled at the same time) significantly resembled each other and preserved the pattern of beta-diversity. We began by calculating the Bray-Curtis dissimilarity between each pair of samples taken from the same material type using the *beta_diversity.py* function from the software QIIME 1.9.1 (Caporaso *et al.*, 2010), producing dissimilarity matrices for each sampling type. Then the Mantel test and false discovery rate adjustment was performed using the *mantel* and *p.adjust* functions in the Vegan and stats R packages. For all comparisons between the sampling types the mantel statistic (which measures the stress in the fit of the two matrices) was significantly high (mantel >= 0.67 for fungi and >= 0.5 bacteria (all $p < 1E-05$) (Table 3 (Fungi) and Table 4 (Bacteria)). Based on the highly significant resemblance between material types, we treated all samples of the same type as technical replicates, meaning that all combinations of material, location, wetting condition, and time point had either 2 (time point 0), 3 (time points 1, 3, and 5), or 4 (time points 2, 4, and 6) replicates.

3.5.9 Rarefaction and statistical analyses

After sequencing and sample merging, bacterial and Fungi OTU tables were rarefied to 1,000 and 10,000 reads respectively for statistical analyses. Rarefaction and statistical analyses were performed using R.

3.5.10 Random forest analyses

Random forest models were implemented using the "randomForest" R package. Samples from time point 0 were removed from the dataset. Models were built with 1000 trees and 10-fold cross-validation. For each of the 10 models for each metadata criterion, a randomly drawn 70% of samples (100 samples) were used for model training and the remaining 30% (44 samples) were used for validation.

3.5.11 Non-metric multidimensional scaling (NMDS)

We visualized sample similarity using non-metric multidimensional scaling (NMDS) ordination based on Bray-Curtis dissimilarity. Metadata vectors were fit onto the ordination using the *envfit* command in the Vegan R package. We converted material, location, and wetting condition into dummy variables (1 = yes, 0 = no) and, in the case of the bacterial and fungal datasets; also fit vectors of relative abundance for the common genera described in Figure 14. We assessed the significance of each of the vectors using 10^5 permutations, and removed non-significant vectors from the figure. The R^2 values for each vector and their significance is presented in Supplementary Table 1 at Lax and Cardona *et al.*, 2019.

3.5.12 Co-occurrence networks

Traditional correlation networks are unsuited to genomic survey data as these data are relative, rather than absolute, measures of community composition. Since the relative abundances of all taxa within each sample must sum to 1, the fractions are not independent and will often exhibit negative correlations to each other regardless of the true correlation in absolute abundance. To avoid these compositional effects, we generated our networks using SparCC (Friedman & Alm, 2012), a correlation metric based on log-ratio transformed data that is specifically suited to compositional genomic surveys. Pseudo-p values for each correlation were generated through comparison from 100 to 1,000 bootstraps of the permuted OTU table.

For the same kingdom and microbe-metabolite networks only samples where either bacteria or fungi and metabolites were found in detectable levels after rarefaction was used (N=83, N= 91 respectively). Additionally, only bacterial OTUs with >9 reads, fungi OTUs with >99 reads, and metabolites with > 5'000,000 abundance in the rarified dataset were used for a total number of 630 bacterial OTUs, 352 fungi OTUs and 426 metabolites. Figures were generated using CAVNet R package (Cardona, 2017) and only displayed the higher correlation threshold (positive or negative) greater than 0.4.

For the network encompassing both bacteria and fungi, the OTUs reads threshold remained the same but only samples with both 16S and ITS data were kept (N = 153) producing a new subset of bacterial and fungi OTUs, 590 and 581 OTUs respectively. This dataset produced a co-occurrence network with 1,171 nodes. Only positive correlations with a pseudo-p < 0.05 were included, resulting in a network with 33,509 edges (density = 0.052). The network was ordinated

using the Fruchterman-Reingold Algorithm (edge-weighted, force-directed) in the *igraph* R package, with node size based on the log read count of each OTU across all samples (with ITS counts first divided by 10 to equalize rarefaction depth between datasets). We used the Walktrap method (Pons & Latapy, 2005) to uncover dense subgraphs (modules) within the network, which may correspond to distinct community structures. We chose Walktrap (which is based on random walks within the network) as our method of community inference due to its computational tractability and its accuracy at uncovering subgraphs regardless of network size (Yang *et al.*, 2016). We used random walks of four steps, which resulted in four distinct modules with network modularity of 0.45.

CHAPTER 4

CO-OCCURRENCE AND METABOLIC NETWORKS TO EXPLORE ECOLOGICAL RELATIONSHIPS ON HIGH RELATIVE HUMIDITY BUILT ENVIRONMENT MICROBIOMES

4.1 Overview

Microbial ecology is a field that has heavily relied on statistics to identify relationships between microbial taxa or genes and different phenotypes or environmental conditions. Many times, these findings guide future hypothesis creation and point scientists to further experimentation with more computation, wet lab or analytical tools. Most recently, a few teams have started to incorporate metabolic models to their microbial ecology studies and that has accelerated the results being derived from statistics methodologies alone. Metabolic models provide a more detailed view of the mechanisms that are more likely driving the observed phenotypes or statistically determined correlations. However, integrating metabolic models into microbial ecology workflows is happening slowly, partly because metabolic models require fully annotated genomes as input data. Many projects do not yet have enough DNA biomass or financial resources to pay for shotgun metagenomics sequencing, and using reference genomes alone cannot provide useful details about a specific system.

Here we are proposing a new method to run metabolic models from inferred full genomes using only 16S rRNA sequence data inputs. The methodology relies on mapping the 16S rRNA sequence to a reference database and identifying a pangenome that is phylogenetically related to that sequence. Then the superset of genes present in the pangenome is used to build the corresponding metabolic models. A more advanced application of this new method is creating

community models, which just like the standard community metabolic models are a relatively simple aggregation of individual metabolic models. For these 16S rRNA based community models two parameters (F and K) need to be defined: F defines a minimum gene frequency cutoff and K the proportion of microbial activity supported per unit of biomass. Expanding our prior work on built environments under high relative humidity conditions (Lax and Cardona *et al.*, 2019), I created various 16S rRNA based community models from microbial OTUs seen in pre-wetted wood samples across 7 time points (TP0 to TP6) and three different materials: oriented strand board (OSB), gypsum wallboard, and mold-free gypsum wallboard (MF). Community model simulations showed that in our built environment controlled system, amino acids and amines are some of the preferred nutrition sources for *Bacillus* OTUs, while sugars and carboxylic acids seem to be preferred by *Erwinia* and *Pseudomonas* taxa. It also shows how cooperation between taxa with the same genus and competition from taxa in different genera tend to be derived from cross feeding or nutrient competition. These observations are in line with statistical and co-occurrence networks results found in our original work. I also analyzed the metabolomics dataset and found that, just like bacterial abundance, the chemical profile had a unique signature for each material. All three materials started from a similar chemical profile of about 50% amino acids and 50% carboxylic acids and sugars and then diverged in different directions. Gypsum maintained about equal abundances of amino acids, sugars and carboxylic acids; MF showed a slight decline of amino acids, followed by a large increase and a second decline; and OSB showed a consistent increase of amino acids and decrease of carboxylic acids and sugars. Looking at the chemical groups abundances variations (what chemical compounds are depleted over time), we observed that the compound consumption is in agreement with the

computationally suggested microbial feeding preferences and the observed microbial growth for each material. Supporting the argument that in our controlled growth environment *Bacillus* thrive when the environments are rich in amino acids, and *Erwinia* and *Pseudomonas* do better when carboxylic acids and sugars are present in high abundance. These experimental findings also provide evidence of the usefulness of 16S rRNA amplicon analysis when combined with community metabolic models to mechanistically explain correlations observed by statistical methods, and offer concrete evidence of how ecological relationships are established within the community.

4.2 Background

In the prior chapter (Chapter 3: "**CO-OCCURRENCE NETWORKS TO CHARACTERIZE HIGH RELATIVE HUMIDITY BUILT ENVIRONMENT MICROBIOMES**"), I collected numerous insights into how microbes colonize and support turnover of metabolites on high relative humidity (RH) wood surfaces over time. Figure 30 shows microbial abundance changes over time and distinctly different abundance for pre-wetted wood coupons in each habitat. Since bacterial growth was often virtually non-existent for the non-wetted environments, in this chapter I focused only on understanding the dynamics of microbial communities for the pre-wetted environments. Based on my definition of metabolic networks (Chapter 1), it is reasonable to apply metabolic models as the next tool to help explain the microbial activities and growth patterns in different wood surfaces and to describe the overall dynamics of our BE microbiome.

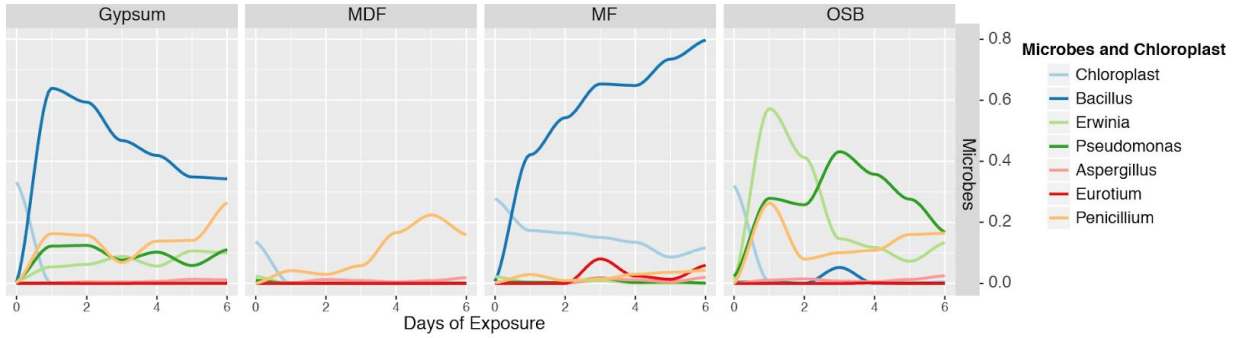

Figure 30: Microbial relative abundance in four wetted materials over time.

Metabolic models mathematically represent the metabolic potential of the cell; they genomically infer a stoichiometric matrix of biochemical reactions and metabolites that could be active in the cell (Baart and Martens, 2012). Extending this methodology to many microbes, metabolite exchanges or competitions can be predicted offering a mechanistic explanation of the interactions between species. They depend, though, on having a nearly complete, annotated genome of the species of interest. Therefore, the construction of most metabolic models relies on automated annotations, such as those provided by RAST (Aziz *et al.*, 2008), and sequence-homology based computational inference of the set of enzymatic genes present in a given genome (Eddy, 2009). Once a genome has been annotated, a metabolic reconstruction can be made linking annotated genes to metabolic-related enzymatic functions and the derived set of potential biochemical reactions (and the stoichiometric matrix, Figure 31). This reconstruction and growth simulations are performed mathematically via a linear optimization method called Flux Balance Analysis or FBA (Orth *et al.*, 2010).

Metabolic models are computational tools widely employed in basic sciences and the biochemical industry. In the basic sciences, metabolic models are a great tool to understand the physiology and metabolism of species of interest and to produce or guide hypothesis generation

(Contador *et al.*, 2019). For the biochemical industry, metabolic models are powerful tools to advance microbe engineering, helping the high throughput screening process and suggesting specific genetic manipulations that can lead to increased biochemical production (Yim *et al.*, 2011). Metabolic models have more recently started being applied in the context of microbial ecology studies (Marshall *et al.*, 2017a), which historically have relied more on statistics and systems biology approaches applied to amplicon and metagenomics datasets. Metagenomic sequencing are very amenable to metabolic modeling, since near full genomes could be reconstructed from metagenomic sequences providing the expected input data to metabolic models. Currently to our knowledge, metabolic models have not been attempted using amplicon data as a base to enable predictions.

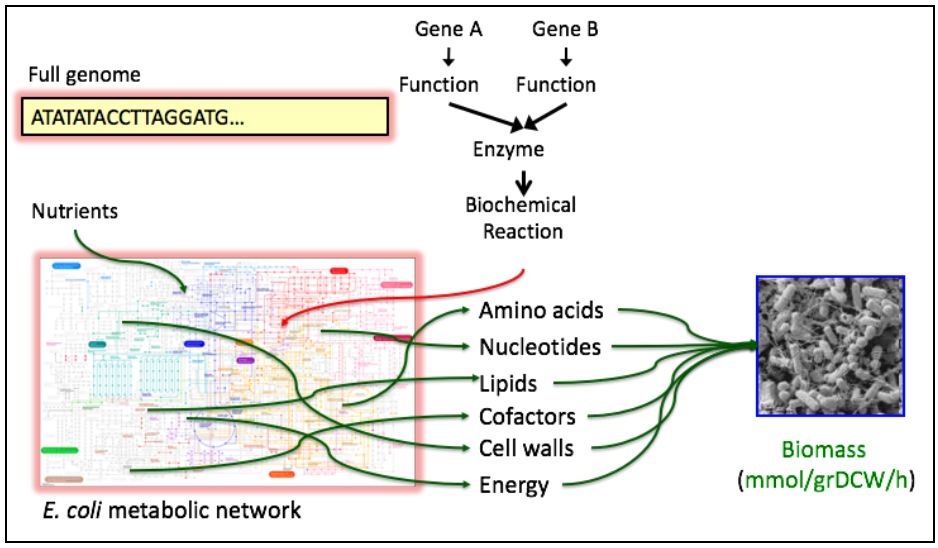

Figure 31: Traditional metabolic modeling schematic from a full-genome.

In this study, I present a new method to infer full pangenomes from 16S rRNA sequences (see Materials and methods). The 16S rRNA sequence is matched to a reference database to identify

closely associated genomes aggregated into a matching pangenome. This pangenome can now be used as an input to create an OTU-inferred, genome-enabled metabolic model and simulate the differential OTU growth rates under different nutrient conditions. Subsequently individual models are combined into a community model where the membership (abundance) of each individual member is inferred and compared with the actual data observed experimentally. I found that community membership generated by this method (starting from 16S rRNA sequences) closely resembles the patterns observed in natural ecosystems.

While I acknowledge that community metabolism is only one of many variables that determine the observed ecological dynamics, I believe that our approach describes the communities' ability to adapt to changing resource availability and to establish inter-member ecological relationships. Therefore, this study provides a community ecology understanding beyond metabolism, capturing the dynamics of the community when it responds to disruptions in local chemical or physical characteristics, and the resulting modification to the community membership. This novel method is the first to infer community-level ecological metabolism information from 16S rRNA samples, and we hope the method can continue being used going forward to help bridge the gap between genomics and ecological sciences.

4.3 Results

4.3.1 Defining the metabolic modeling approach

My analysis was focused on the three most abundant bacterial genera (*Bacillus*, *Erwinia*, and *Pseudomonas*) from the experimental microbial growth observed over time in the built environment microbiome (Figure 30). Later on, we removed OTUs from these most abundant

genera if they were outside of the top 10 overall OTUs, effectively producing two relative abundance datasets and figures (Figure 32A all OTUs and Figure 32B only the 10 most abundant OTUs for the most abundant genera). The growth pattern for the reduced set was highly correlated with the pattern of all OTUs for the genera of interest. We chose to continue working with the smaller set of only 10 OTU members for ease of computational simulations going forward.

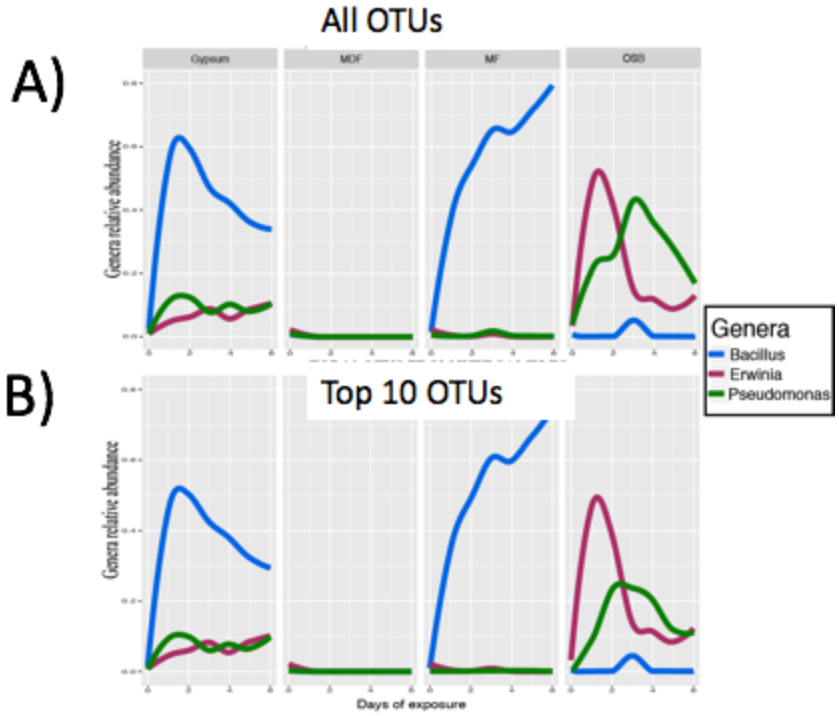

Figure 32: Bacterial abundances over time for (A) all OTUs and (B) top 10 OTUs for selected genera and all organisms from other genera.

We transformed the abundance data to display the genera (Figure 33A) and individual OTU proportion (Figure 33B) only relative to the top 10 members (such that the sum of contributions add to one) for the three materials with bacterial growth (gypsum, MF, and OSB).

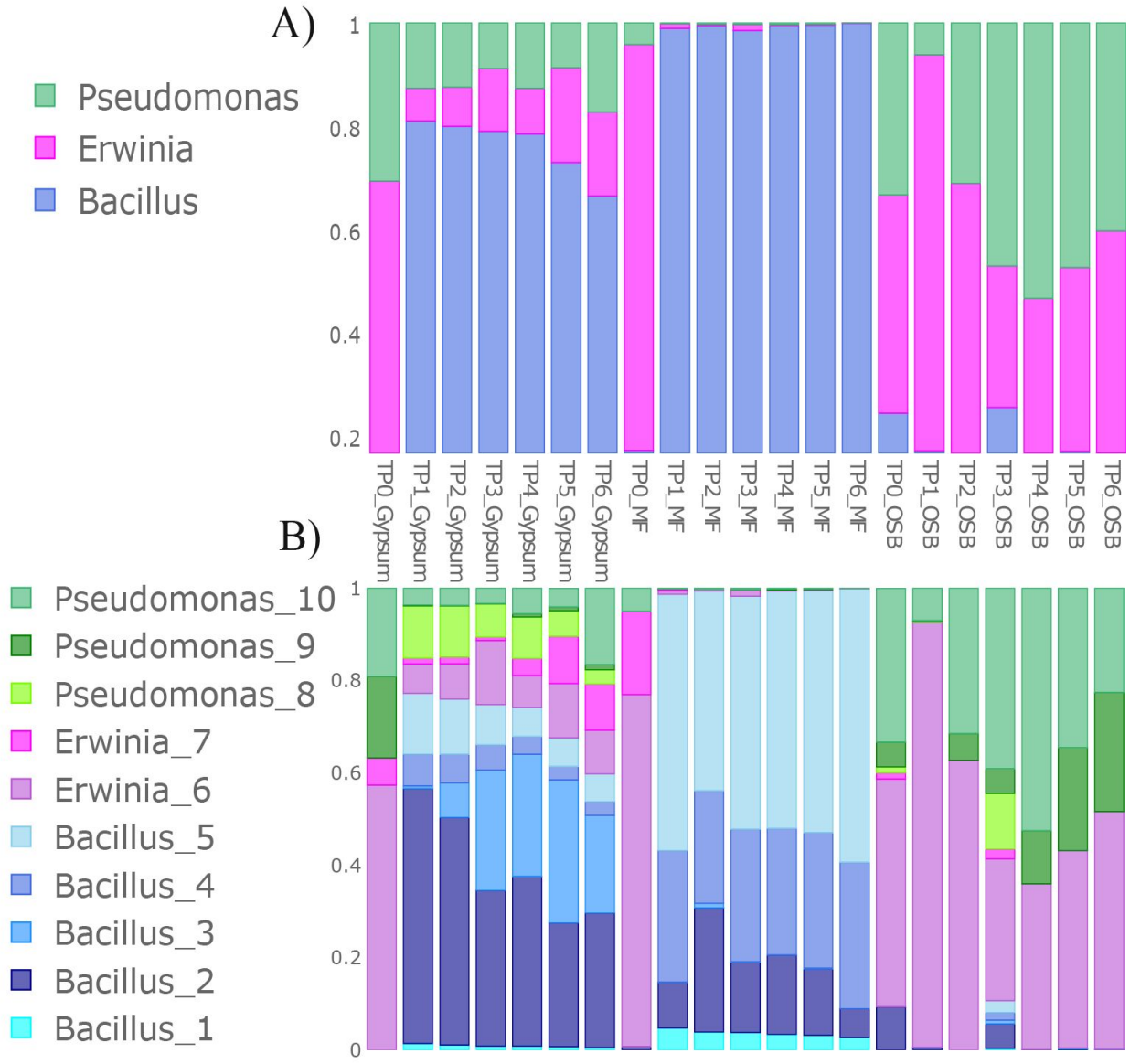

Figure 33: Relative measured bacterial abundance distributions for A) most abundant genera and B) most abundant 10 OTUs.

These abundance profiles served as the ground truth of our system, microbial profiles as they were observed in the controlled experimentation. Then the question of understanding the dynamics of the system was rewritten to something more measurable and computationally tractable: Can computational models reproduce a similar microbial abundance pattern by

simulating microbial community growth feeding from the metabolites that are present in the environment, namely wood coupons compounds and microbial byproducts? Or alternatively: Can computational models test a large number of possible nutrient sources for the community and identify those that create the pattern observed experimentally? While the first approach sounds more intuitive, there is a large bottleneck in the process of identifying the various metabolites present in the environment at each moment. The second approach requires more computational exploration, but resulting metabolites from simulations could be more easily identified in the metabolomics data. I opted for this second approach and simulated the growth of the microbial community with 384 compounds common in microbiology identification assays (i.e. Biolog plates).

4.3.2 Probabilistic metabolic models for top 10 OTUs

A new method was developed to infer the underlying metabolic network of a bacterium from its 16S rRNA sequence by identifying the set of genomes in our reference database (PATRIC, Wattam *et al.*, 2017) that was most closely phylogenetically related to the 16S rRNA sequence. All selected genomes are grouped into a representative pangenome and the frequency of genes in the pangenome determines how likely it is to find such a gene (and related functionality) in the inferred genome (Materials and methods).

The method was used to infer metabolic networks for the 10 most abundant OTUs in our study and included OTUs from *Bacillus* (B), *Erwinia* (E) *and Pseudomonas* (P) genera. Subsequently, using these network reconstructions metabolic models (B1, B2, B3, B4, B5, E6, E7, P8, P9, and

P10) were created making it possible to simulate the individual OTUs growth on different nutrient conditions and paving the way to create community metabolic models.

In addition, in order to assess the quality of the OTU models created, we also built full genome (master) models from *Bacillus*, *Erwinia*, and *Pseudomonas* isolates cultivated in our lab. The full genomes for these isolates were sequenced, and the resulting 16S rRNA gene sequence was compared against the amplicon gene sequence to confirm 100% sequence match to the OTU from the experimental dataset. Growth curves were generated for each isolate, using 384 different media conditions via Carbon, Nitrogen, Sulphur, and Phosphate Biolog plates. Growth for at least one isolate was observed for 157 of the tested media conditions. These growth experiments allowed the computational identification of biochemical reactions (biolog reactions), beyond the master model reactions, that are likely to be present in the isolates metabolism and provide them with the ability to grow in a specific subset of nutrient conditions.

Table 5 summarizes the findings of comparing the OTU models to the full genome models. The number of reactions inferred by the OTU models are on par with the master models, sometimes slightly larger, sometimes slightly smaller, but overall capturing between 75% to 93% of the same reactions when compared to the master models. A visual Jaccard similarity heatmap for the OTU models and master models is shown in Figure 34. Within each genus, OTU models are more similar to each other than to master models and they preserve phylogenetic similarities observed after also reconstructing the closest phylogenetic tree (Figure 35). The tree was inferred by identifying the closest PATRIC full genome corresponding to each OTU 16S rRNA sequence (using up to three sequences in the case of a score tie) and using the KBase tool 'Insert Set of Genomes Into Species Tree 2.2.10' (Price *et al.* 2010). We observed, for instance, how B4 and

B5 models are essentially identical, and how they are more similar to B1 model. Also B2 and B3 models are more similar to each other and *Erwinia* models are, in general, more similar to *Pseudomonas* than *Bacillus* models .

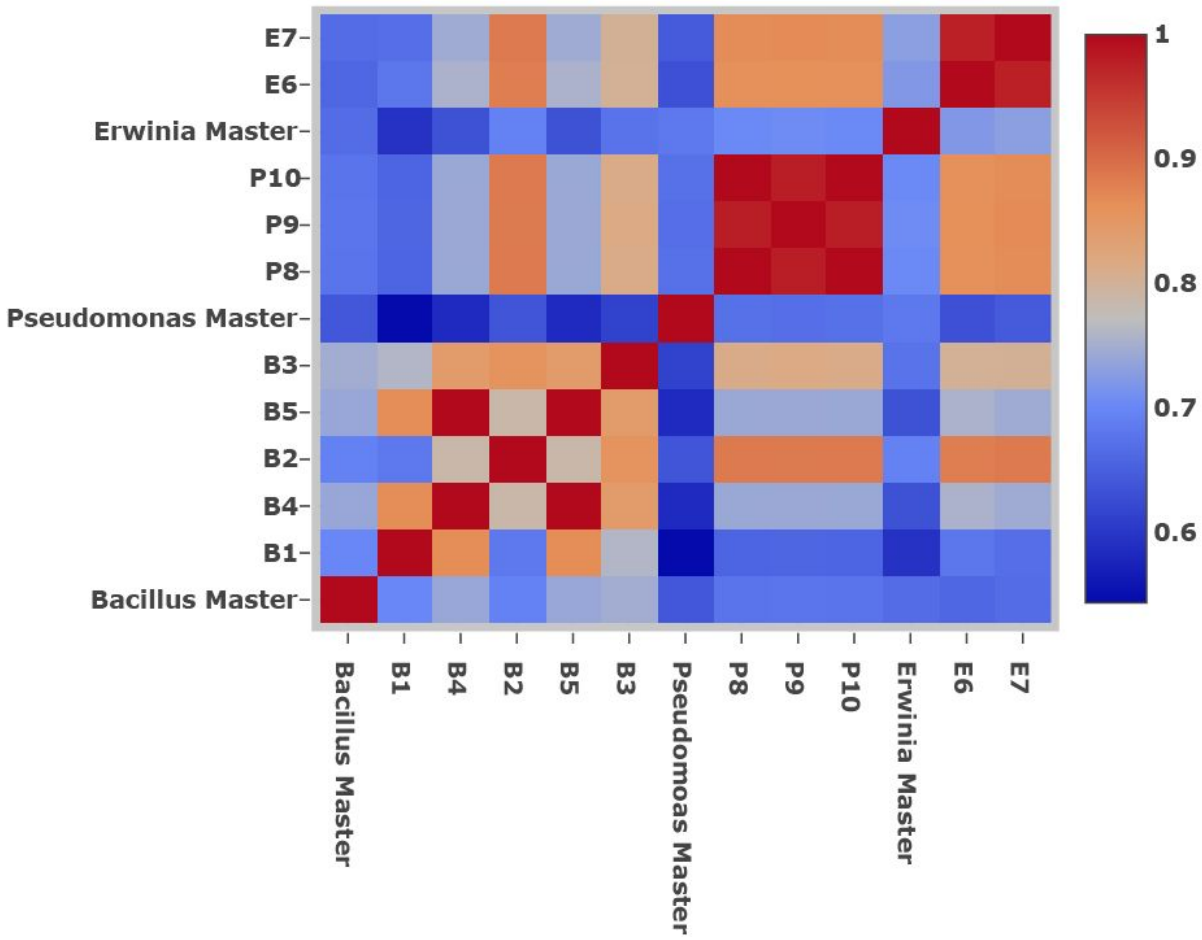

Figure 34: Jaccard similarity for OTU models and master models. Model similarity ranges from 0.54 (dark blue) to 1 (dark red).

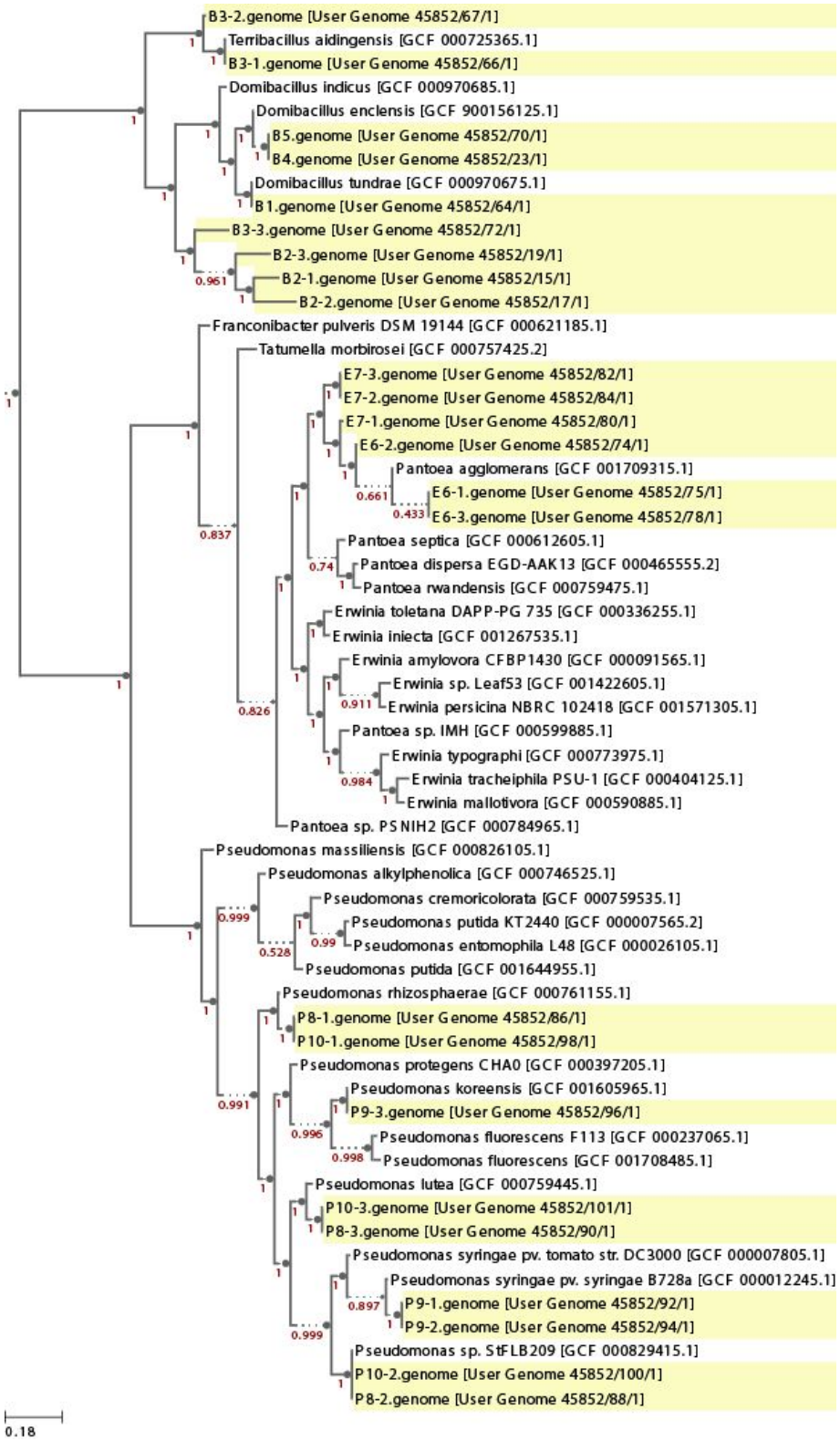

Figure 35: Phylogenetic tree reconstruction for genomes associated to OTU models (highlighted yellow) in the context of other genomes in the KBase database.

The number of biolog reactions added to each master model ranges from 36 to 142 new reactions, and out of that set, the OTU models captured between 6-32% of them. Therefore, the OTUs models capture the core reactions that full master models have, and some more specialized reactions that are only identified by experimental exercises. This is likely a result of the access to the pangenome of metabolic capability, rather than limiting selection to that which the specific clonal isolate maintains.

Differences in the number of reactions were observed between the different genera, for example, *Erwinia* and *Pseudomonas* models were quite similar in complement of reactions, but both had a greater number than the *Bacillus* models. There are several variables that determine the number of reactions in an OTU model, including the size of the selected pangenome and the diversity of the annotations available for each group. *Erwinia* and *Pseudomonas* had more similar genomes in their pangenome and a reduced number of reactions compared to *Bacillus* in the PATRIC database.

Model Type	Model ID	OTU ID	Master model number of reactions	OTU model number of reactions	OTU/master overlap rate	Num additional biolog reactions	Biolog reactions captured by OTU models	Percent biolog reactions captured by OTU models
Bacillus	B1	16S_41451	1385	1148	75%	36	2	6%
Bacillus	B2	16S_1078248	1385	1682	91%	36	10	28%
Bacillus	B3	16S_567725	1385	1447	88%	36	3	8%
Bacillus	B4	16S_179393	1385	1325	83%	36	3	8%
Bacillus	B5	16S_344495	1385	1325	83%	36	3	8%
Erwinia	E6	16S_922761	1328	1563	91%	53	17	32%
Erwinia	E7	16S_687940	1328	1589	93%	53	17	32%
Pseudomonas	P8	16S_738230	1411	1655	87%	142	28	20%
Pseudomonas	P9	16S_974121	1411	1665	88%	142	28	20%
Pseudomonas	P10	16S_589597	1411	1655	87%	142	28	20%

Table 5: Summary of OTU, and master models for top 10 OTUs models (*Bacillus*, *Erwinia*, and *Pseudomonas* genera).

4.3.3 Microbial community metabolic model

Using microbial community metabolic models to describe natural or lab controlled microbiome dynamics is a novel and promising field (Marshall *et al.*, 2017b). Community models have been implemented with different methodologies, most notably: mix bag, multi-species dynamic, and compartmentalized models (Faria *et al.*, 2016; Henry *et al.*, 2016). Regardless of the implementation, they allow researchers to simulate the community metabolic exchange and to

identify likely species interactions. The optimization problem of community models compared to those examining single species is more complex, not only for the scale of data (more compounds, reactions and fluxes), but also because of the need to balance the community optimization goal (usually biomass growth) given the limited resources and accounting for reasonable interspecies interactions (Henry *et al.*, 2016).

For this study we chose to implement a compartmentalized community model, where each one of our OTU models occupied a compartment, with the possibility of metabolite exchanges between the 10 OTU compartments (see Figure 36).

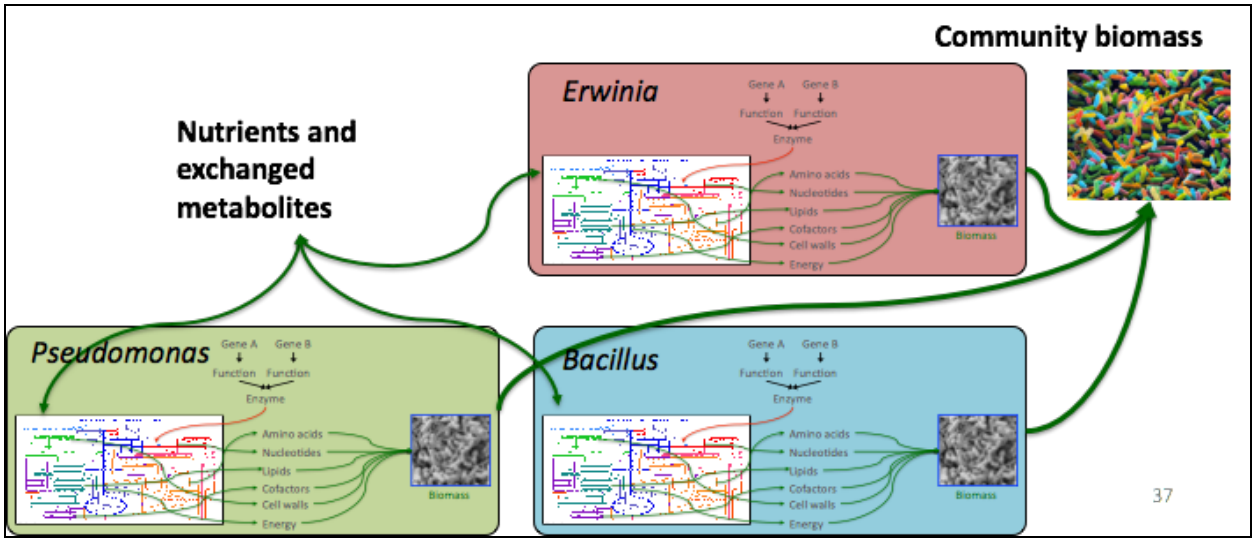

Figure 36: Schematic of a 3 member community model, where the community has a common community biomass to maximize while the member species interact for nutrients acquisition and metabolite exchanges.

Creating the community models requires the definition of a new parameter that balances the optimization of individual and community biomass growth. If focusing only on the community growth, we run the risk that solutions are not biologically reasonable. To avoid that, we defined a parameter K that refers to the proportion of microbial activity supported per unit of biomass. We assumed that the total metabolic activity of an organism within a microbiome is proportional

to the amount of biomass of that organism in the community. A higher K means more metabolic activity. Metabolic activity defines the amount of byproducts and organism can produce beyond its own internal production of biomass and it also offers the possibility of metabolic tradeoffs. We experimented with K=2000 to 3900.

If NIF_j is the number of internal fluxes for the j^{th} OTU (community member) then, the relationship of biomass, fluxes and K is given by the formula:

$$\sum_{i}^{NIF_j} (flux_{ji}) \leq K \times biomass_j$$

Also in our particular case, since our OTU models were inferred from their closest pangenomes, each reaction carries a probability score that tells the frequency of the reaction occurring in the matching pangenome. This probability of the reaction could also be used to prioritize (filter out) reactions with lower support and possibly define different subsets of chemical reactions associated to genes with high, medium, and low frequency. I defined a frequency threshold parameter F to explore the different subsets of gene frequencies for a given probabilistic OTU model.

High gene frequency genes are those core to the organisms that have been preserved at the population level. Medium and low frequency genes are those that provide flexibility and specialization to each organism, for they enable ecological dynamics and diversity to respond to new situations or novel environments (Cordero and Polz, 2014). We used F values of 0.9 for only the high frequency genes, 0.7 and 0.5 for and high and medium frequency genes and 0.0 for all high, medium, and low frequency genes.

The 157 nutrient conditions where at least one of the isolates showed detectable growth from the Biolog plate experiments were used as input media for the community metabolic model growth simulations. Exploring several combinations of K, F and input nutrients, we expected some simulation results to resemble the microbial composition observed experimentally.

4.3.4 Simulation of community microbial growth with various media and parameters

We explored the effect of the F and K parameters in the community growth, simulating growth for 157 selected nutrients. For each F and K parameter pair, we selected the 21 nutrients that generated the closest microbial community profile to experimentally measured data (10 dimensional Euclidean distance) and only kept that subset for comparisons. We used PCA plots and mantel tests to quantify the similarity of simulated versus observed community profiles. PCA plots visually presented various ordination patterns depending on the input parameters, however they did not undoubtedly show that a particular set of parameters was best at reconstructing the microbial communities profiles measured experimentally (Figure 40, Figure 37, Figure 38, Figure 39). The two PCA axis only explained between 28% and 42% of the different microbial abundance variances. It was at least evident that small K values tended to explore the search space a little more than larger ones, therefore presenting seemingly good candidates for further simulations.

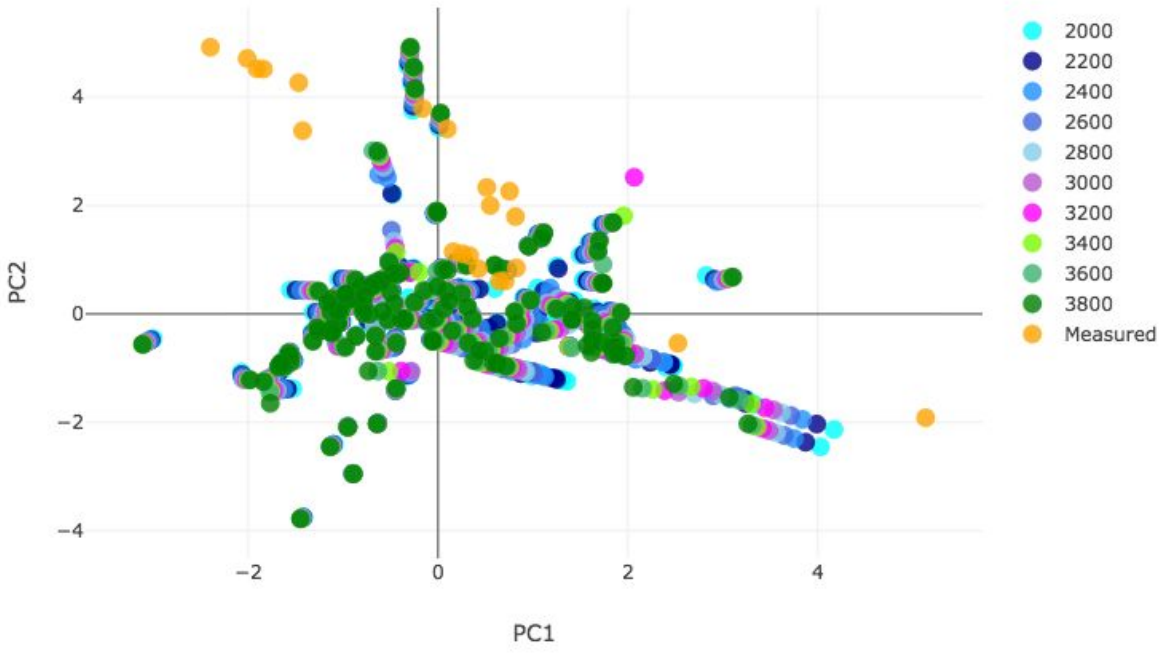

Figure 37: PCA figures maps the different microbial abundance of simulated and observed growth to two dimensions when run with F=0.9 and K between 2000 to 3800 values. Two axes account for 30% of the variation of the data.

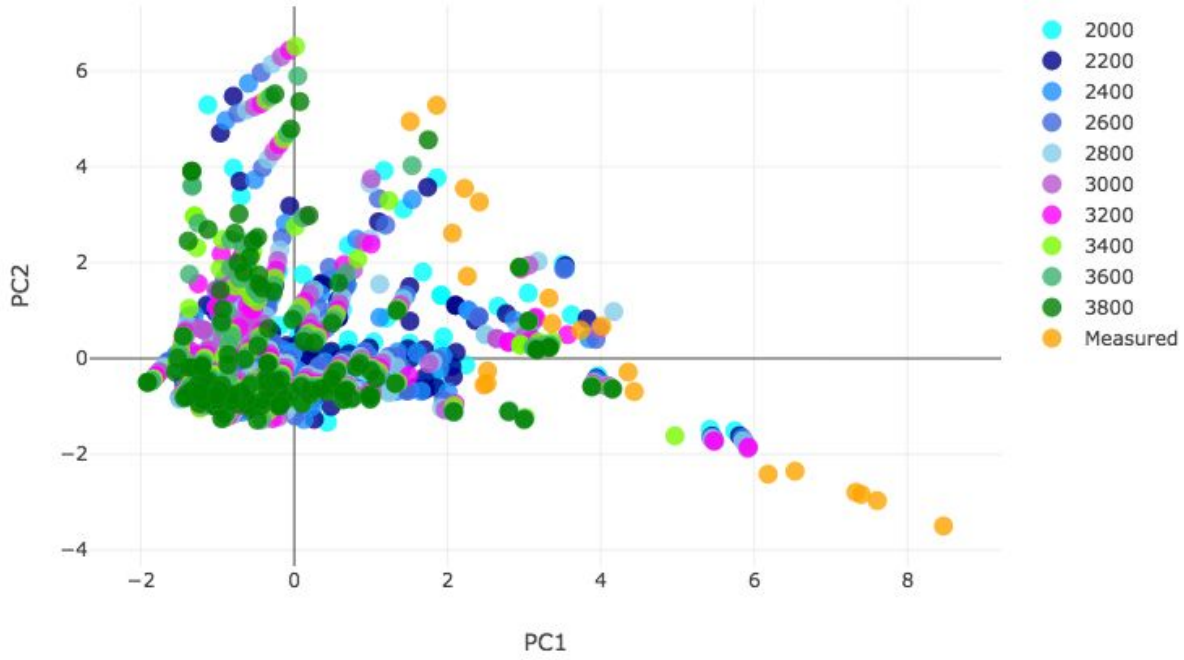

Figure 38: PCA figures maps the different microbial abundance of simulated and observed growth to two dimensions when run with F=0.7 and K between 2000 to 3800 values. Two axes account for 34% of the variation of the data.

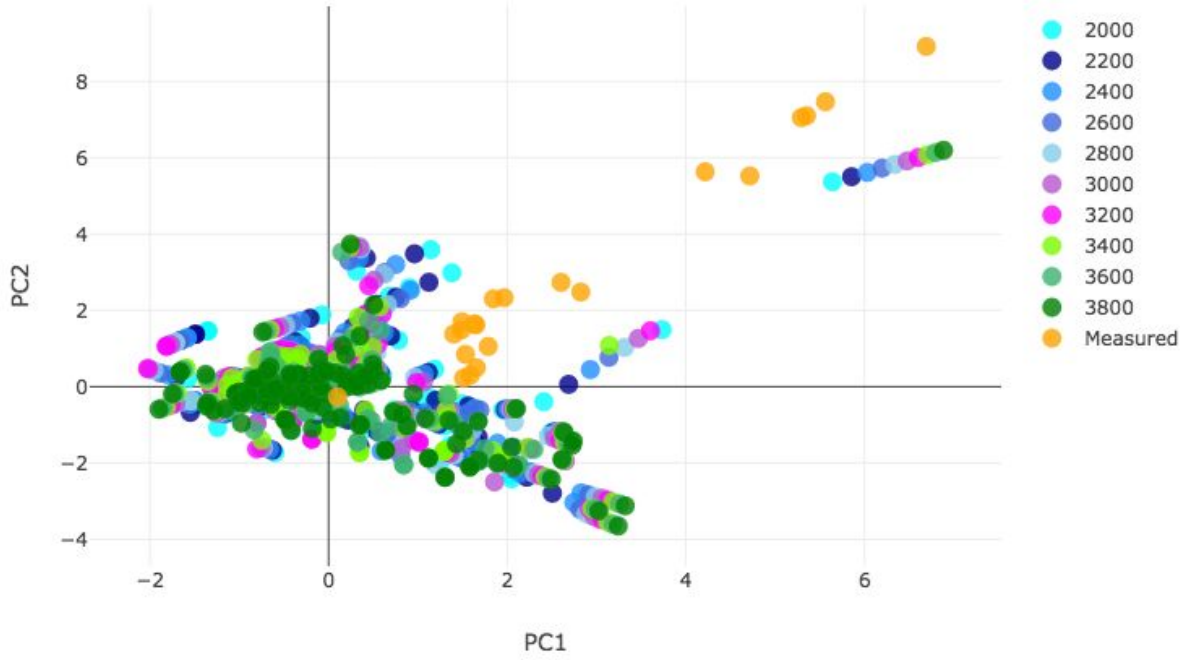

Figure 39: PCA figures maps the different microbial abundance of simulated and observed growth to two dimensions when run with F=0.5 and K between 2000 to 3800 values. Two axes account for 28% of the variation of the data.

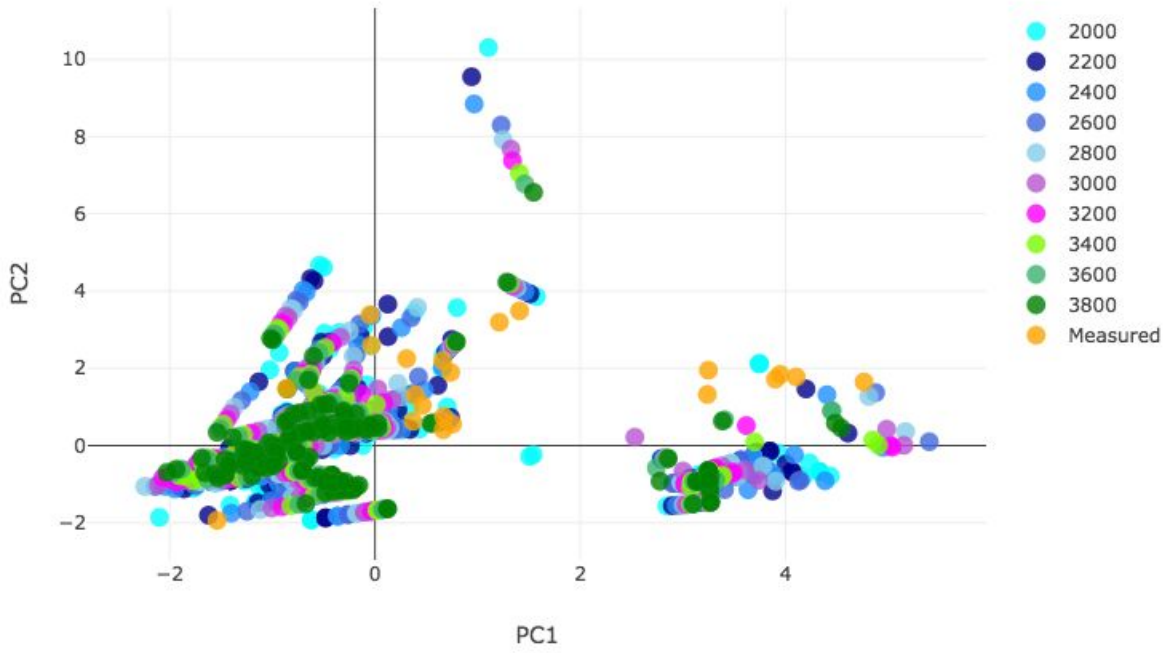

Figure 40: PCA figures maps the different microbial abundance of simulated and observed growth to two dimensions when run with F=0.0 and K between 2000 to 3800 values. Two axes account for 42% of the variation of the data.

After experimenting with constant F and variable K parameter values, we observed that simulations with smaller K values tended to produce higher mantel coefficients values, suggesting that the similarity between the inferred community abundance was slightly enhanced with smaller K values. The overall highest mantel coefficient (0.66) was produced by the simulation with parameters F=0.5 and K=2000, and while the microbial abundance of all simulations was fairly similar to each other and tended to select similar compounds or similar chemistry, we chose this pair of parameters (F=0.5 and K=2000) as the most representative for further analyses.

Parameters		Mantel coefficient	P-value	n
F=0.0	K=2000	0.55	0.001	21
F=0.0	K=3000	0.36	0.001	21
F=0.0	K=3800	0.42	0.001	21
F=0.5	K=2000	0.66	0.001	21
F=0.5	K=3000	0.60	0.001	21
F=0.5	K=3800	0.63	0.001	21
F=0.7	K=2000	0.48	0.001	21
F=0.7	K=3000	0.47	0.001	21
F=0.7	K=3800	0.44	0.001	21
F=0.9	K=2000	0.54	0.001	21
F=0.9	K=3000	0.55	0.001	21
F=0.9	K=3800	0.54	0.001	21

Table 6: Mantel test displays the correlation between the abundance profile of simulated community growth with different K and F parameters and the observed profiles. F=0.5 and K=2000 present the highest Mantel coefficient=0.66. Mantel test was run on the distance matrix for the inter-sample (material/time point) comparisons at each abundance profile.

4.3.5 Selected representative profile and its chemistry

Figure 41B represents the abundance profile at OTU level for the simulation that uses the representative parameters F=0.5 and K=2000, as it resembles the observed abundance pattern from the experimental data (Figure 41D). The main difference seems to be driven by *Bacillus* simulated growth which never reached a level below one-third of the overall community growth,

while a few times this is observed in the actual experimental profiles. Increased growth *Bacillus* in the simulations is at the expense of *Pseudomonas* growth.

We also ran an additional simulation to visualize the best matching at the genera level (3 dimensional Euclidean distance) between computer generated profiles (Figure 41A) and observed abundance profiles (Figure 41C). This was performed maintaining the same representative parameters from the OTU level match. Just like before *Bacillus* simulated growth never reached a level below one third of the overall community growth, but this time it was done at the expense of *Erwinia* growth.

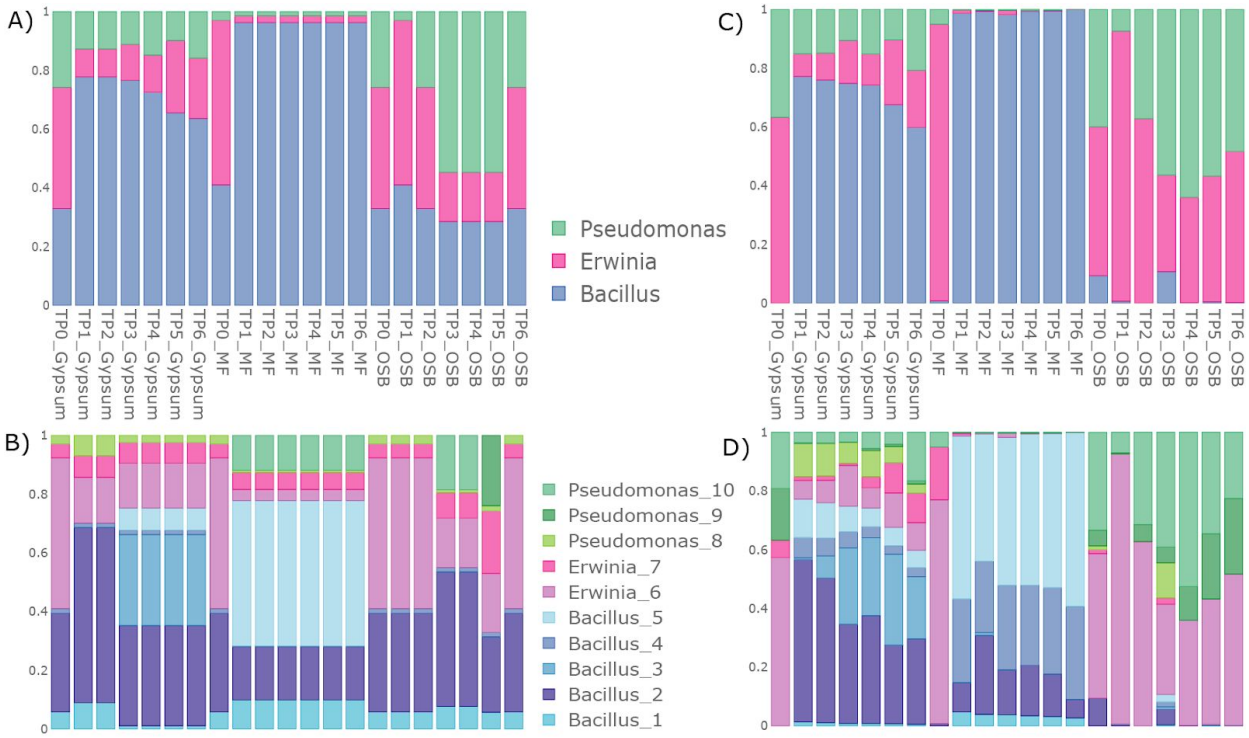

Figure 41: Microbial abundance profiles for representative simulation results, with F=0.5 and K=2000 at genera A) and OTU level B). Observed abundance profiles are placed side to side for comparison at genera C) and OTU level D).

Subsequently, I looked at the common nutrient types that resulted from the simulations and the original OTU level abundance simulation (Figure 41B) with the representative values F=0.5 and K=2000. I observed two different regimes, one dominated by *Bacillus* and a second one dominated by *Erwinia* or *Pseudomonas*. In the first scenario, the most common chemistry was usually amino acids or small molecules, while for the second one, that chemistry was primarily sugars, larger molecules, fatty acids, carboxylic acids and also sulfur rich compounds. While a more in depth analysis of these preferences remains to be done, the table below summarizes the nutrient types selected by our representative simulation (Table 7).

Time point and Material	Main nutrient	Main chemistry	Dominant genera observed
TP0_Gypsum	Glycerol	Carbohydrate, Sugar, Sugar alcohol	*Erwinia*
TP1_Gypsum	g-Amino-Butyric-Acid	Carboxylic acid, Amino acid	*Bacillus*
TP2_Gypsum	g-Amino-Butyric-Acid	Carboxylic acid, Amino acid	*Bacillus*
TP3_Gypsum	Acetamide	The simplest amine	*Bacillus*
TP4_Gypsum	Acetamide	The simplest amine	*Bacillus*
TP5_Gypsum	Acetamide	The simplest amine	*Bacillus*
TP6_Gypsum	Acetamide	The simplest amine	*Bacillus*
TP0_MF	Glycerol	Carbohydrate, Sugar, Sugar alcohol	*Erwinia*
TP1_MF	L-Leucine	Amino acid	*Bacillus*

Table 7: Microbes (genera), nutrients and general chemistry most dominant for each material at each sampled point with the closest abundance profile to the observed data.

Time point and Material	Main nutrient	Main chemistry	Dominant genera observed
TP2_MF	L-Leucine	Amino acid	*Bacillus*
TP3_MF	L-Leucine	Amino acid	*Bacillus*
TP4_MF	L-Leucine	Amino acid	*Bacillus*
TP5_MF	L-Leucine	Amino acid	*Bacillus*
TP6_MF	L-Leucine	Amino acid	*Bacillus*
TP0_OSB	Glycerol	Carbohydrate, Sugar, Sugar alcohol	*Erwinia, Pseudomonas*
TP1_OSB	Glycerol	Carbohydrate, Sugar, Sugar alcohol	*Erwinia*
TP2_OSB	Glycerol	Carbohydrate, Sugar, Sugar alcohol	*Erwinia, Pseudomonas*
TP3_OSB	Caproic-Acid	Fatty acid	*Erwinia, Pseudomonas*
TP4_OSB	Caproic-Acid	Fatty acid	*Erwinia, Pseudomonas*
TP5_OSB	L-Methionine	Sulfur amino acid	*Erwinia, Pseudomonas*
TP6_OSB	Glycerol	Carbohydrate, Sugar, Sugar alcohol	*Erwinia, Pseudomonas*

Table 7 (continued).

4.3.6 Validation of simulation results and ecological implications

I sought to validate our findings by checking if similar patterns were seen at the overall metabolomics data collected along the microbial abundance data experimentally. We merged the

5,990 compounds (peaks) found in the metabolomics data and the 157 known nutrient compounds used for simulations, (representing the only subset for which we knew their chemical group annotations), and identified 144 unique compounds matching both sets (144 were not necessarily mapped to 157 different nutrients). From this merged dataset we were able to identify the dominant chemistry on the different sampling points which is summarized in Figure 42. It is clearly not trivial to interpret the metabolomic profiles, as the chemistry reported at each time point is a combination of the dominant nutrients that are making specific microbes thrive as well as community byproducts that are being accumulated while depleting the favorite food source at that time point. This duality is what allows the dynamical environments to exist; some find a more stable equilibrium and some others do not.

To attempt an interpretation of the metabolite evolution, I focused on the points at which the microbial community transitions from a dominant species to another (Figure 42). It can be seen that initially at TP0 for the three materials there is a similar chemical profile: about 50% amino acids and another 50% carboxylic acids (including fatty acids) and sugars. *Erwinia* and *Pseudomonas* seem to thrive in this media mix, being the most abundant in all TP0 data points. Then the microbial growth pattern starts to significantly diverge between different materials. *Bacillus* takes over for gypsum and MF while *Erwinia* and *Pseudomonas* are able to maintain their dominance in the OSB wood.

At the chemical level for gypsum woods, we see that while there is a fluctuation in the chemical groups ratio, the overall pattern remains minimally changed. The abundance of the three taxa on the six data points after TP0 seems to be fairly conserved, of which the most abundant chemical groups are amino acids and monosaccharides, which seem to favor the *Bacillus* dominance while

allowing for somewhat stable small percent (20 to 30%) of *Erwinia* and *Pseudomonas*. For MF the dominance of *Bacillus* is unquestioned. They make close to 100% of the community from TP1 to TP6, thereby reducing diversity is expected to show reduction of some nutrients and accumulation of some byproducts, and that is what we see in the metabolomics chemistry over time, from TP0 to TP3 the availability of amino acids is slightly going down, while monosaccharides are consistently going up. Interestingly, there is an abrupt change in TP4 with increased amino acid abundance, and the pattern restarts again with amino acids decreasing and sugars increasing in TP5 and TP6. When looking back at the antimicrobial metabolites time series (Figure 26), there is an event in MF materials exactly at TP4 that could have caused a disruption of the microbial and chemical pattern. MF at TP4 is the only data point from the MF time series where nigragillin and fumigaclavine C compounds are detected and they are well known antimicrobial agents that could be disrupting the community. While more evidence needs to be collected to confirm the causation of this is correlation, a possible explanation to all these observations is an attack to the bacterial cells causing cell lysis and increase of amino acids back to the environment. Even though profound changes to the environmental chemistry happened at TP4, the MF relative microbial abundance didn't reflect any significant shift. This could be justified by the effect of relative data measurements or that the microbial community indeed exhibits a resistance quality by remaining minimally changed in the presence of a disturbance. Similarly, it can also be stated that since the chemical composition of the community bounces back to their original state, this could be seen as the community resilience after a disturbance. Regardless, these patterns of growth also support the observation that amino acids are one of the preferred nutrient sources for *Bacillus*.

For the OSB material the dominance is clearly for *Erwinia* and *Pseudomonas* and the chemical profile shows the rapid use of sugars, and carboxylic acids. Figure 43 highlights these chemicals and visually confirms the depletion of these compound types only for the OSB materials.

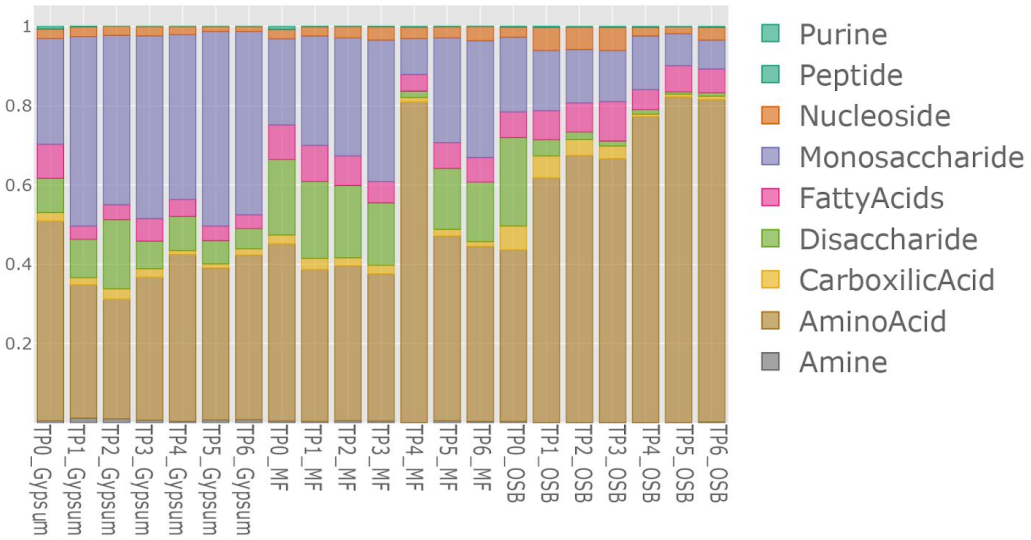

Figure 42: Distribution profiles of dominant chemistry for 144 metabolites (unique peaks) from our metabolomics experimental data in pre-wetted wood samples for built environment microbiome. Amino acids and monosaccharides are the highest abundance chemical groups.

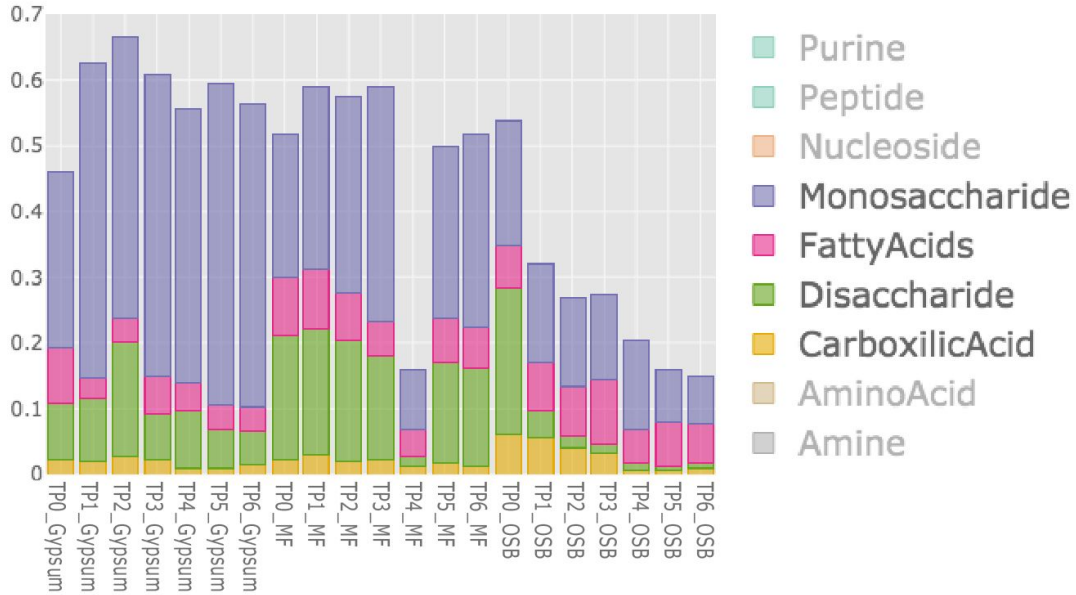

Figure 43: Distribution profiles of sugars and carboxylic acids for 144 metabolites (unique peaks) from metabolomics experimental data in pre-wetted wood samples for built environment microbiome.

In summary, the observed microbiome-metabolomics patterns in the experimental data do correlate well with the simulation results and while more rigorous analysis ought to be done, I feel that there is enough evidence here to continue advancing this novel approach.

Another way to help understand the dynamics of the system is to look at the results of a particular sample point simulation. While this doesn't directly reproduce the dynamics patterns observed four days apart (as the measured data), it can explain from its instantaneous perspective what could be the biochemical exchanges that rule the community dynamics at that specific data point and suggest what chemistry could become available in the environment for future exchanges. Here we show the metabolic exchange within the community when 20 nmol mmol/grDCW/h of glycerol are used as the initial and main Carbon nutrient source to feed the community. Glycerol is of special interest for our system as it was found (in the simulations) to be the nutrient that provides the most similar microbial growth profiles between experimentally measured and computational simulations for all materials at time TP0. The result of the metabolic exchanges (compounds taken or excreted by each OTU model) are presented in Table 8 and allows us to observe several interesting patterns.

Compound	B1	B2	B3	B4	B5	E6	E7	P8	P9	P10
L-Isoleucine	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	0.87	(0.07)	(0.09)	(0.09)	(0.09)
CO2	2.76	59.64	0.13	1.60	0.16	-	(92.11)	10.93	9.84	7.04
Glycerol	-	(56.46)	(0.14)	1.16	-	35.43	0.00	-	0.00	
L-Leucine	(0.12)	(0.10)	(0.12)	1.04	(0.10)	(0.13)	(0.13)	(0.11)	(0.11)	(0.11)
L-Lactate	-	(51.94)	2.63	-	7.62	-	17.25	13.87	10.58	
L-Alanine	-	(114.40)	3.23	4.74	-	7.02	99.41	-	-	-
L-Glutamate	(41.20)	(44.33)	(0.27)	(0.26)	(0.08)	8.21	68.80	-	2.84	6.30
Succinate	(19.86)	122.04	0.06	(0.55)	(0.14)	(0.16)	(50.65)	(20.03)	(18.05)	(12.66)
D-Mannitol	-	509.09	-	-	-	(25.90)	(483.19)	-	-	-
L-Proline	23.56	43.89	0.00	-	(0.18)	(5.48)	(68.02)	3.65	1.61	0.96
Cytidine	-	383.13	-	-	-	(15.63)	(367.50)	-	-	-
D-Fructose	(1.69)	-	(2.81)	(5.35)	(4.05)	14.10	-	-	(0.20)	-
Na+	11.78	(0.76)	(0.67)	(0.74)	(0.74)	7.94E-06	0.00	-	1.42	3.15
Phosphate	(0.54)	(0.36)	(0.36)	(0.36)	(0.36)	(0.92)	(0.26)	-	-	-
H2O	25.83	28.30	14.35	17.63	13.04	-	-	-	-	0.86

Table 8: Table shows the metabolic exchanges for a microbial community simulation with 20 nmol mmol/grDCW/h of glycerol as the main nutrient source. Red cells show represent the

Table 8 (continued): intake of the compound at the left by the OTU at the top, blue cells represent byproducts following the same left, top pattern.

First all glycerol is taken (red cells) by a single bacterium model, the OTU B2 (*Bacillus_2)*. However, the intake of glycerol is almost three times higher than the initial supplied input. It turns out that this could be explained by observing that B4 (*Bacillus_4*) and E6 (*Erwinia_6*) are producing (blue cells) additional glycerol for B2 (*Bacillus_2*). But why would they do that and how are they (and the rest of the community members) growing if they are not taking the initial input nutrient? The answers could be inferred by the exchange table itself. B2 (*Bacillus_2*) is making a carboxylic acid (succinate), a sugar alcohol (mannitol), an amino acid (proline), and a nucleoside (cytidine) making those available for the rest of the community members. Succinate is taken by B1 (*Bacillus_1*) and all *Erwinia* and *Pseudomonas* OTU models. E6 (*Erwinia_6*) in exchange also provides a sugar (fructose) to the remaining slow growing bacilli B3 (*Bacillus_3*), B4 (*Bacillus_4*), and B5 (*Bacillus_5)*. This multi-tier dynamic allows all community members to grow and shows how collaboration patterns could arise. There is also clear chances for competition when multiple members want to intake the same compound, for instance in the case of the three bacilli B3 (*Bacillus_3*), B4 (*Bacillus_4*), and B5 (*Bacillus_5)*, as they all seek to acquire the same sugar to support their individual growth.

Also, these observations align with the microbe-chemistry preferences observed so far. The time points where glycerol was chosen as the best match to the experimental data are also the time points where *Erwinia* and *Pseudomonas* growth is more dominant, and this metabolite exchange table reflects how for these genera, the OTU models have a preferred metabolite (higher red cell

value) either a sugar alcohol or the carboxylic acid compound, in agreement with all patterns observed so far.

4.4 Discussion

This work is an extension of the earlier published study (Lax and Cardona *et al.*, 2019). Here I explored the biochemical mechanisms that underpin the ecological dynamics of the microbial communities in high relative humidity built environments, depending on the material type and the underlying chemistry of the environment. I have shown how the addition of metabolic models increase our understanding of how the microbial community is structured and the mechanisms that could lead to ecological features. Our prior work suggested that same taxa (Figure 18) usually co-occur in multiple samples together and some different taxa, most notably Bacillaceae and Pseudomonadaceae, have frequent anti-correlated occurrence patterns (Figure 19 and Figure 20). My new work presented here shows how cross feeding between different OTUs can explain such ecological relationships (Table 8).

Originally (Chapter 3) we identified a few antifungals at different time points for different materials, which we hypothesized could be shaping the community structure. In this work, by grouping some of our metabolomics datasets into common chemical types, we were able to see what changes on the microbial community and chemical composition occurred at TP4 for the MF woods. TP4 is of interest because it is the single time point where antimicrobial compounds were identified (nigragillin and fumigaclavine C) on the MF material. We observed that while the microbial composition was minimally changed, which could represent the ecological feature of resistance, the chemical composition was drastically changed at that same time point. We are not

able to show causality between the presence of the two antibacterial and the chemical profile changes; however, as the chemical profiles start to trend back to their prior state following the disappearance of antibacterials, it is possible that this represents an ecological feature, namely community resilience.

In addition to this finding, the chemical types identified for our metabolites provide an explanation of how the nutrient availability might shape microbial abundance. We saw how the chemistry of all wood surfaces were initially very similar, composed of about 50% amino acids and 50% carboxylic acids and sugars. However, depending on the material type, two very different patterns diverged: the solid wood materials (Gypsum and MF) were colonized primarily by gram positive *Bacillus* taxa, and the OSB, an agglomerate material, was mostly dominated by gram negative *Erwinia* and *Pseudomonas* OTUs. These two different growth patterns seem to be driven by the underlying chemical compounds present at each wood type, with *Bacillus OTUs* seemingly feeding mostly from amino acids and *Erwinia* and *Pseudomonas* OTUs favoring all sugars and fatty acids. This pattern of genera-chemical type preference is also often detected by our computational experiments and shown in Table 7 as one of the representative simulation examples.

Future work should focus on improving the weaker parts of this first exploratory analysis: 1) use a full size chemical groups dataset that matches most compounds on metabolomics data; 2) further study the effect of F and K parameters, in particular the connection of F values with the high, medium, and low abundance genes in the pangenomes and community models; 3) explore further differences between experimental and simulated microbial profiles, in our case, why

Bacillus growth was always so dominant, even in places where that was not expected, and 4) automate the workflow to make it more accessible to others.

I am, however, optimistic that the system I presented here to infer genomes from 16S rRNA sequences and the combination of OTU models into community models could become standard tools for microbial ecology so scientists can start explaining the correlative effects observed with statistics and co-occurrence networks. Adding the power of metabolic models opens the access to mechanistic information that can accelerate our understanding of causality and ecological relationships for built and any other microbiomes, and now this can be done starting from only amplicon sequence data.

4.5 Materials and methods

4.5.1 Biolog phenotype assays

Bacillus subtilis (ATCC 6056), *Erwinia herbicola* (ATCC 27155), and *Pseudomonas putida* (ATCC 12633) were spread on nutrient broth (NB) media petri dishes at room temperature until colonies of significant size (approximately 1 mm or larger) had appeared. Colonies were scraped off and thoroughly resuspended in inoculating fluid. Each strain was tested against Biolog plates containing a variety of compounds: carbon (PM01, PM2A), nitrogen (PM3B), as well as sulphate and phosphate (PM4B). Following the manufacturer's protocol ("PM for *B. subtilis* and other GP Bacteria" for *B. subtilis*, and "PM for *E. coli* and other GN Bacteria" for *E. herbicola* and *P. putida*), the inoculating fluid containing the diluted cell suspension was added to each well of the

Biolog plates. A redox dye present in each well changes color as the cells grow, indicating that the compound can be utilized. Each assay was ran for 72 hours. Output from Biolog's camera data was pre-processed first with the Biolog OmniLog software, then analyzed with the omp R package to determine the utilization by each microbe of the various compounds, and thus their phenotypic potential.

4.5.2 DNA extraction and genomic sequencing

Strains were grown overnight in liquid nutrient broth, then the culture centrifuged, and the cell pellet collected. DNA was extracted using the DNeasy PowerSoil kit (Qiagen), and following the manufacturer's protocol. Genomes were sequenced using paired-end libraries that were prepared using the Nextera XT kit (Illumina). Sequencing was performed on a MiSeq 2x150 nt run.

4.5.3 Probabilistic metabolic network reconstruction from 16S rRNA sequences

Metabolic models are traditionally built from nearly full genomes and that has limited the microbial ecology community ability to use them more extensively despite its proven value to study microbial strains and communities. Most microbial ecology projects rely only on amplicon sequencing because of limitations on the DNA biomass availability or budget to pay for shotgun metagenomics sequencing that can provide these full genomes. With our new multi-step approach proposed here we infer the most likely (probabilistic) bacterium genome only from its 16S rRNA sequence. The probabilistic genome provides a way to define the microorganism most

likely (probabilistic) metabolic network and the ability to perform growth simulations for both individual microorganisms and microbial communities via metabolic models.

The inference of a microorganism probabilistic metabolic network using its 16S rRNA sequence (queried sequence) requires four steps, three to create a probabilistic genome and one to expand the genome into the associated metabolic network (Figure 44). In the first step, the queried sequence is matched to the 16S rRNA PATRIC (Wattam *et al.*, 2017) database to identify all PATRIC genomes that are at least 97% similar (a different similarity threshold could be used too) to the input sequence. Secondly, the full genomes associated with these PATRIC microorganism identities are grouped into a "matched pangenome" set. Thirdly, all genes present in the matched pangenome are annotated using PATRIC databases and the occurrence frequency for each gene is calculated. Now, the superset of the pangenome genes are used to represent the probabilistic genome genes for the queried sequence with each gene probability equal to the occurrence frequency of such a gene on the pangenome set.

Once the probabilistic genome is created, enzymes and biochemical reactions likely to be present in the bacterium are inferred based on standard Boolean rules from the available genes. The probability of biochemical reactions is calculated as a sum (which should not exceed 1) of the genes (enzymes) catalyzing each particular reaction. This is the fourth and final step that completes the metabolic network reconstruction from a 16S rRNA sequence and metabolic models could be created and run going forward.

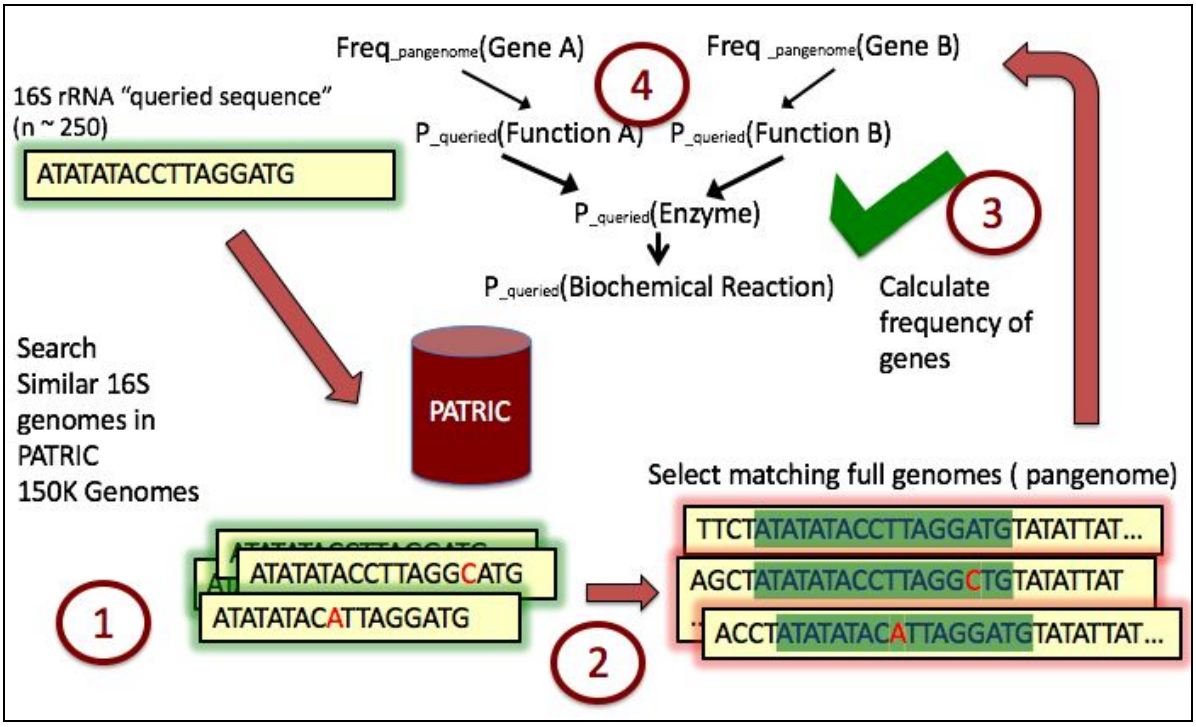

Figure 44: Schematic of the probabilistic reconstruction of the metabolic network of a bacterium from its 16S rRNA sequence. 1) The queried 16S sequence is mapped to the PATRIC 16S database 2) the matching PATRIC full genomes are grouped into a 'matched pangenome'. 3) Genes are extracted from the genomic features of the matched pangenome and an inferred genome is created as the aggregation of all reactions on the pangenome with a probability of each gene to be present equal to the frequency of that gene in the matched pangenome. 4) Gene probabilities are combined to obtain biochemical reactions probabilities completing the probabilistic inference of the metabolic network.

However, by using the probabilistic metabolic reconstruction, the biochemical reactions have a probability score that could be harnessed to filter out reactions with lower probability support and possibly define different subset of chemical reactions associated to genes with high, medium, and low frequency. High gene frequency genes are those core to the organisms that have been preserved at the population level. Medium and low frequency genes are those that provide flexibility and specialization to each organism, for they enable ecological dynamics and diversity to respond to new situations or novel environments (Cordero and Polz, 2014).

4.5.4 Genome scale metabolic models simulations

We used the Department of Energy Systems Biology Knowledgebase (KBase, http://kbase.us) to create the OTU models and community models. KBase is a software and data platform designed to predict and design biological functions for microbes and for plants and their communities. It integrates data and apps in a friendly user interface so users can combine multiple tools in sophisticated workflows (narratives). KBase is the first large-scale bioinformatics system that enables users to upload their own data, analyze it, build increasingly realistic models, and share and document their workflows and conclusions for publication (Arkin *et al.*, 2018).

Our project used four main narratives, with BE4 containing the final community models reported in this document.

- BE1: Media repository storing the 157 media used for community metabolic modeling simulations: https://narrative.kbase.us/narrative/ws.44082.obj.1
- BE2: OTU and community models repository: https://narrative.kbase.us/narrative/ws.41899.obj.774
- BE3: Gap-filling community models: https://narrative.kbase.us/narrative/ws.44087.obj.3182
- BE4: FBA community composition simulations: https://narrative.kbase.us/narrative/ws.44080.obj.4305

CHAPTER 5

CONCLUSION AND FUTURE DIRECTIONS

Currently, microbial co-occurrence networks and microbial metabolic modeling each provide unique insight into aspects of microbial community structure and assembly. Microbial co-occurrence networks are useful for identifying hub or keystone species, examining the relative influence of environmental factors and biotic interactions in determining community structure, and exploring community stability, while the strengths of metabolic modeling lie in its potential to establish a biochemical mechanism explaining the metabolic subset of species interactions, generate predictive models of organism and community behavior, and explore community dynamics through the application of approaches such as dynamic FBA. Each of these methods can be used independently to address relevant and important questions in microbial ecology. However, the mechanistic understanding of microbial community stability and dynamics necessary in order to engineer and manipulate microbial functions relevant to human endeavors requires integration of these approaches. Such integration would enable us to take advantage of the mechanistic understanding of metabolic models and pair it with the broader relationships that can be discovered through examination of microbial co-occurrence networks. One possible pipeline is to first use microbial co-occurrence networks to identify keystone or hub OTUs and the organisms with which they interact most strongly. Then use metagenomic binning and automatic annotation to generate draft metabolic reconstructions and first iteration metabolic models. These early models could be used to predict changes in those species over time or with changes in environmental conditions, using dynamic FBA and media with altered nutrient inputs.

These predictions could then be validated with additional sampling, or methods such as PRMT could be applied to evaluate the emergent effects of altered conditions on the keystone species as well as the remainder of the community. Such an approach would enable *in silico* evaluation of engineering and management strategies to direct microbial community behavior and function.

REFERENCES

Adams, Rachel I., *et al.* "Ten questions concerning the microbiomes of buildings." *Building and Environment* 109 (2016): 224-234.

Adams, Rachel I., *et al.* "Dispersal in microbes: fungi in indoor air are dominated by outdoor air and show dispersal limitation at short distances." The ISME journal 7.7 (2013): 1262.

Apprill, A. (2017). Marine Animal Microbiomes: Toward Understanding Host–Microbiome Interactions in a Changing Ocean. Front. Mar. Sci. *4*.

Arkin, A.P., Cottingham, R.W., Henry, C.S., Harris, N.L., Stevens, R.L., Maslov, S., Dehal, P., Ware, D., Perez, F., Canon, S., *et al.* (2018). KBase: The United States Department of Energy Systems Biology Knowledgebase. Nat. Biotechnol. 36, 566–569.

Attramadal, K.J.K., Minniti, G., Øie, G., Kjørsvik, E., Østensen, M.-A., Bakke, I., and Vadstein, O. (2016). Microbial maturation of intake water at different carrying capacities affects microbial control in rearing tanks for marine fish larvae. Aquaculture *457*, 68–72.

Avalos-Téllez, R., Suárez-Güemes, F, Carillo-Casas, EM, and Hernandez-Castro, R Bacteria and yeast normal microbiota from respiratory tract and genital area of bottlenose dolphins (Tursiops truncatus). (2010). Research Gate.

Baart, G.J.E., and Martens, D.E. (2012). Genome-Scale Metabolic Models: Reconstruction and Analysis. In Neisseria Meningitidis, M. Christodoulides, ed. (Totowa, NJ: Humana Press), pp. 107–126.

Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Li, Y., Xia, Y., Xie, H., Zhong, H., *et al.* (2015). Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell Host Microbe *17*, 690–703.

Balba, Hamdy. "Review of strobilurin fungicide chemicals." *Journal of Environmental Science and Health Part B* 42.4 (2007): 441-451.

Benítez-Páez, A., Belda-Ferre, P., Simón-Soro, A., and Mira, A. (2014). Microbiota diversity and gene expression dynamics in human oral biofilms. BMC Genomics *15*, 311.

Berry, D., and Widder, S. (2014). Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Microb. Symbioses *5*, 219.

Bik, E.M., Costello, E.K., Switzer, A.D., Callahan, B.J., Holmes, S.P., Wells, R.S., Carlin, K.P., Jensen, E.D., Venn-Watson, S., and Relman, D.A. (2016). Marine mammals harbor unique microbiotas shaped by and yet distinct from the sea. Nat. Commun. *7*, 10516.

Brooks, A.W., Kohl, K.D., Brucker, R.M., Opstal, E.J. van, and Bordenstein, S.R. (2016). Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History. PLOS Biol. *14*, e2000225.

Cahill, P.L., Fidler, A.E., Hopkins, G.A., and Wood, S.A. (2016). Geographically conserved microbiomes of four temperate water tunicates. Environ. Microbiol. Rep. *8*, 470–478.

Campbell, D.J., and Koch, M.A. (2017). Living in Peace: Host-Microbiota Mutualism in the Skin. Cell Host Microbe *21*, 419–420.

Caporaso, J. Gregory, *et al.* "Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample." *Proceedings of the national academy of sciences* 108.Supplement 1 (2011): 4516-4522.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods *7*, 335–336.

Caporaso, J.G., Lauber, C.L., Costello, E.K., Berg-Lyons, D., Gonzalez, A., Stombaugh, J., Knights, D., Gajer, P., Ravel, J., Fierer, N., *et al.* (2011). Moving pictures of the human microbiome. Genome Biol. *12*, R50.

Cardona C. 2017. CAVNet: Creation Analysis and Visualization of Networks Package. Bitbucket. https://bitbucket.org/cesarcardonau/cavnet

Cardona, C., Lax, S., Larsen, P., Stephens, B., Hampton-Marcell, J., Edwardson, C.F., Henry, C., Van Bonn, B., and Gilbert, J.A. (2018). Environmental Sources of Bacteria Differentially Influence Host-Associated Microbial Dynamics. MSystems 3.

Cardona, C., Weisenhorn, P., Henry, C., and Gilbert, J.A. (2016). Network-based metabolic analysis and microbial community modeling. Curr. Opin. Microbiol. 31, 124–131.

Chase, John, *et al.* "Geography and location are the primary drivers of office microbiome composition." *MSystems* 1.2 (2016): e00022-16.

Clausen, Carol A., and Vina Yang. "Protecting wood from mould, decay, and termites with multi-component biocide systems." *International Biodeterioration & Biodegradation* 59.1 (2007): 20-24.

Cole, Richard J., *et al.* "Mycotoxins produced by *Aspergillus fumigatus* species isolated from molded silage." *Journal of Agricultural and Food Chemistry* 25.4 (1977): 826-830.

Contador, C.A., Rodríguez, V., Andrews, B.A., and Asenjo, J.A. (2019). Use of genome-scale models to get new insights into the marine actinomycete genus Salinispora. BMC Syst. Biol. 13.

Coombs, Kanistha, *et al.,* "Fungal microbiomes associated with green and non-green building materials." *International biodeterioration & biodegradation* 125 (2017): 251-257.

Cordero, O.X., and Polz, M.F. (2014). Explaining microbial genomic diversity in light of evolutionary ecology. Nat. Rev. Microbiol. 12, 263–273.

Costello, E.K., Lauber, C.L., Hamady, M., Fierer, N., Gordon, J.I., and Knight, R. (2009). Bacterial Community Variation in Human Body Habitats Across Space and Time. Science *326*, 1694–1697.

Curtis, T.P., Sloan, W.T., and Scannell, J.W. (2002). Estimating prokaryotic diversity and its limits. Proc. Natl. Acad. Sci. 99, 10494–10499.

Dannemiller, Karen C., *et al.,* "Combining real-time PCR and next-generation DNA sequencing to provide quantitative comparisons of fungal aerosol populations." *Atmospheric environment* 84 (2014): 113-121.

David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A., *et al.* (2014a). Diet rapidly and reproducibly alters the human gut microbiome. Nature *505*, 559–563.

David, L.A., Materna, A.C., Friedman, J., Campos-Baptista, M.I., Blackburn, M.C., Perrotta, A., Erdman, S.E., and Alm, E.J. (2014b). Host lifestyle affects human microbiota on daily timescales. Genome Biol. *15*, R89.

Department of Agriculture. e-CFR: Title 9: Animals and Animal Products PART 3—STANDARDS Subpart E—Specifications for the Humane Handling, Care, Treatment, and Transportation of Marine Mammals.

Diaz, M., Bik, E., Carlin, K., Venn-Watson, S., Jensen, E., Jones, S., Gaston, E., Relman, D., and Versalovic, J. (2013). Identification of *Lactobacillus* strains with probiotic features from the bottlenose dolphin (Tursiops truncatus). J. Appl. Microbiol. *115*, 1037–1051.

Diomande, Sara Esther, *et al.,* "Role of fatty acids in *Bacillus* environmental adaptation." *Frontiers in microbiology* 6 (2015): 813.

Dominguez-Bello, M.G., De Jesus-Laboy, K.M., Shen, N., Cox, L.M., Amir, A., Gonzalez, A., Bokulich, N.A., Song, S.J., Hoashi, M., Rivera-Vinas, J.I., *et al.* (2016). Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. Nat. Med. *22*, 250–253.

Dubois-Brissonnet, Florence, Elsa Trotier, and Romain Briandet. "The biofilm lifestyle involves an increase in bacterial membrane saturated fatty acids." *Frontiers in microbiology* 7 (2016): 1673.

Eddy, S.R. (2009). A New Generation Of Homology Search Tools Based On Probabilistic Inference. In Genome Informatics 2009, (Pacifico Yokohama, Japan: Published By Imperial College Press And Distributed By World Scientific Publishing Co.), pp. 205–211.

Edwards, Michael R., *et al.* "The microbiology of asthma." Nature Reviews Microbiology 10.7 (2012): 459.

Eren, A.M., Maignien, L., Sul, W.J., Murphy, L.G., Grim, S.L., Morrison, H.G., Sogin, M.L., and Freckleton, R. (2013). Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. Methods Ecol. Evol. *4*, 1111–1119.

Fair, P.A., Schaefer, A.M., Houser, D.S., Bossart, G.D., Romano, T.A., Champagne, C.D., Stott, J.L., Rice, C.D., White, N., and Reif, J.S. (2017). The environment as a driver of immune and endocrine responses in dolphins (Tursiops truncatus). PLOS ONE *12*, e0176202.

Faria, J.P., Khazaei, T., Edirisinghe, J.N., Weisenhorn, P., Seaver, S.M.D., Conrad, N., Harris, N., DeJongh, M., and Henry, C.S. (2016). Constructing and Analyzing Metabolic Flux Models of Microbial Communities. In Hydrocarbon and Lipid Microbiology Protocols, T.J. McGenity, K.N. Timmis, and B. Nogales, eds. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 247–273.

Feighery, L.M., Smith, P., O'Mahony, L., Fallon, P.G., and Brayden, D.J. (2008). Effects of *Lactobacillus salivarius 433118* on Intestinal Inflammation, Immunity Status and In vitro Colon Function in Two Mouse Models of Inflammatory Bowel Disease. Dig. Dis. Sci. *53*, 2495–2506.

Fisk, William J., Ekaterina A. Eliseeva, and Mark J. Mendell. "Association of residential dampness and mold with respiratory tract infections and bronchitis: a meta-analysis." *Environmental Health* 9.1 (2010): 72.

Friedman, Jonathan, and Eric J. Alm. "Inferring correlation networks from genomic survey data." *PLOS computational biology* 8.9 (2012): e1002687.

Gibbons, Sean M. "The built environment is a microbial wasteland." *MSystems* 1.2 (2016): e00033-16.

Gibbons, Sean M., *et al.,* "Ecological succession and viability of human-associated microbiota on restroom surfaces." *Applied and environmental microbiology* 81.2 (2015): 765-773.

Gilbert, J.A., Jansson, J.K., and Knight, R. (2014). The Earth Microbiome project: successes and aspirations. BMC Biol. *12*.

Gilbert, J.A., Quinn, R.A., Debelius, J., Xu, Z.Z., Morton, J., Garg, N., Jansson, J.K., Dorrestein, P.C., and Knight, R. (2016). Microbiome-wide association studies link dynamic microbial consortia to disease. Nature *535*, 94–103.

Gnonlonfin, GJ Benoit, Ambaliou Sanni, and Leon Brimer. "Review scopoletin–a coumarin phytoalexin with medicinal properties." *Critical Reviews in Plant Sciences* 31.1 (2012): 47-56.

Godoy-Vitorino, F., Rodriguez-Hilario, A., Alves, A.L., Gonçalves, F., Cabrera-Colon, B., Mesquita, C.S., Soares-Castro, P., Ferreira, M., Marçalo, A., Vingada, J., *et al.* (2017). The microbiome of a striped dolphin (Stenella coeruleoalba) stranded in Portugal. Res. Microbiol. *168*, 85–93.

Gravesen, Suzanne, *et al.* "Microfungal contamination of damp buildings--examples of risk constructions and risk materials." *Environmental Health Perspectives* 107.Suppl 3 (1999): 505.

Handley, Kim M., *et al.,* "Metabolic and spatio-taxonomic response of uncultivated seafloor bacteria following the Deepwater Horizon oil spill." *The ISME journal* 11.11 (2017): 2569.

Harper, C.G., Feng, Y., Xu, S., Taylor, N.S., Kinsel, M., Dewhirst, F.E., Paster, B.J., Greenwell, M., Levine, G., Rogers, A., *et al.* (2002). Helicobacter cetorum sp. Nov., a Urease-Positive Helicobacter Species Isolated from Dolphins and Whales. J. Clin. Microbiol. *40*, 4536–4543.

Hartemink, A., and *et al.* (2005). Banjo: Bayesian Network Inference with Java Objects.

Hay, D. B., B. J. Hart, and A. E. Douglas. "Evidence refuting the contribution of the fungus Aspergillus penicillioides to the allergenicity of the house dust mite Dermatophagoides pteronyssinus." *International archives of allergy and immunology 97.1* (1992): 86-88.

Hegarty, Bridget, K. C. Dannemiller, and Jordan Peccia. "Gene expression of indoor fungal communities under damp building conditions: Implications for human health." *Indoor air* (2018).

Henry, C.S., Bernstein, H.C., Weisenhorn, P., Taylor, R.C., Lee, J.-Y., Zucker, J., and Song, H.-S. (2016). Microbial Community Metabolic Modeling: A Community Data-Driven Network Reconstruction. J. Cell. Physiol. 231, 2339–2345.

Herter, T., *et al.* "Glucose-1-phosphatase (AgpE) from Enterobacter cloacae displays enhanced phytase activity." *Applied microbiology and biotechnology* 70.1 (2006): 60-64.

Hoang, Chi P., *et al.,* "Resistance of green building materials to fungal growth." *International Biodeterioration & Biodegradation*64.2 (2010): 104-113.

Holmes, Gerald J., and Joseph W. Eckert. "Sensitivity of *Penicillium digitatum* and *P. italicum* to postharvest citrus fungicides in California." *Phytopathology* 89.9 (1999): 716-721.

Houser, D., Finneran, J., and Ridgway, S. Research with Navy Marine Mammals benefits animal care, conservation and biology. (2010). Research Gate.

Hyde, E.R., Navas-Molina, J.A., Song, S.J., Kueneman, J.G., Ackermann, G., Cardona, C., Humphrey, G., Boyer, D., Weaver, T., Mendelson, J.R., *et al.* (2016). The Oral and Skin Microbiomes of Captive Komodo Dragons Are Significantly Shared with Their Habitat. MSystems *1*, e00046-16.

Hyvärinen, Anne, *et al.,* "Fungi and actinobacteria in moisture-damaged building materials—concentrations and diversity." *International Biodeterioration & Biodegradation* 49.1 (2002): 27-37.

Institute of Medicine (US) Committee on Damp Indoor Spaces and Health. Damp Indoor Spaces and Health. Washington (DC): National Academies Press (US); 2004. Available from: https://www.ncbi.nlm.nih.gov/books/NBK215643/doi: 10.17226/11011

Isogai, Akira, *et al.,* "Isolation and identification of Nigragillin as an insecticidal metabolite produced by a *Aspergillus niger*." *Agricultural and Biological Chemistry* 39.3 (1975): 739-740.

Jaakkola, Maritta S., *et al.,* "Association of indoor dampness and molds with rhinitis risk: a systematic review and meta-analysis." *Journal of Allergy and Clinical Immunology* 132.5 (2013): 1099-1110.

Jaing, C., Thissen, J., Gardner, S., McLoughlin, K., Slezak, T., Bossart, G., and Fair, P. (2015). Pathogen surveillance in wild bottlenose dolphins Tursiops truncatus. Dis. Aquat. Organ. *116*, 83–91.

Jensen, Peter Ø., *et al.,* "Formation of hydroxyl radicals contributes to the bactericidal activity of ciprofloxacin against Pseudomonas aeruginosa biofilms." *Pathogens and disease* 70.3 (2014): 440-443.

Johansson, Pernilla, *et al.,* "Laboratory study to determine the critical moisture level for mould growth on building materials." *International Biodeterioration & Biodegradation* 73 (2012): 23-32.

Johnson, W.R., Torralba, M., Fair, P.A., Bossart, G.D., Nelson, K.E., and Morris, P.J. (2009). Novel diversity of bacterial communities associated with bottlenose dolphin upper respiratory tracts. Environ. Microbiol. Rep. *1*, 555–562.

Kim, C.H. (2018). Immune regulation by microbiome metabolites. Immunology 154, 220–229.

Kong, Na-Na, *et al.* "Flavonoids from the halophyte Apocynum venetum and their antifouling activities against marine biofilm-derived bacteria." *Natural product research* 28.12 (2014): 928-931.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics 9*, 559.

Lax, S., Cardona, C., Zhao, D., Winton, V.J., Goodney, G., Gao, P., Gottel, N., Hartmann, E.M., Henry, C., Thomas, P.M., *et al.* (2019). Microbial and metabolic succession on common building materials under high humidity conditions. *Nature Communications*. 10.

Lax, S., Sangwan N., Smith D., Larsen P., Hadley K.M., *et al.* (2017). Bacterial colonization and succession in a newly opened hospital. *Science translational medicine* 9.391 (2017): eaah6500.

Lax, S., Smith, D.P., Hampton-Marcell, J., Owens, S.M., Handley, K.M., Scott, N.M., Gibbons, S.M., Larsen, P., Shogan, B.D., Weiss, S., *et al.* (2014). Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science 345*, 1048–1052.

LBNL Indoor Environment Group (2019a). Prevalence of Building Dampness. *IAQ Scientific Report*. https://iaqscience.lbl.gov/dampness-prevalence.

LBNL Indoor Environment Group (2019b). Nature and Causes of Building Dampness. *IAQ Scientific Report*. https://iaqscience.lbl.gov/dampness-nature.

LeBlanc, J.G., Laiño, J.E., del Valle, M.J., Vannini, V., van Sinderen, D., Taranto, M.P., de Valdez, G.F., de Giori, G.S., and Sesma, F. (2011). B-Group vitamin production by lactic acid bacteria - current knowledge and potential applications: Vitamin production by LAB. *J. Appl. Microbiol.* 111, 1297–1309.

Lerat, Sylvain, *et al.,* "Streptomyces scabiei and its toxin thaxtomin A induce scopoletin biosynthesis in tobacco and Arabidopsis thaliana." *Plant cell reports* 28.12 (2009): 1895-1903.

Liang, S., Webb, T., and Li, Z. (2014). Probiotic antigens stimulate hepatic natural killer T cells. Immunology *141*, 203–210.

Love, M., Anders, S., and Huber, W. (2014). Differential analysis of count data–the DESeq2 package. Genome Biol *15*, 550.

Lutz, J.K., Crawford, J., Hoet, A.E., Wilkins, J.R., and Lee, J. (2013). Comparative performance of contact plates, electrostatic wipes, swabs and a novel sampling device for the detection of Staphylococcus aureus on environmental surfaces. J. Appl. Microbiol. 115, 171–178.

Ma, X., Hua, J., and Li, Z. (2008). Probiotics Improve High Fat Diet-induced Hepatic Steatosis and Insulin Resistance by Increasing Hepatic NKT cells. J. Hepatol. *49*, 821–830.

Magdy, Wesam, *et al.* "Nigragillin, Nigerazine B and Five Naphtho-γ-pyrones from *Aspergillus japonicus* Isolated from Hot Desert Soil." *The Natural Products Journal* 7.3 (2017): 216-223.

Marshall, C.W., Ross, D.E., Handley, K.M., Weisenhorn, P.B., Edirisinghe, J.N., Henry, C.S., Gilbert, J.A., May, H.D., and Norman, R.S. (2017a). Metabolic Reconstruction and Modeling Microbial Electrosynthesis. Sci. Rep. 7, 8391.

Mendell, Mark J., *et al.,* "Respiratory and allergic health effects of dampness, mold, and dampness-related agents: a review of the epidemiologic evidence." *Environmental health perspectives* 119.6 (2011): 748.

Mille‐Lindblom, Cecilia, Helmut Fischer, and Lars J. Tranvik. "Antagonism between bacteria and fungi: substrate competition and a possible tradeoff between fungal growth and tolerance towards bacteria." *Oikos* 113.2 (2006): 233-242.

Miller, J. David, and David R. McMullin. "Fungal secondary metabolites as harmful indoor air contaminants: 10 years on." *Applied microbiology and biotechnology* 98.24 (2014): 9953-9966.

Nascimento, M. S., *et al.,* "Phenolic extractives and natural resistance of wood." *Biodegradation-Life of Science*. InTech, 2013.

Neilson, Julia W., *et al.,* "Significant Impacts of Increasing Aridity on the Arid Soil Microbiome." *MSystems* 2.3 (2017): e00195-16.

Orth, J.D., Thiele, I., and Palsson, B.Ø. (2010). What is flux balance analysis? *Nat. Biotechnol. 28*, 245–248.

Pasanen, A-L., *et al.,* "Occurrence and moisture requirements of microbial growth in building materials." *International Biodeterioration & Biodegradation* 30.4 (1992): 273-283.

Petras, Daniel, *et al.,* "Mass spectrometry-based visualization of molecules associated with human habitats." *Analytical chemistry* 88.22 (2016): 10775-10784.

Pinheiro, Eduardo Antonio A., *et al.,* "Antibacterial activity of alkaloids produced by endophytic fungus *Aspergillus sp. EJC08* isolated from medical plant Bauhinia guianensis." *Natural product research* 27.18 (2013): 1633-1638.

Pons, Pascal, and Matthieu Latapy. "Computing communities in large networks using random walks." *International symposium on computer and information sciences*. Springer, Berlin, Heidelberg, (2005).

Powers, Matthew J., *et al.,* "Inhibition of cell differentiation in *Bacillus subtilis* by *Pseudomonas protegens*." *Journal of bacteriology* (2015): JB-02535.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE* 5, e9490.

Quansah, Reginald, *et al.,* "Residential dampness and molds and the risk of developing asthma: a systematic review and meta-analysis." *PLOS ONE* 7.11 (2012): e47526.

Raaijmakers, Jos M., *et al.,* "Natural functions of lipopeptides from *Bacillus* and *Pseudomonas*: more than surfactants and antibiotics." *FEMS microbiology reviews* 34.6 (2010): 1037-1062.

Rand, Thomas G., *et al.* "Inflammation-associated gene expression in RAW 264.7 macrophages induced by toxins from fungi common on damp building materials." *Toxicology in Vitro* 43 (2017): 16-20.

Raveh-Sadka, T., Thomas, B.C., Singh, A., Firek, B., Brooks, B., Castelle, C.J., Sharon, I., Baker, R., Good, M., Morowitz, M.J., *et al.* (2015). Gut bacteria are rarely shared by co-hospitalized premature infants, regardless of necrotizing enterocolitis development. *ELife 4*, e05477.

Ronan, Evan, *et al.* "Interspecies interaction extends bacterial survival at solid–air interfaces." *Biofouling* 29.9 (2013): 1087-1096.

Rose, L., Jensen, B., Peterson, A., Banerjee, S.N., and Arduino, M.J. (2004). Swab Materials and *Bacillus anthracis* Spore Recovery from Nonporous Surfaces. *Emerg. Infect. Dis.* 10, 1023–1029.

Round, J.L., and Mazmanian, S.K. (2009). The gut microbiota shapes intestinal immune responses during health and disease. *Nat. Rev. Immunol*. *9*, 313–323.

Roze, Ludmila V., Sung-Yong Hong, and John E. Linz. "Aflatoxin biosynthesis: current frontiers." *Annual review of food science and technology* 4 (2013): 293-311

Saez-Lara, M.J., Gomez-Llorente, C., Plaza-Diaz, J., and Gil, A. (2015). The Role of Probiotic Lactic Acid Bacteria and Bifidobacteria in the Prevention and Treatment of Inflammatory Bowel Disease and Other Related Diseases: A Systematic Review of Randomized Human Clinical Trials. *BioMed Res. Int*. *2015*, 1–15.

Sender, R., Fuchs, S., and Milo, R. (2016). Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. *Cell* 164, 337–340.

Shank, Elizabeth A., *et al.* "Interspecies interactions that result in *Bacillus subtilis* forming biofilms are mediated mainly by members of its own genus." *Proceedings of the National Academy of Sciences* (2011): 201103630.

Smelt, M.J., de Haan, B.J., Bron, P.A., van Swam, I., Meijerink, M., Wells, J.M., Faas, M.M., and de Vos, P. (2012). L. plantarum, L. salivarius, and L. lactis Attenuate Th2 Responses and Increase Treg Frequencies in Healthy Mice in a Strain Dependent Manner. *PLOS ONE 7*, e47244.

Soverini, M., Quercia, S., Biancani, B., Furlati, S., Turroni, S., Biagi, E., Consolandi, C., Peano, C., Severgnini, M., Rampelli, S., *et al.* (2016). The bottlenose dolphin (Tursiops truncatus) faecal microbiota. FEMS *Microbiol. Ecol*. *92*.

Stein, M.M., Hrusch, C.L., Gozdz, J., Igartua, C., Pivniouk, V., Murray, S.E., Ledford, J.G., Marques dos Santos, M., Anderson, R.L., Metwali, N., *et al.* (2016). Innate Immunity and Asthma Risk in Amish and Hutterite Farm Children. *N. Engl. J. Med*. *375*, 411–421.

Stephens, Brent. "What have we learned about the microbiomes of indoor environments?." *MSystems* 1.4 (2016): e00083-16.

Thaiss, C.A., Zeevi, D., Levy, M., Zilberman-Schapira, G., Suez, J., Tengeler, A.C., Abramson, L., Katz, M.N., Korem, T., Zmora, N., *et al.* (2014). Transkingdom Control of Microbiota Diurnal Oscillations Promotes Metabolic Homeostasis. *Cell 159*, 514–529.

Thurber, Rebecca V., *et al.,* "Laboratory procedures to generate viral metagenomes." *Nature protocols* 4.4 (2009): 470.

Trosvik, P., and de Muinck, E.J. (2015). Ecology of bacteria in the human gastrointestinal tract—identification of keystone and foundation taxa. *Microbiome 3*.

Van Bonn, W., LaPointe, A., Gibbons, S.M., Frazier, A., Hampton-Marcell, J., and Gilbert, J. (2015). Aquarium microbiome response to ninety-percent system water change: Clues to microbiome management: Aquarium Microbiome Management. *Zoo Biol. 34*, 360–367.

Venn-Watson, S., Smith, C., and Jensen, E. (2008). Primary bacterial pathogens in bottlenose dolphins Tursiops truncatus: needles in haystacks of commensal and environmental microbes. *Dis. Aquat. Organ. 79*, 87–93.

Venn-Watson, S.K., Jensen, E.D., and Ridgway, S.H. (2011). Evaluation of population health among bottlenose dolphins (Tursiops truncatus) at the United States Navy Marine Mammal Program. J. Am. *Vet. Med. Assoc*. *238*, 356–360.

Viitanen, Hannu, *et al.,* "Moisture and bio-deterioration risk of building materials and structures." *Journal of Building Physics*33.3 (2010): 201-224.

Vlasova, A.N., Kandasamy, S., Chattha, K.S., Rajashekara, G., and Saif, L.J. (2016). Comparison of probiotic lactobacilli and bifidobacteria effects, immune responses and rotavirus vaccines and infection in different host species. *Vet. Immunol. Immunopathol. 172*, 72–84.

Walters, William, Embriette R. Hyde, Donna Berg-Lyons, Gail Ackermann, Greg Humphrey, AlmaParada, Jack A. Gilbert, Janet K. Jansson, J. Gregory Caporaso, Jed A. Fuhrman, Amy Apprill, Rob Knight. "Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys" *mSystems* Dec 2015, 1 (1) e00009-15; DOI: 10.1128/mSystems.00009-15

Wattam, A.R., Davis, J.J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., Conrad, N., Dietrich, E.M., Disz, T., Gabbard, J.L., *et al.* (2017). Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res. 45,* D535–D542.

Wells, R.S. (2009). Learning from nature: bottlenose dolphin care and husbandry. *Zoo Biol*. *28*, 635–651.

Wells, R.S., McHugh, K.A., Douglas, D.C., Shippee, S., McCabe, E.B., Barros, N.B., and Phillips, G.T. (2013). Evaluation of Potential Protective Factors Against Metabolic Syndrome in Bottlenose Dolphins: Feeding and Activity Patterns of Dolphins in Sarasota Bay, Florida. Front. *Endocrinol*. *4*.

Wood, A.P., and Kelly, D.P. (2010). Skin Microbiology, Body Odor, and Methylotrophic Bacteria. In Handbook of Hydrocarbon and Lipid Microbiology, K.N. Timmis, ed. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 3203–3213.

Yang, Zhao, René Algesheimer, and Claudio J. Tessone. "A comparative analysis of community detection algorithms on artificial networks." *Scientific Reports* 6 (2016): 30750.

Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J.D., Osterhout, R.E., Stephen, R., *et al.* (2011). Metabolic engineering of *Escherichia coli.* for direct

production of 1,4-butanediol. Nat. Chem. Biol. 7, 445–452.

Zhang, G., Ma, L., and Doyle, M.P. (2007). Salmonellae Reduction in Poultry by Competitive Exclusion Bacteria *Lactobacillus salivarius* and *Streptococcus cristatus*. J. Food Prot. *70*, 874–878.

Zhang, J., Deng, J., Wang, Z., Che, C., Li, Y., and Yang, Q. (2011). Modulatory Effects of *Lactobacillus salivarius* on Intestinal Mucosal Immunity of Piglets. Curr. Microbiol. *62*, 1623–1631.