RELATIONSHIPS AMONG AIRBORNE MICROBIAL COMMUNITIES, URBAN LAND USES AND VEGETATION COVER: IMPLICATIONS FOR URBAN PLANNING AND HUMAN HEALTH

by

GWYNHWYFER MHUIREACH

A DISSERTATION

Presented to the Department of Landscape Architecture
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2018

DISSERTATION APPROVAL PAGE

Student: Gwynhwyfer Mhuireach

Title: Relationships Among Airborne Microbial Communities, Urban Land Uses and Vegetation Cover: Implications for Urban Planning and Human Health

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Landscape Architecture by:

Bart R. Johnson	Chairperson
Jessica L. Green	Core Member
Roxi Thoren	Core Member
Deb Johnson-Shelton	Core Member
G.Z. Brown	Institutional Representative

and

Sara D. Hodges	Interim Vice Provost and Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2018

DISSERTATION ABSTRACT

Gwynhwyfer Mhuireach

Doctor of Philosophy

Department of Landscape Architecture

June 2018

Title: Relationships Among Airborne Microbial Communities, Urban Land Uses and Vegetation Cover: Implications for Urban Planning and Human Health

Variation in exposure to environmental microbial communities has been implicated in the etiology of allergies, asthma and other chronic and immune disorders. In particular, preliminary research suggests that exposure to a high diversity of microbes during early life, for example through living in highly vegetated environments like farms or forests, may have specific health benefits, including immune system development and stimulation. In the face of rapidly growing cities and potential reductions in urban greenspace, it is vital to clarify our understanding of the relationship between vegetation and microbial communities so that we can better design cities that support human health. To explore whether and how urban airborne bacterial communities vary with the amount and structural diversity of nearby vegetation, I used passive air sampling and culture-independent microbial DNA sequencing combined with more traditional landscape architecture tools, including geographic information systems (GIS) and remote sensing data. The results indicated that locations with little vegetation (i.e., paved parking lots) were marked by significantly different bacterial composition from areas that were heavily vegetated (parks and forests). These differences were largely driven by taxonomic groups and indicator species that were enriched at certain sites. My work also shows that regional agricultural activities during the summer may have a substantial effect on airborne bacterial communities in the Eugene-Springfield metropolitan area (Oregon), specifically through elevated abundance of *Sphingomonas faeni*, a taxon previously isolated from hay dust.

The second part of my work focused on building a conceptual bridge between scientific findings and potential design principles that can be tested in practical application. I performed

a narrative review of vegetation-health, vegetation-microbe, and microbe-health relationships, which formed the foundation of a framework to translate scientific findings into design-relevant concepts. Strengthening this linkage between science and design will help ensure that research questions are relevant to design practice and that new scientific knowledge is accessible to designers.

This dissertation includes previously published and unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Gwynhwyfer Mhuireach

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon

University of Washington, Seattle, Washington

DEGREES AWARDED:

Doctor of Philosophy, Landscape Architecture, 2018, University of Oregon

Master of Science, Architecture, 2012, University of Oregon.

Bachelor of Science, Biology (Ecology, Evolution and Conservation Track), 1999, University of Washington.

AREAS OF SPECIAL INTEREST:

Microbial Ecology

Environmental Health

PROFESSIONAL EXPERIENCE:

Graduate Teaching Fellowship, Department of Landscape Architecture, 2012 - 2017.

Graduate Research Fellowship, Energy Studies in Buildings Laboratory, 2012 - 2017.

GRANTS, AWARDS, AND HONORS:

University of Oregon Dissertation Research Fellowship, 2017

Travel grant, Department of Defense Microbiome Symposium, US Air Force Academy, Colorado Springs, CO, 2017.

Travel grant, National Academies of Science, 3rd Meeting of Microbiomes of the Built Environment Committee, Irvine, CA, 2016.

National Academies Fellow, graduated from National Academies Summer Institute on Undergraduate Education, Eugene, OR June 2015.

EPA Science To Achieve Results (STAR) Fellowship, 2014.

Invited participant, Evolution of Built Environment Microbiome Working Group, National Evolutionary Synthesis Center, Raleigh, NC, 2013.

Donna V. Sundberg Scholarship in Architecture, 2011.

PUBLICATIONS:

Mhuireach, G., Johnson, B. R., Altrichter, A. E., Ladau, J., Meadow, J. F., Pollard, K. S., & Green, J. L. (2016). Urban greenness influences airborne bacterial community composition. Sci Total Environ, 571, 680–687.

Brown, G.Z., Kline, J., Mhuireach, G., Northcutt, D. and Stenson, J., (2016). Making microbiology of the built environment relevant to design. Microbiome, 4(1), p.1.

Martin, L. J., Adams, R. I., Bateman, A., Bik, H. M., Hawks, J., Mhuireach, G., ... Dunn, R. R. (2015). Evolution of the indoor biome. Trends in Ecology & Evolution, 30(4), 223–232.

Kembel, S. W., Meadow, J. F., O'Connor, T. K., Mhuireach, G., Northcutt, D., Kline, J., ... Green, J. L. (2014). Architectural Design Drives the Biogeography of Indoor Bacterial Communities. PLOS ONE, 9(1), e87093.

Meadow, J. F., Altrichter, A. E., Kembel, S. W., Kline, J., Mhuireach, G., Moriyama, M., ... Bohannan, B. J. M. (2013). Indoor airborne bacterial communities are influenced by ventilation, occupancy, and outdoor air source. Indoor Air, 24(1), 41–48.

Mhuireach, G., Bohannan, B. J. M., Brown, G. Z., Green, J. L., Jones, E., Kembel, S. W., ... Womack, A. M. (2012). The relationship between ventilation energy use and indoor microbial communities. In Proceedings of the 10th International Conference on Healthy Buildings. Brisbane, Australia.

ACKNOWLEDGMENTS

I profoundly and sincerely appreciate the assistance of Professor Bart Johnson, without whose penetrating insight and sustained enthusiasm this project would never have been completed. Doctor Jessica Green, my co-advisor, deserves special thanks for her extraordinary wisdom and for always having my best interests in mind. I am deeply grateful to Professor Roxi Thoren for her wonderful advice and genuine interest in designing urban landscapes for microbes. I would like to recognize Doctor Deborah Johnson-Shelton, my friend and collaborator, with whom I look forward to many new research endeavors. Professor G.Z. Brown, my long-time mentor, started me on this path and has been a guiding light throughout my graduate career; I appreciate and honor him more than I can express.

I wish to thank all of the members, past and present, of the Biology and Built Environment (BioBE) Center and the larger international community of researchers studying microbiology of the built environment, who almost feel like family after working together and meeting at conferences for so many years. In particular, I am grateful to James Meadow, Joshua Ladau, Adam Altrichter, Clarisse Betancourt-Román, for their assistance with the microbiology aspects of my project and their patience in teaching me DNA processing and bioinformatics skills. My heartfelt thanks also goes out to all of the Energy Studies in Buildings staff and researchers, with whom I've shared many wonderful (and some stressful) times throughout my graduate career. I am deeply beholden to Kevin Van Den Wymelenberg, Director of ESBL and Co-Director of BioBE, for continuing to consider me part of the team and offer support during the final stages of my PhD. I also gratefully acknowledge the assistance of numerous volunteers in collecting microbial samples, as well as the public and private landowners, including the City of Eugene Parks and Open Space and Willamalane Park and Recreation District, who permitted the collection of samples on their land.

My research was supported by the Department of Landscape Architecture at the University of Oregon, by the Biology and Built Environment Center, which was funded through generous grants from the Alfred P. Sloan Foundation, and by a Science to Achieve Results (STAR)

Fellowship, #91769501-0, from the US Environmental Protection Agency.

To Lyric and Cadence, who have coped with my graduate studies throughout their early years. All my love.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

Figure Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

Urbanization is a global phenomenon unprecedented in extent and magnitude. By 2050, the world population is expected to reach 9.3 billion. All population growth is predicted to occur in urban areas (U.N. 2012). Urban development is the most rapidly increasing land use type, and over 50% of humanity now resides in urban areas. We are becoming "*Homo urbanus*," in the words of Crane and Kinzig (2005), and the large-scale migration from rural to urban over the past century has been called "the greatest human-environmental experiment of all time" (Meyerson et al. 2007). This suggests that the urban condition should be a high-priority focus of study if we are to understand the implications of these changes to human habitat.

Cities are unlike environments in which humans evolved. Although the first cities appeared around 4000 BC, it wasn't until the rise of industrialization that a significant proportion of the global human population resided in cities (Davis 1955). Much of our evolutionary history was spent in forests, savannas, meadows and farmland, but over the last few centuries we have increasingly inhabited urban environments. We have only a rudimentary understanding of how this change has affected our health and well-being, yet we continue to modify urban environments. Cities are complex amalgamations of interdependent systems that are costly and difficult to alter once they are set on a particular trajectory. Urban policy, planning and design decisions that occur now will influence the lives of many generations to come.

Vegetation is a major component of the urban environment with recognized benefits for human well-being. Some of the many benefits of urban vegetation, or greenspace (see definitions of terms used in dissertation in Figure 1), include improving property values, increasing social cohesion, reducing the urban heat island effect, improving air quality, providing areas for recreation, promoting psychological well-being, and increasing biodiversity (Breuste 2008). Due to population growth and urbanization, it is predicted that in the future most people will have most of their interactions with nature in urban settings rather than wilderness.

Figure 1. Definitions of terms used.

Aerobiome: The consortium of airborne microbes in a given environment.

Commensal: Microbes that live on and in the body of other organisms.

Culture-based approaches: Methods that involve the use of sterile dishes containing growth substrates (e.g., agar) to grow and identify microbes.

Densification policies: Governmental policies that encourage or require population growth to be accommodated within urban boundaries to the extent possible.

Dysbiosis: An imbalance in the human commensal microbiome.

Greenspace: Any vegetated area within an urbanized region, including private yards and gardens, public parks, street trees, vacant lots, and waste ground.

Helminths: Internal parasites, such as tapeworms, that live within higher organisms.

High-throughput genome sequencing: Parallelized process to identify nucleotides within many DNA or RNA molecules concurrently.

Macrobiodiversity: Number of species of large (visible) plants and animals.

Management regimes: The frequency, duration, and type of human activities performed to maintain desired characteristics of a site.

Microbiodiversity: Number of species of tiny (invisible) microbes, including bacteria, fungi, and viruses.

Microbiome: The consortium of microbes found in a given environment.

Saprophytes: Microbes commonly associated with soil.

However, increasing urban density through infill development can result in overall reductions of vegetation within metropolitan areas. In highly urbanized regions of England, for instance, there has been significant overall loss of greenspace since implementation of densification policies in 2000 (Dallimer et al. 2011). Despite the recognized importance of urban greenspace, there are currently no clear national guidelines in the US for its provision. In 1981, the National Recreation and Parks Association published a guideline stating that cities should provide 10 acres of parkland for every thousand residents, but in 1995 replaced it with

a "level of service" recommendation suggesting that open space provision should be specifically tailored to each community. According to the Trust for Public Land (2011), "Some cities have plenty of parkland that's well distributed around town; others have enough land but an inequitable distribution; others are short of even a basic amount of park space for their citizens." A survey of the 50 largest cities in the US conducted by the Trust for Public Land in 2003 found wide disparities in the recommended maximum distance any given household should be from the nearest park (Harnik and Simms 2004). Only 18 of the surveyed cities had any standards in place and, of those 18, the standards ranged widely from 1/8–1 mile. The combination of vague guidelines, growing urban populations and shrinking budgets (Baur et al. 2014) has led to an overall trend of losses in urban greenspace per resident (American Planning Association 2013).

Furthermore, there is wide disparity in the provision of greenspace across different socioeconomic groups (Sherer 2003). Researchers have found that residents of disadvantaged neighborhoods tend to have less vegetation near their homes and poorer access to public greenspace (Wolch et al. 2005; Harlan et al. 2006; García and White 2006; Coalition for a Livable Future 2007; Fitzpatrick and LaGory 2000; Bell and Ebisu 2012; Dai 2011; Jennings et al. 2012; Landry and Chakraborty 2009; Dai et al. 2011; Wen et al. 2013; Astell-Burt et al. 2014; Lakes et al. 2014). New research shows that not only is there an existing gap between greenspace provision in advantaged versus disadvantaged neighborhoods, but that the gap is increasing. From 2001 to 2011, US neighborhoods with higher ethnic populations lost overall greenness while those with higher white populations gained greenness (Casey et al. 2017). There is also evidence that urban biodiversity varies with socioeconomic status—people living in higher socioeconomic status neighborhoods (i.e., more household income, less unemployment, lower population density) have significantly greater species richness near their homes than do those living in lower socioeconomic status neighborhoods (Strohbach et al. 2009; Hope et al. 2003; Kinzig et al. 2005).

The Biodiversity Hypothesis posits that macrobiodiversity (e.g., urban greenspace) is intrinsically related to microbiodiversity, and that human health is indivisibly linked with both

macro- and microbiodiversity (von Hertzen et al. 2011). This hypothesis builds from earlier work suggesting that modern lifestyles have led to a severe reduction in exposure to "Old Friends," or the various microbes and helminths with which humans co-evolved, thus disrupting immune training and development. Development of these theories was, in part, prompted by observations that increasing susceptibility to a suite of maladies associated with modern lifestyles has paralleled growth in urbanization (Jackson 2003).

The increasing prevalence of chronic and autoimmune disorders that were relatively rare or unknown before industrialization, including asthma, allergies, autism, depression, obesity, inflammatory bowel disorder (IBD), diabetes, and cancer, has been linked to lack of greenspace exposure, although results are sometimes mixed (Alcock et al. 2017; Ruokolainen et al. 2015; Wu and Jackson 2017; Cox et al. 2017; Dadvand et al. 2014; Timm et al. 2014; Astell-Burt et al. 2014; Li et al. 2008). This apparent relationship between health and urbanization may be partially explained by lifestyle changes, such as decreased physical activity and increased exposure to air pollutants, both of which are also associated with nearby greenspace. According to the Biodiversity Hypothesis, reduced exposure to microbial diversity may be another important factor, since all of the disorders listed have also been shown to be associated with microbial dysbiosis, or imbalance in the human commensal microbiome—allergies and asthma (Fujimura and Lynch 2015), autism (Vuong and Hsiao 2017), mental health disorders (Lowry et al. 2016), obesity (Walters et al. 2014), IBD (Round and Mazmanian 2009), diabetes (Tilg and Moschen 2014), and cancer (Bultman 2016).

Thus, provision of biodiverse urban greenspaces may promote increased microbial diversity, which may generate health benefits. For example, Ege et al. (2011) compared children living on farms with those in a reference group and linked exposure to higher microbial diversity with a lower prevalence of asthma and allergic sensitivity. In a separate study, Hanski et al. (2012) found significant correlations between exposure to highly vegetated environments and skin microbiota diversity, both of which were associated with lower likelihood of allergic disposition. These empirical studies are supported by theoretical work seeking to elucidate how the observed

protective effects of diverse microbial exposure might occur (Eder et al. 2006; Rook et al. 2013; Raison et al. 2010; Raison and Miller 2013; von Hertzen et al. 2011). Current hypotheses propose that particles of microbial origin, such as endotoxins, beta-glucans, polysaccharides and muramic acid, may play a key role in protecting against chronic and autoimmune disorders by signaling components of the immune system to tolerate environmental exposures (Heederik and von Mutius 2012). It has been suggested that one of the most important classes of microbes we have lost contact with are saprophytic, or soil-living, microbes and that this class may be particularly protective against development of allergic disorders (von Hertzen and Haahtela 2005). The bacterial phyla Proteobacteria and Actinobacteria include many members of common saprophytic microbes (e.g., *Acinetobacter* sp., *Mycobacterium vaccae*), which may have beneficial health effects (Fyhrquist et al. 2014; Lowry et al. 2016). Vegetation and soil-associated microbes such as these are typically enriched in farming and natural environments, and this could help explain why asthma and allergy are typically more prevalent in urban areas, since cities have largely paved over and otherwise degraded the majority of their soils. As an example, societies that maintain a primarily hunter-gatherer lifestyle tend to have greater exposure to environmental microbes and lower incidence of allergic sensitivity than modernized societies (Brown et al. 2013).

Evolutionary theory may help explain why humans require this exposure to microbial diversity to maintain good health. Humankind evolved under constant exposure to outdoor air containing a complex mixture of microbes originating from plants, soil, water, and other animals. Every day we move through a veritable ocean of microbes, and we coexist with countless numbers of them throughout our lives. Over the past hundred years we have implemented strategies of impermeable surface construction, water treatment, industrialized agriculture, urban inhabitation, and landscape maintenance regimes (e.g., pesticides, fertilizers, mowing) that have almost certainly changed our degree of exposure to the environmental microorganisms associated with natural elements (Brown et al. 2013). Blaser and Falkow (2009) describe this condition as a case of our "disappearing microbiota," wherein we have largely lost

contact with the microbes that have accompanied us throughout evolutionary history. It is possible that urban environments comprise relatively new assemblages of microbes on Earth, given these environments are characterized by unique combinations of biotic and abiotic conditions that have not been found historically in natural habitats (McKinney 2006; Young et al. 2009; Kowarik 2011).

The prevailing view of microbes as generally harmful creatures that should be eliminated from primary human environments (with the exception of those few microbes involved in food and drink fermentation processes) has begun to change as we find increasing evidence of their ubiquity and importance to our health. In contrast to culture-based approaches, which are only effective for a small percentage of microbial taxa, high-throughput genome sequencing allows us to "see" entire communities of microbes and begin to tease out the ecological relationships among microbes, their functional roles, and their interactions with larger organisms. Burgeoning scientific theory suggests that humans (and other macroorganisms) should be understood not only in terms of "self," but also as inseparable associations of host and commensal microbes (Gilbert et al. 2012). This marks a shift toward a more nuanced understanding of microbial interactions with humans and their role in promoting health (Sachs 2007). Although our understanding of relationships among macrobiodiversity, microbiodiversity, and human health continues to grow, numerous questions remain to be answered before designers, planners, and policymakers can implement health-promoting urban design strategies. The following are among the most pressing design-relevant knowledge gaps identified in the literature (specifically Rook et al. 2013; Mills et al. 2017):

- How to determine optimal vegetation species for green spaces;
- Whether plant (and animal) diversity is associated with health-relevant microbes;
- Whether ecological restoration of degraded urban sites leads to re-establishment of a vegetation-associated microbiome;
- What are the relative contributions of soil, air, leaf surfaces, and animals or animal products to urban microbial communities; and

- What activities promote human interaction with urban microbial communities.

With my dissertation research I sought to begin filling these gaps by exploring how nearby vegetation may influence the structure and dynamics of airborne microbial communities (collectively, the aerobiome). I focused on the aerobiome because it is a largely unavoidable exposure route for urban residents and its composition may be highly dependent on local sources, such as vegetation. To accomplish this, I first executed a pilot study on the variation in airborne bacterial communities across disparate urban land cover types (parks and parking lots), which constitutes Chapter II of this dissertation. Building on the foundations laid by the pilot study, for my third chapter I conducted a large-scale aerobiome sampling campaign of three land cover types (forest, grass, paved). In this campaign, samples were collected from 50 sites at four time points, with the objective of unraveling the complex interactions between vegetation types, management regimes, and bacterial communities across space and time. In Chapter IV, I explored how the scientific results from Chapters II and III, as well as evidence from the literature, can be translated into a conceptual framework linking landscape design with microbial ecology and human health.

Ultimately, my objectives were to a) discover whether the airborne bacterial communities vary with urban vegetation cover or structure, and b) explore how landscape design could effect variation in the urban aerobiome, particularly as cities continue to expand and increase in population density. I hope that this work will further our understanding of how and why vegetation might influence the microbial communities urban residents interact with every day, and introduce a new facet of ecological design that has been largely unrecognized by landscape architects and planners.

This dissertation includes previously published and unpublished co-authored material. Chapter II was previously published in Sci Total Environ with Bart R. Johnson, Adam E. Altrichter, Joshua Ladau, James F. Meadow, Katherine S. Pollard, and Jessica L. Green as co-authors. Chapter III was prepared for submission to Environ Int with Clarisse M. Betancourt-Román, Jessica L. Green, and Bart R. Johnson as co-authors.

CHAPTER II

URBAN GREENNESS INFLUENCES AIRBORNE BACTERIAL COMMUNITY COMPOSITION

This paper was published in *Science of the Total Environment* in September 2016. I conceived the study, collected the samples, performed bioinformatics, analyzed the data, and wrote the paper. Assistance in experimental design, sampling, processing samples, data analysis, and writing the manuscript was provided by my co-authors: Bart R. Johnson, Adam E. Altrichter, Joshua Ladau, James F. Meadow, Katherine S. Pollard, Jessica L. Green.

1. Introduction

Human well-being in urban areas is linked to the abundance and degree of access to nearby greenspace (e.g., Maas 2006; Maas et al. 2009; Villenueve et al. 2012; Mitchell and Popham 2007; Dadvand et al. 2012; Donovan et al. 2013). However, the specific mechanisms linking health and greenspace are not well understood. New evidence indicates that exposure to microbial diversity, especially from soil, plants and some animals, is an understudied pathway through which health benefits may arise (von Hertzen and Haahtela 2006; Hanski et al. 2012; Fall et al. 2015).

Humans evolved under constant exposure to air, water, and soil containing a diversity of environmental microbes. However, over the past few centuries our lifestyles have shifted dramatically (indoor living, antibiotic use, processed food, chemical treatment of water, etc.) and, in the process, this has altered the abundance, diversity, and composition of the microbial communities to which we are exposed on a daily basis (Blaser and Falkow 2009). While these shifts in lifestyle have been associated with reduced incidence of many diseases, greater longevity, and other benefits, it is also now widely recognized that early life immunological experiences, including exposures to various environmental substances as well as the lack of exposures, are associated with the development of later life immune-mediated disease, such as

asthma, allergy and other inflammatory disorders (Russell et al. 2012; Rook et al. 2013; Ege et al. 2012).

There is evidence that airborne microbial communities vary across major land use types, e.g. forest, agricultural land and urban areas (Bowers et al. 2010; Shaffer and Lighthart 1997; Burrows et al. 2009). At local scales, spatial proximity is an important predictor of microbial community similarity among outdoor samples (Adams et al. 2013; Adams et al. 2014), although long-range transport is also known to play a significant role in shaping airborne bioaerosol composition (Barberán et al. 2015). Little is known, however, about how airborne microbial composition varies within an urban area, nor what factors influence its variation. Beginning to describe fine-scale biogeographic patterns, such as distance-decay relationships (i.e., the spatial distance at which similarity of microbial community composition breaks down) at the scale of urban blocks and neighborhoods, would be a valuable contribution to the scientific knowledge base (Womack et al. 2010).

Vegetation structure and composition could play a role in the localized variability of microbial communities. Plants are important sources of airborne microorganisms (Lindemann and Upper 1985; Lindemann et al. 1982; Bowers et al. 2011). It is estimated that leaf surfaces comprise the largest biological surface type on the planet -- over a billion km^2 -- and may host up to 10^8 bacteria per cm^2 (Delmotte 2009; Vorholt 2012; Peñuelas and Terradas 2013). These leaf-inhabiting microorganisms become airborne during plant processes, like evapotranspiration, as well as by meteorological processes, such as rain splash, wind gusts and thermal plumes (Lighthart et al. 2009; Whipps et al. 2008). Different species of plants are associated with different compositions and emission rates of microbes (Lambais et al. 2014; Jumpponen and Jones 2010; Lindow and Brandl 2003; Kembel et al. 2013; Vokou et al. 2012). Although culture-based investigations of the population dynamics of leaf-surface microbes concluded that, "nearby vegetation strongly influences the atmospheric microbial concentration and composition at a given location" (Kinkel 1997), there have been few studies using modern molecular techniques. A notable exception is the recent work by Lymperopoulou et al. (2016)

investigating the abundance and composition of airborne microbes in relation to the sample's proximity to vegetated versus non-vegetated areas. Their results showed that local vegetation could contribute up to half of the airborne bacteria found at a distance of 50 meters downwind.

The primary goal of this study was to explore how urban vegetation, spatial proximity of sample sites, and other factors influence the composition of airborne microbial communities, specifically focusing on bacteria (fungal analysis to follow in a separate paper). We hypothesized that locations with large amounts of vegetation would have different airborne bacterial communities than areas with little or no vegetation, and that spatial proximity would have less influence on composition than the amount of vegetation. In the longer term, this course of investigation has the potential to substantially change our understanding of how to design healthy urban neighborhoods.

2. Materials & Methods

2.1. Field sampling

We collected air samples for an eight-hour period on July 24th, 2013, beginning at 08:00. Six samples from each site were collected simultaneously at five pairs of parks and nearby parking lots in Eugene, Oregon (Figure 2a) for a total of 60 samples. The sampling station consisted of a custom tray (sterilized prior to use with 99 percent isopropyl alcohol) containing three passive settling dishes with their lids and three vacuum pump-powered button filters attached to the sides of the tray (Figure 2b-2c) placed approximately 2 meters above ground level in a relatively open area (i.e., not directly underneath tree canopy or other obstruction). SKC Button Samplers and SKC AirChex XR5000 Pumps (SKC Inc., Eighty Four, PA, USA) were set to draw 4 l/min (~1,920 l total for the sampling period) through 25 mm-diameter cellulose ester filters (1.4 lm pore diameter; autoclaved prior to sample collection), and HOBO U52 dataloggers (Onset Corporation) were used to measure temperature and relative humidity at 1-minute intervals. Technicians were present at each site to monitor the sampling equipment

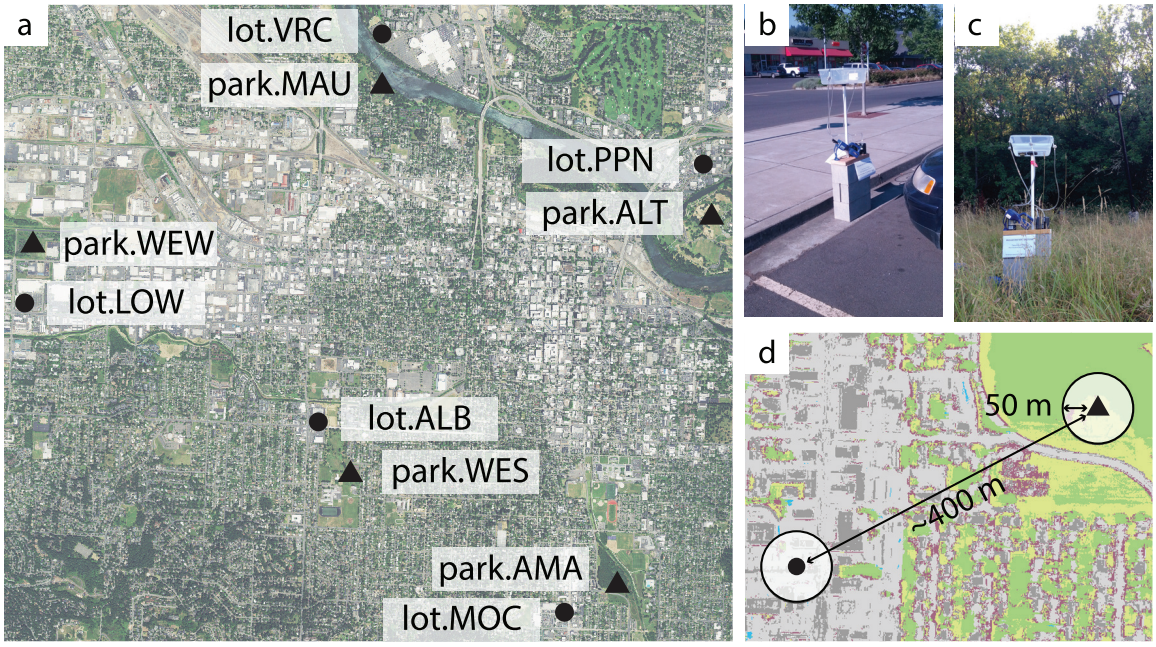


Figure 2. a) Map of paired park and parking lot sites; b) photo of sampling station in parking lot; c) photo of sampling station in park; and d) classification image showing 50-m buffer zones and approximate distance between park-lot pairs.

and perform hourly wind speed and direction measurements. All air samples were frozen at -80°C immediately following sampling and stored frozen until processing.

2.2. GIS data

Urban environmental characteristics were measured using ArcGIS 10.2 with geospatial data accessed from the Lane Council of Governments and the National Agriculture Imagery Program. All data layers were imported into a new geodatabase and re-projected to the NAD 1983 HARN StatePlane Oregon South (Feet Intl) Coordinate System, based on a Lambert Conformal Conic Projection. Six primary land cover types were identified (built, paved, dirt, grass, trees and shrubs, and water) using supervised maximum likelihood classification of aerial 4-band orthoimagery at 1-meter resolution. To assess the amount of surrounding vegetation, buffer zones of 50, 100, 200, 400, and 800 meter radii were created around each sampling point and the proportion of vegetated area (grass + trees and shrubs) within each buffer zone was calculated using the classified raster image (Figure 2d and Table 4 in Appendix A).

2.3. DNA amplification and sequencing

The petri dishes and their lids were swabbed with nylon flocked swabs (copanusa.com; 552C), and DNA was extracted directly from the swabs and filter samples using the MO BIO PowerWater DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA, USA) according to manufacturer's instructions. Protocols followed those of Meadow et al. (2013), and negative controls were included at each step to evaluate potential contamination. We amplified the V4 region of the bacterial 16S rRNA gene using F515/R806 primers (5' - GTGCCAGCMGCCGCGG - 3' , 5' - TACNVGGGTATCTAATCC - 3') (Caporaso et al. 2012; Claesson et al. 2010). The samples were sequenced as paired-end reads at the Dana-Farber/ Harvard Cancer Center DNA Resource Core (Boston, MA, USA; dnaseq.med.harvard.edu) using the Illumina MiSeq platform.

2.4. Analyses

Raw sequences were processed using the FastX Toolkit and QIIME pipeline (Caporaso et al. 2010). After recombining the barcodes from paired-end reads, forward reads were used for analysis due to lower quality reverse reads. In quality filtering, sequences that did not meet a 30 quality score over at least 75% of the read, or had ambiguous bases, or more than one primer mismatch, were removed. Sequence read lengths were trimmed to 250 bp, and taxonomy assignment was performed on a reference set of high-quality sequences using the open-reference OTU picking function in QIIME, which uses UCLUST (Edgar 2010). OTU clusters with a 97% similarity were identified using the Greengenes 13.5 database (DeSantis et al. 2006). All sequence files and metadata can be found in the FigShare data repository (FigShare DOI: 10.6084/m9.figshare.3362344).

2.5. Statistical analyses

Plant and mitochondrial sequences, sequences occurring fewer than three times, and the top three most abundant potential contaminants observed in our negative control samples

were removed prior to statistical analyses. The potential contaminants that were excluded were *Alicyclobacillus* sp., *Bradyrhizobium* sp., and *Shewanella algae*, altogether comprising slightly more than 60% of the sequences recovered from negative controls. It should be noted that the negative controls contained approximately 1% of the number of sequences that the study samples contained, therefore it is unlikely that any remaining contamination skewed our results. Three of the actively collected samples (two from the same park and one from a parking lot) did not meet the minimum criteria of 25,000 reads and were also eliminated from the analyses. Statistical analyses were implemented in R (R Development Core Team 2010) using chiefly the DESeq2, phyloseq, and vegan packages (Love et al. 2014; McMurdie and Holmes 2013; Oksanen et al. 2016). All analysis code is shown in Appendix B. The variance-stabilizing transformation function in the DESeq2 package was used to adjust for unequal sample library sizes. The three samples for each site and sampling method were pooled before executing statistical analyses. We performed a Mantel test using Spearman rank correlation to test for spatial autocorrelation among sampling sites. The Morisita-Horn dissimilarity index was employed for beta diversity calculations because it has been shown to perform well when there is variability in sampling depth and when under-sampling is suspected (Huse et al. 2013). We used constrained analysis of principal coordinates (CAP), which requires the cloud of sample points to be plotted along orthogonal vector(s) that directly represent the explanatory variable(s). This method can uncover significant environmental effects on compositional differences even in "noisy" data (Erb-Downward et al. 2012). The adonis function from the vegan package, which performs Permutational Multivariate Analysis of Variance (PERMANOVA), was used to examine the statistical significance of compositional differences between parks and parking lots. Testing for differentially abundant taxa was achieved using the DESeq function in the DESeq2 package, which adjusts for testing multiple hypotheses by applying the Benjamini–Hochberg method (Benjamini & Hochberg 1995). This function automatically filters out data that is unlikely to have statistical significance and does so independently of the factors being studied (Love et al. 2014). All differentially abundant taxa that distinguish parks from parking lots were

further identified in the NCBI 16S isolate database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to obtain putative species assignments.

3. Results

We collected air samples from five pairs of parks and parking lots to assess whether nearby vegetation influences airborne bacterial composition. After quality control and initial filtering there were 5,762,173 total reads, representing 16,633 operational taxonomic units (OTUs) from 40 unique bacterial phyla, based on 97 percent sequence similarity. The number of sequences recovered from each site ranged from 379,687 - 721,208. One OTU (*Sphingomonas* sp.) dominated the samples, comprising almost a quarter of all observed sequences. *Hymenobacter*, *Pedobacter*, *Agrobacterium*, and *Rhodococcus* spp. were also in the top ten most abundant OTUs and are common soil-associated bacteria (Oren 2006; Steyn et al. 1998; Matthysse 2006; Bell et al. 1998). About 83% of taxa were found in both parks and parking lots, and there was no significant difference in alpha-diversity as measured by the Shannon-Weaver index (Figure 24 in Appendix A; lot mean = 4.76, park mean = 4.90, t = -1.66, df = 15.1, p = 0.12).

3.1 Passive and active sampling methods give comparable results

At each site we collected three active and three passive samples to verify whether the passive settling dish method gives satisfactory results. Active samples had slightly higher alpha diversity than passive samples (active mean = 4.94, passive mean = 4.72, p = 0.0086), but composition was quite similar (Figure 25 in Appendix A). To further evaluate the correspondence of active versus passive collection, we used PERMANOVA to test the null hypothesis that sampling method does influence composition for the top 50 OTUs, representing over 68% of sequences. Results of this test showed that sampling method explains nearly zero percent of observed variation (R^2 = 0.001, p = 0.89). Since the two methods were found to be comparable, all further analyses were performed on passive samples only. We chose to analyze the passive

samples instead of active samples because: a) three of the active samples failed; b) passive samples had higher biomass; and c) passive sampling is more cost-effective and easier to deploy for future studies.

3.2 Nearby vegetation influences community composition

Airborne bacterial communities from parks and parking lots were significantly different ($R^2 = 0.148$, $p = 0.032$) in our PERMANOVA analysis. When we ran the same analysis using vegetation cover within 50 meters of the sampling station instead of the site type, the model gave a similar result ($R^2 = 0.15$, $p = 0.023$). A constrained PCoA shows this result in ordination space (Figure 3), where the x-axis is constrained by the 50 meter vegetation cover gradient. No

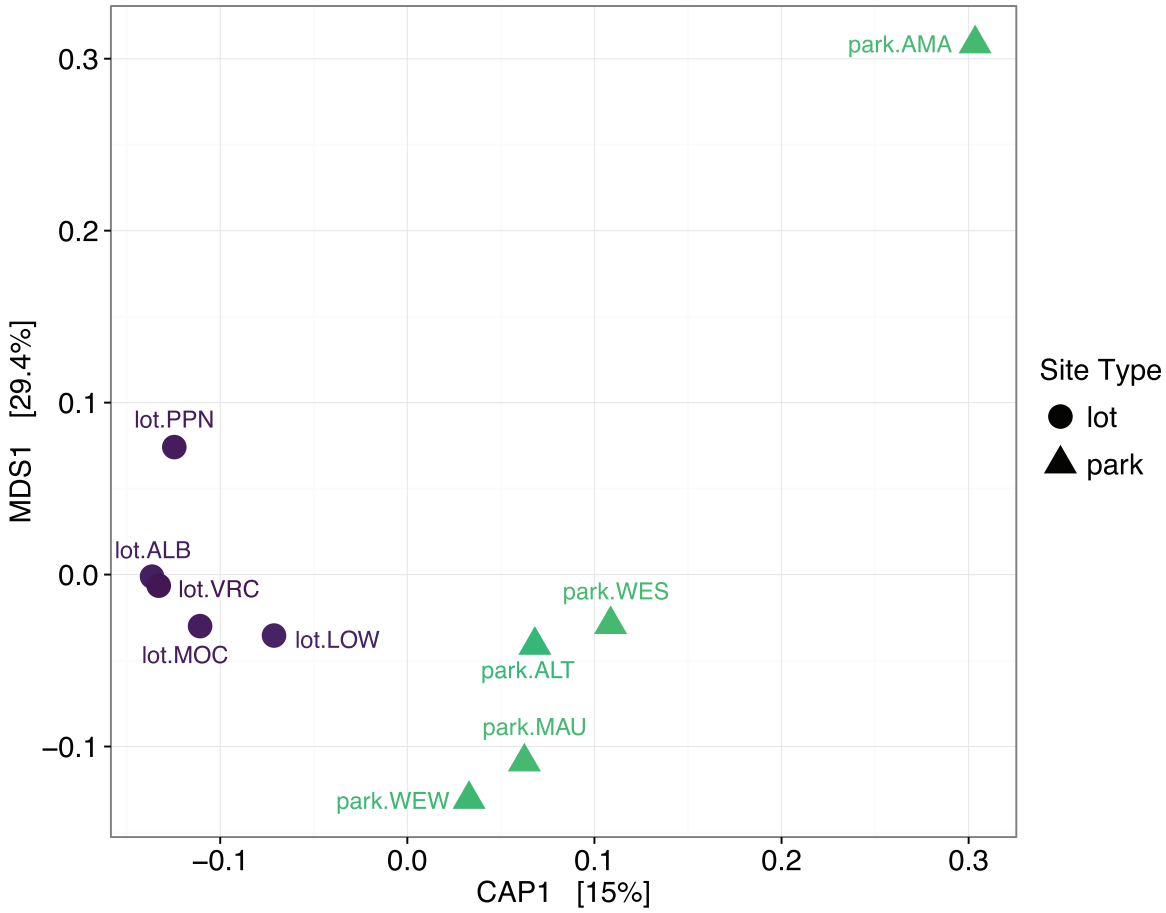


Figure 3. Constrained ordination of samples from parks (green triangles) and parking lots (purple circles), where the horizontal axis uses proportion of vegetated area within 50 m as a constraint.

other buffer zone radius (100, 200, 400, or 800 meters) improved the model fit and, in fact, we found significant negative linear relationships between buffer radius and both R^2 and *p*-value (Figure 4 and Table 5 in Appendix A). The potential model improvement using the 50 meter radius vegetation was strongly constrained because all parking lots were close to 0% vegetated area and all parks were close to 100% as a result of the sampling design (Table 4 in Appendix A).

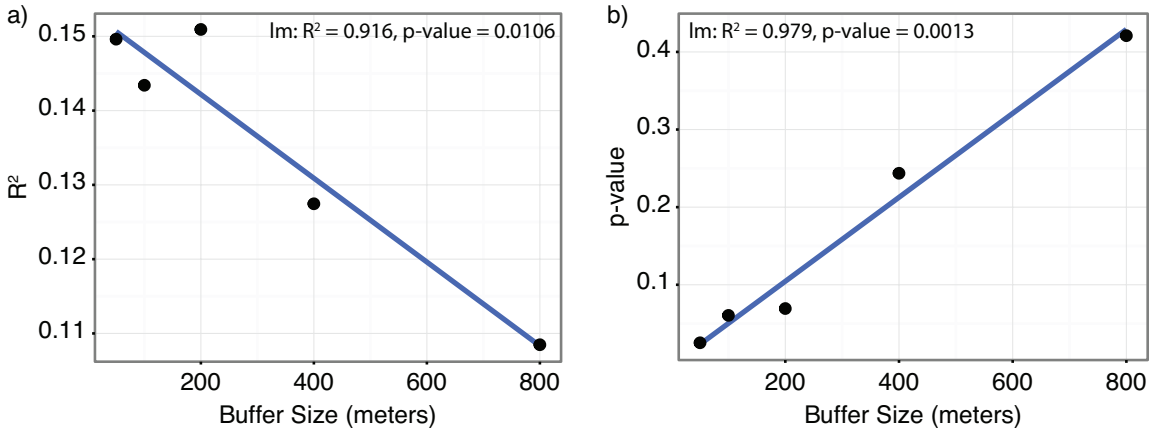


Figure 4. Change in a) explanatory power and b) statistical significance of the PERMANOVA model with varying buffer zone radii.

3.3 Key bacterial families differentiate parks from parking lots

Several key bacterial taxa were identified as differentially abundant either in parks or parking lots using a generalized linear model based on the negative binomial distribution. There were 23 OTUs identified as significantly more abundant in parking lots, seven of which were Acetobacteraceae, and 44 OTUs that were significantly more abundant in parks, 15 of which were Acidobacteriaceae (Figure 5 and Table 6 in Appendix A; note that some OTUs were unable to be matched precisely in the NCBI database and therefore taxa names may occur more than once). Altogether, despite accounting for only 0.4% of all OTUs, these 67 differentially abundant OTUs for both parks and parking lots comprised 13.6% of all the sequences recovered from passive sampling. Individual OTUs in this group ranged from 0.0087% - 1.62% relative abundance, which may be rare in comparison to the most abundant

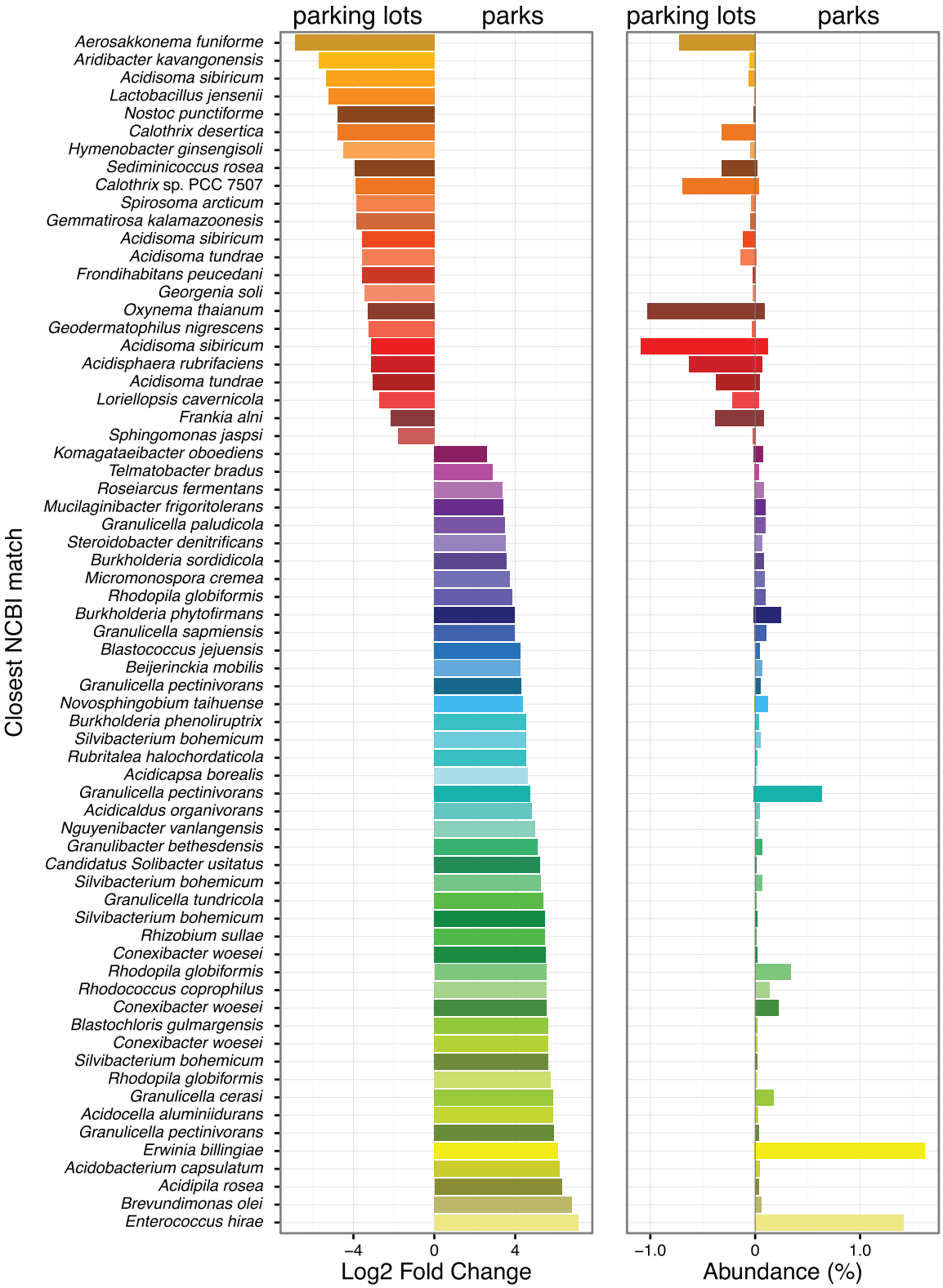


Figure 5. Differentially abundant bacterial OTUs in parks (purples, blues, greens, yellows) and parking lots (reds, oranges), labeled by closest match in NCBI database. Left panel indicates the degree to which each OTU is differentially abundant, right panel shows the actual relative abundance of each OTU.

OTUs but are not inconsequential, especially given the fact that microbial communities tend to be typified by having a handful of highly abundant taxa and an extremely long 'tail' of low-abundance taxa (Shade et al. 2014). Parking lots tended to have similar compositions of these differentially abundant OTUs, whereas parks were more variable (Figure 6). It should be noted that the OTUs shown in Figure 5 are only those that were found to be differentially abundant and do not include the entire community. The fact that the samples were paired spatially did not significantly influence community composition (Spearman correlation = 0.013, significance = 0.497), suggesting that the distance to which the influence of any site-scale characteristics might extend is less than ~400 meters (the average distance separating park-parking lot pairs).

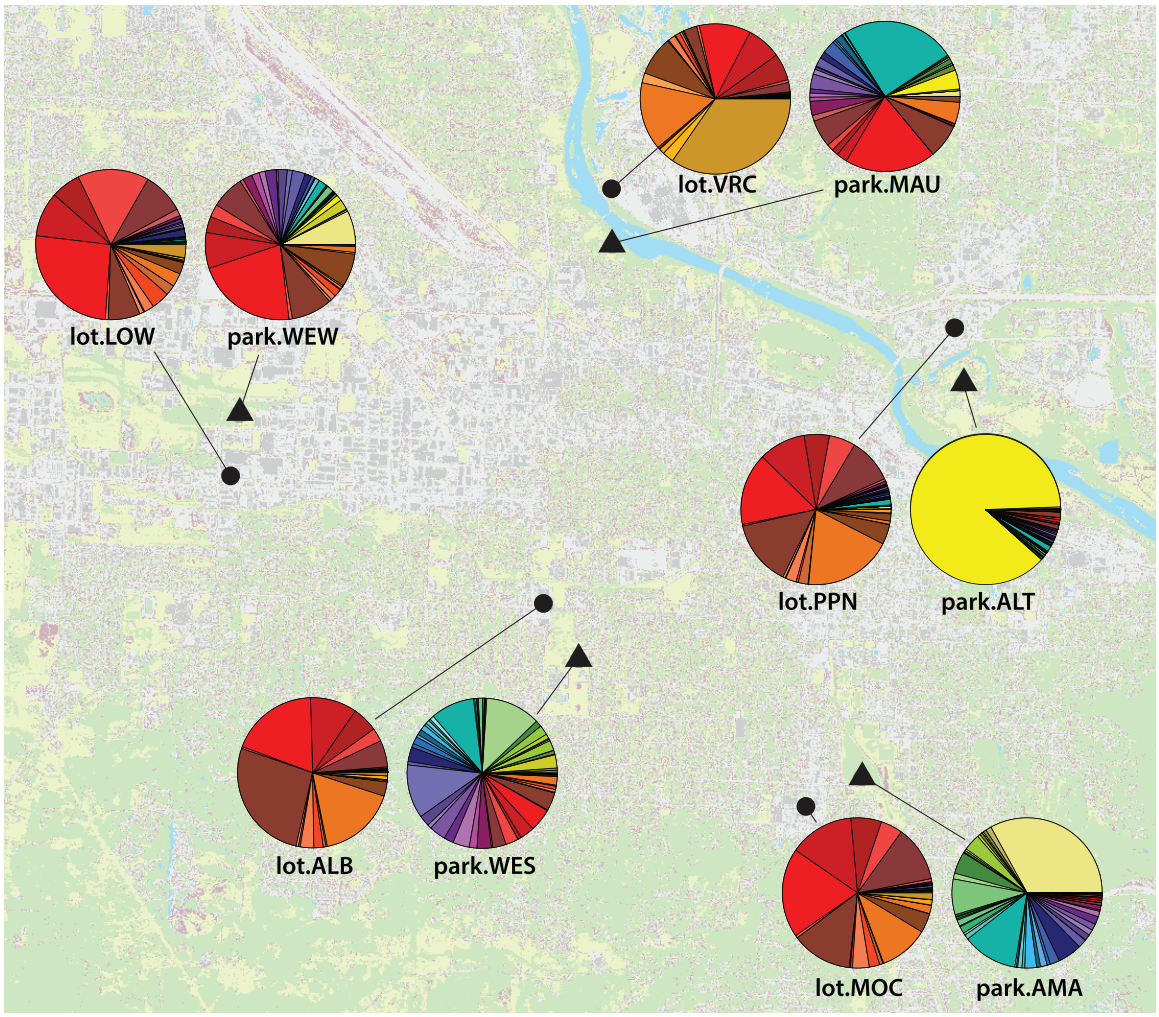


Figure 6. Composition of differentially abundant taxa for each site, colored same as Figure 5.

4. Discussion

We investigated the heterogeneity of airborne bacterial communities within an urban area and the role of vegetation as a potential driver of variation. We found that the most prevalent taxa were highly abundant at all sites and were primarily comprised of plant- and soil-associated bacteria. The consistent abundance of this large suite of common taxa may have been influenced by our decision to sample only in open areas with short herbaceous vegetation but no shrub layer or overhanging tree canopy to influence air movement. In these open areas, it is likely that the larger air mass moving through the region has a stronger influence on the composition of urban airborne microbial communities than individual site characteristics. For example, the most abundant taxon observed in this study, *Sphingomonas* sp., comprised almost a quarter of all sequences collected. Sphingomonads are commonly present in soil and on plant surfaces; they are considered ubiquitous across numerous species of higher plants (Kim et al. 1998; Innerebner et al. 2011). A BLAST search identified this OTU as either *S. faeni* or *S. aurantiaca*, both of which had been previously observed from hay dust (Busse et al. 2003). Our sampling period coincided with prime grass harvesting season in the region and Linn County, colloquially known as "the grass seed capital of the world," is immediately north of Eugene. As the dominant wind direction is from the north during the time of year in which we sampled it is perhaps likely that airborne microbial community composition at our sampling locations was influenced by upwind regional agricultural activities. In fact, earlier researchers estimated that grass harvesting in Linn County may contribute up to 40% of the total bacterial load in the Willamette Valley airshed (Lighthart 1984).

At the same time, a relatively small component (13.6%) of the microbial communities clearly distinguished parks from parking lots. Related to this, each park tended to have its own unique bacterial signature of indicator taxa, whereas parking lots were more similar to each other (Figure 6). These observations may be explained by the relatively homogeneous environmental conditions of parking lots—they are dry, covered in asphalt, exposed to high amounts of solar radiation, and receive continual inputs of heavy metals and fossil fuel products

from automobiles. Thus, we might expect that parking lot microbial signatures are determined by the ability of certain taxa to persist in extreme conditions. Parks, on the other hand, vary widely in the plant species present, vegetation structure and layering, human management regimes, and landscape design. We conjecture that park microbial signatures may be governed by some, or all, of these factors, which would explain their wide variation. Of the five parks that were chosen for this exploration, one was mown and irrigated, two were undergoing prairie restoration, and one was sheltered from wind by a small urban forest. In particular, it seems likely that vegetation serves both as a source, emitting microbes, as well as a modifier of airflow, which could tend to retain locally emitted microbes in some situations.

As yet, we know little about the spatial scale of such influences, but it is notable that the proportion of vegetated area in the smallest buffer radius (50 meters) provided the best separation of parks and parking lots, and successively larger buffer sizes produced poorer results. In terms of urban design, the distinction between the relatively strong association of vegetation and bacterial communities at smaller buffer sizes (50, 100, and 200 meters) and the weak relationship at larger buffer sizes (400 and 800 meters) suggests that the "park-like microbiome" extends less than 400 meters. However, we also noted that, in contrast to the distinct separation of parks and parking lots shown in Figure 3, an unconstrained PCoA ordination showed two samples collected from parks (park.MAU and park.WEW) clustering near the parking lot samples (Figure 26 in Appendix A). These two parks had less vegetation in the larger surrounding area (Table 4 in Appendix A, 800 meter buffer) than any other site sampled, including the parking lots. Stated differently, in this study parking lot communities always resembled those from other parking lots, regardless of vegetation in the larger area; park communities were generally distinct from parking lots, but those that had less vegetation in their larger surrounding area were more similar to parking lots than those that had plentiful vegetation within the 800 meter buffer zone. Additionally, the differentially abundant bacterial signatures in Figure 6 show that both park.MAU and park.WEW have a relatively high proportion of one OTU (closest NCBI match: *Acidisoma sibiricum*, Accession #:

NR_042706.1, 95% match) that is also prevalent in all parking lots. Possibly this indicates built and paved environments as a distinct source of that OTU (and other OTUs identified by reds and browns in Figure 6).

Currently the major evidence that suggests a potential linkage between human health and vegetation-associated microbes comes from studies at coarse spatial scales where differences in vegetation are confounded with differences in land use. For example, Hanski et al. (2012) found a relationship among diversity of skin Gammaproteobacteria, prevalence of atopic sensitization, and land use (forest, agriculture, built, water bodies, and wetlands), where the spatial scale of measure was a 3 kilometers radius around the home. A larger follow-up study also documented associations among the relative abundances of several classes of Proteobacteria on the skin of healthy (versus atopic) individuals and land use within 2–5 kilometers around the home (Ruokolainen et al. 2015). It remains to be seen whether such effects are related to specific vegetation factors, such as biomass, structural diversity, or species composition, and whether fine-scale differences such as those found in cities would play a role. Our results suggest that if plant-associated microorganisms are shown to be beneficial for human health, planners and designers might consider provisioning urban residents with greenspace within 400 meters of their homes.

5. Conclusion

Although we know that urban greenspace has significant health benefits, we don't know the exact mechanism(s) through which those benefits arise. By 2050, the world population is expected to reach 9.3 billion and all population growth in the next 35 years is projected to occur in urban areas, bringing the percentage of people living in cities to about 66% (U.N. 2012). At the same time, cities are being built more densely so as to reduce their impact on surrounding landscapes by creating a smaller spatial footprint. This may have the side effect of reducing large green open spaces as well as the amount of vegetation in residential neighborhoods. We cannot predict how further loss of urban vegetation would affect human health, nor do we

know enough about the mechanisms through which vegetation influences human health and well-being to design urban greenspace to maximize health benefits. Here, we provide the first evidence of fine-scale variability in outdoor urban microbiomes due to difference in vegetation. Future research may be able to elucidate how urban vegetation composition and structure, and open space distribution, influence urban airborne microbial communities, and in turn the degree to which this may influence human health. The current study thus provides a foundation for understanding how urban greenspace design impacts microbial communities, which could in time provide landscape architects and other urban design professionals the ability to better design cities and neighborhoods to foster human health.

Bridge to Chapter III

In the foregoing chapter, I established that there is a significant compositional difference in airborne bacterial communities from urban parks and parking lots. I also demonstrated that passive collection of airborne microbes via settling dishes provided similar results to active collection via vacuum pumps and was a defensible method to use in the large-scale collection campaign that follows in Chapter III. In the next chapter I will report the results of our large-scale study, which investigates the dynamics of airborne bacterial communities across space and time. This work endeavors to answer practical questions that will help designers and planners build healthier cities. To accomplish this, I will assess whether landscape features with which designers and planners work (e.g., land cover type, vegetation amount and composition, management activities) influence airborne bacterial diversity and/or composition. Finally, I will attempt to untangle factors that drive airborne bacterial community assembly across different scales of space and time.

CHAPTER III

SPATIOTEMPORAL CONTROLS ON THE URBAN AEROBIOME

This paper was prepared for submission to *Environment International*. I conceived the study, collected the samples, processed the samples, performed bioinformatics, analyzed the data, and wrote the paper. Assistance in experimental design, sampling, processing samples, data analysis, and writing the manuscript was provided by my co-authors: Clarisse M. Betancourt-Román, Jessica L. Green and Bart R. Johnson.

1. Introduction

The ethical code of landscape architecture requires practitioners to protect public health, safety, and welfare. In a written address on the Leadership and Governance page of the American Society of Landscape Architects website (2013), Richard Joseph Jackson, the Director of the National Center for Environmental Health says,

> "Despite the United States' expenditure of more than one of every seven dollars on medical care, the prevalence of chronic diseases, and the costs for their care in an aging population, will continue to increase. Of the strategies to limit these burdens, none will be as cost effective and improve the quality of life more than giving much more attention to how we plan and design our living environments."

The challenge is that we have limited knowledge of how to design urban environments to support optimal health. There is extensive evidence that vegetation is a key factor in creating healthy neighborhoods (Kuo 2015; Frumkin et al. 2017), but causal mechanisms and spatial effect sizes to support specific design recommendations remain elusive. The aerobiome (the consortium of airborne microbiota in a given environment) may represent a new dimension of environmental quality to be considered by planners and designers in much the same way as air pollution. A key difference, however, is that microbial exposures may have positive, as well as negative, health effects (Rook 2013).

Humankind evolved in the presence of airborne microbes associated with vegetation, soil,

water, and wildlife. Our immune systems are not only adapted to coexist with these microbes but may even require those interactions to function properly (Sorci et al. 2016; Rook et al. 2017). Indeed, it has been suggested that microbes may represent a direct pathway linking nature with health, since a lack of adequate childhood immune priming through exposure to commensal and environmental microbes has been implicated in a variety of autoimmune disorders later in life, including allergies, asthma, and mental disorders (Haahtela et al. 2013; Rook 2013; Raison and Miller 2013). For instance, studies have found that children growing up in biodiverse rural environments, like farms and forests, tend to have higher exposure to extremely diverse and novel microbial communities, compared to children in urban environments (Ege et al. 2011; Hanski et al. 2012; Ruokolainen et al. 2014). These rural children are also less likely to have allergies and asthma. In particular, greater diversity of genera within the class Gammaproteobacteria has been found on the skin of healthy individuals compared with atopic individuals, and abundance of *Acinetobacter* spp. (class Gammaproteobacteria) has been associated with elevated expression of anti-inflammatory compounds in healthy individuals (Hanski et al. 2012).

Human interactions with the aerobiome can occur through several different pathways—direct skin contact, inhalation, and ingestion of inhaled microbes after they are cleared from airways by mucociliary action. We now know that human skin acts not only as a barrier to environmental threats, but can actively respond to different external conditions, including exposure to microbes, by producing hormones and other biological compounds that influence whole-body state (Prescott et al. 2017). Similarly, cells lining airway passages and gastrointestinal surfaces are capable of signaling both the immune system and the neurological system following exposure to environmental microbes (Bene et al. 2017; Naik et al. 2012; Mulder et al. 2009). Previous work has noted that commensal microbial colonization appears to differ across environmental gradients, such as urban–rural (Ying et al. 2015; Lehtimaki et al. 2017; Camarinha-Silva et al. 2014; Kraemer et al. 2018). In sum, it appears possible that elements of the external environment, such as vegetation and soil, can affect human health even

when no physical contact occurs, because we constantly bathe in, inhale, and ingest ambient airborne microbes that are associated with those environmental elements.

A number of factors have been associated with variation in the aerobiome across space and time, including land use type, meteorological conditions, and season. Vegetation and soil, for instance, are important sources of microbes to nearby air (Bowers et al. 2013; Lindemann and Upper 1985; Bowers et al. 2011), and different plant species host distinct microbial communities (Laforest-Lapointe et al. 2016). Differences in vegetation composition and soil conditions may thus be partially responsible for observed variation in airborne microbial communities across regional-scale land use types, like forest, urban, and agricultural (Bowers et al. 2010; Bowers et al. 2013). At a finer spatial scale, vegetation has been shown to contribute a significant fraction of bacteria to the air within 50 meters downwind (Lymperopoulou et al. 2016). Mhuireach et al. (2016) suggested a link between these two spatial scales in a study of urban parks and parking lots: the composition of abundant airborne bacterial taxa was largely similar across the two land cover types, potentially indicating an influence of bacteria carried by the regional air mass, while differential abundances of some rare and moderately rare taxa suggested an effect of site-scale features, like vegetation.

These initial studies investigated only a limited number of sites and land cover types at single points in time. Whether certain land cover types consistently harbor distinctive microbial communities that maintain a detectable signal over time remains poorly understood. Furthermore, although differences in microbial communities have been observed across large-scale land uses like agricultural versus urban, few studies have examined whether there is substantial variation in composition across typical urban land cover types or vegetation cover at the scale of the city block or neighborhood. Critically, we also lack definitive mechanistic evidence linking variation in microbial composition to human health. Resolving these uncertainties is vital if urban designers and planners are to have sound evidence on which to base decision-making.

Importantly, the most appropriate vegetation metric(s) for capturing site-scale features

that exert influence on airborne microbial communities are unknown. A key challenge is that the spatiotemporal dynamics of these communities are complex due to the constant interplay between microbes transmitted by large regional and continental air masses and those emitted from local sources (Seifried et al. 2015; Innocente et al. 2017). We conjecture that microbial composition of the regional air mass is primarily influenced by large-scale periodic events, human activities, biotic processes, land use/land cover types, and physical geography, as well as macroclimatic influences such as prevailing winds and tropospheric transport from distant continents (see Womack et al. 2010; Lighthart 1984; Mhuireach 2016; Smith et al. 2012). At individual sampling sites, however, airborne microbial composition is likely driven by fine-scale land cover, microclimate, management, biotic processes, and physical geography features with greater retention or dilution of locally-sourced microbes depending on the degree of mixing with the regional air mass. The latter is likely affected by both local topography and nearby vegetation structure (Wuyts et al. 2008), suggesting a second role for site-scale vegetation in shaping airborne microbial communities. Disentangling these interactions at different scales of time and space is critical to understanding factors that control the assembly of, and variation within, the urban aerobiome, and to furnishing concrete evidence about which landscape features are important and how to best harness them for design interventions.

A key step to discerning whether urban design has the capacity to meaningfully influence airborne microbial communities is to ascertain the degree to which the site-scale microbiome is structured based on factors occurring across a gradient of relevant spatial and temporal scales from local to regional influences and in relation to daily to seasonal change. The challenge of assessing the processes that structure airborne microbial communities is that they are similar to those structuring oceanic systems where localized phenomena are dramatically influenced by large scale mixing and movement in a soluble media. With this in mind, we organized our sampling design to allow us to begin to infer site-scale versus landscape-scale influences in relation to change over time. Over the course of the study, our results and analysis led us to formulate a more specific framework to organize and present our current thinking, as well as

proposals for future investigation (Figure 7).

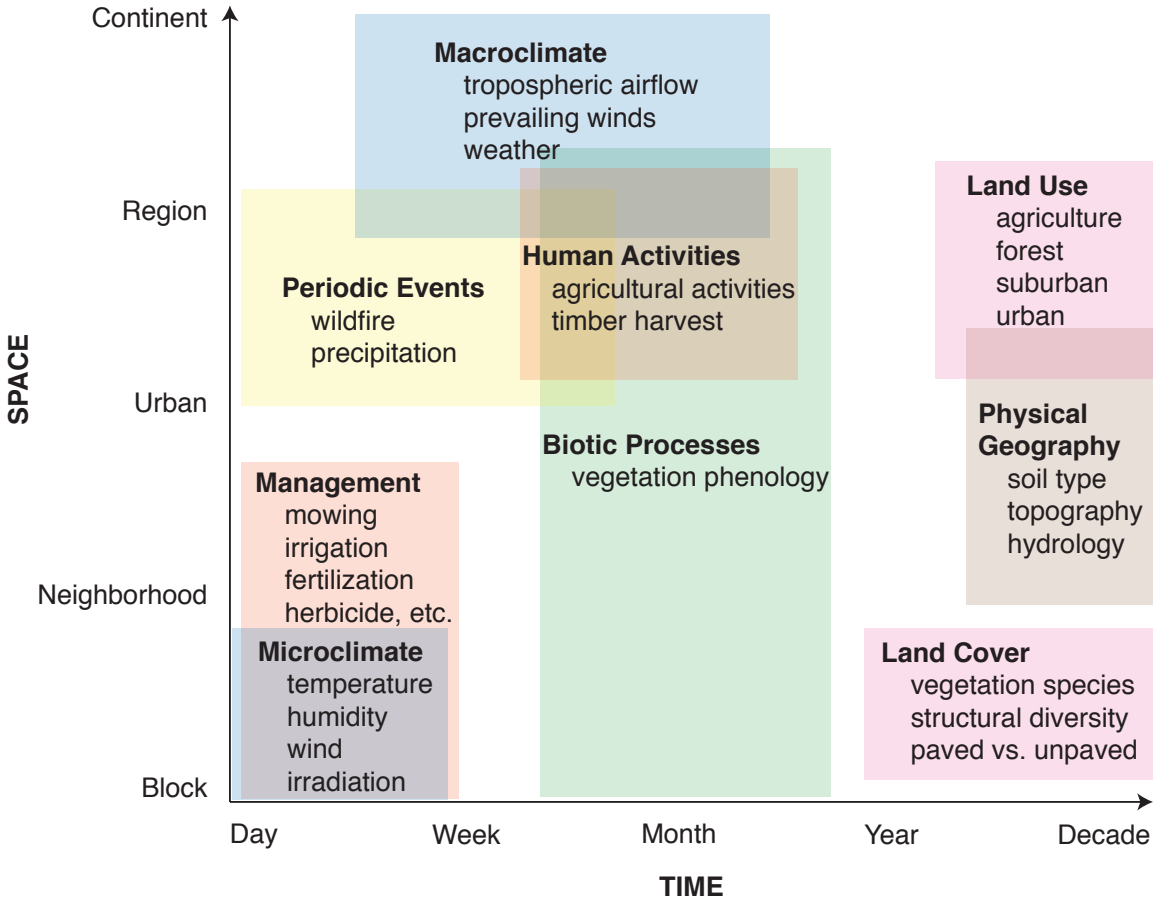


Figure 7. Conceptual model of potential controls on microbial community assemblage across spatiotemporal scales. The scale of this study is indicated by the burgundy star, although it should be recognized that factors across all scales bear on our results.

Providing a foundation for evidence-based design of urban vegetation is critical for contemporary urban planning, since rapid population growth and increasing density have decreased greenspace in many cities (Chen et al. 2017; Gan et al. 2014; Dallimer et al. 2011). Additionally, vegetation is often unequally distributed within urban areas and disproportionate vegetation loss often occurs in vulnerable and disadvantaged neighborhoods (Casey et al. 2017). If vegetation surrounding residents' homes does indeed alter airborne microbial communities in ways that could affect health, then promoting equitable and sufficient greenspace in cities is a public health necessity rather than a luxury. A more complete understanding of the relationships

among vegetation, the aerobiome, and human health, will help urban designers, planners, and policymakers effectively target investments in public infrastructure projects. To optimize investments, we must also supply practical design guidelines, such as how much and what type of vegetation and at what distance from residents' homes provides the greatest benefits.

In this study, we investigated the influences of urban vegetation on airborne bacterial communities across space and time. Our primary objectives were to understand: 1) whether common urban land cover types representing extremes of vegetation cover and structural diversity (i.e., forest, grassland, paved) exhibited differences in airborne bacterial richness or composition over time; 2) which metric(s) best captured environmental features that influenced bacterial community composition over time; and 3) which bacterial taxa contributed to differences in community composition across land cover types. We focused especially on landscape features that are human-controlled (e.g., land use, urban vegetation types, management activities), since these may represent prime lever points for influencing the aerobiome. As our knowledge of urban aerobiome structure and assembly grows, landscape architects and urban designers may be able to engineer greenspaces that help strengthen human immune systems, promote mental well-being, and protect against the chronic maladies that plague modern cities.

2. Methods

2.1. Overview and site descriptions

The study took place in the Eugene-Springfield metropolitan area of Oregon, USA, during July-September, 2015. Eugene-Springfield is located in the Willamette Valley, a region known for its agriculture, especially grass seed production. Weather during these months is typically warm and dry, with winds blowing primarily from the north. A total of 36 urban sites were selected, representing three land cover types — forest, grassland (e.g. urban turf or meadow) and paved. We attempted to select sampling sites such that vegetated sites (forest and grassland)

had greater than 75% vegetation cover within a 50-meter radius and non-vegetated sites (paved) had less than 25%. Within the vegetated sites, forest and grassland differed by both species composition and vertical structural diversity (Figure 8). Furthermore, we tried to stratify the sites across a broad range of vegetation cover within the greater surrounding context (i.e., within 800 meters), so that not all vegetated sites sat within large areas of high vegetation cover and not all paved sites sat within large areas of low vegetation cover.

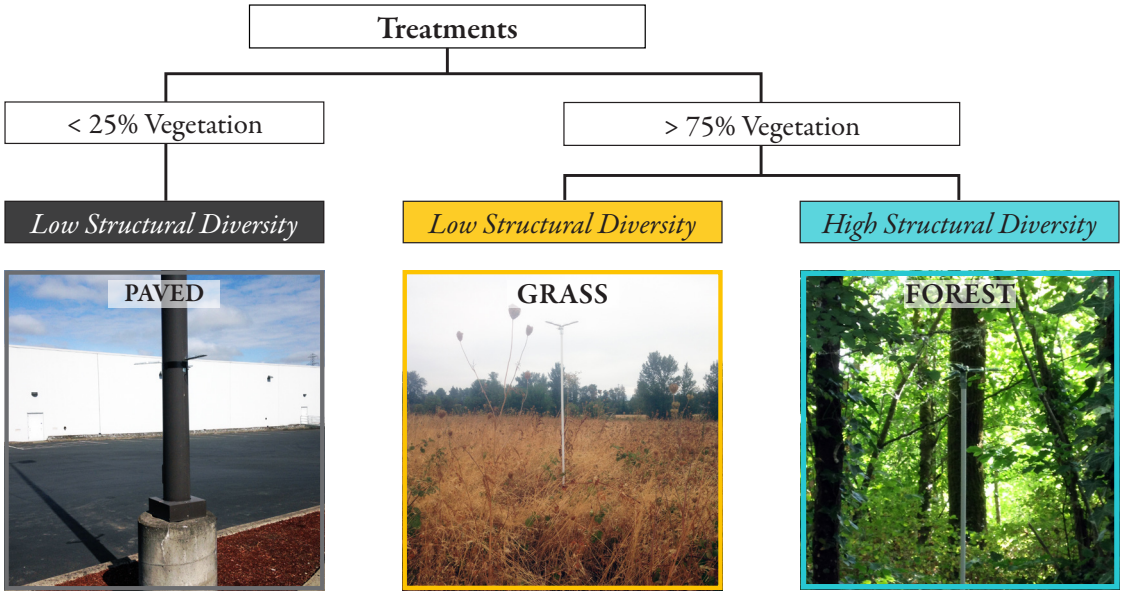


Figure 8. Sample selection criteria for the three land cover types.

Site locations were chosen to disperse samples from each land cover type across the study area (Figure 9). The majority of forest and grass sites were publicly owned parks and natural areas, while the paved sites were all privately owned business parking lots. Samples were collected in 2015 on July 14 (pilot), August 4, 14, 18 and 25, and September 1, 15, 22 and 29. The 36 sites were split into two groups of 18 to allow for set-up and take-down of all sampling stations within a specified time window while achieving adequate spatial representation. Samples were collected from each group every other sampling day during August and September, for a total of four samples per site, not including the July samples (Table 7 in Appendix C).

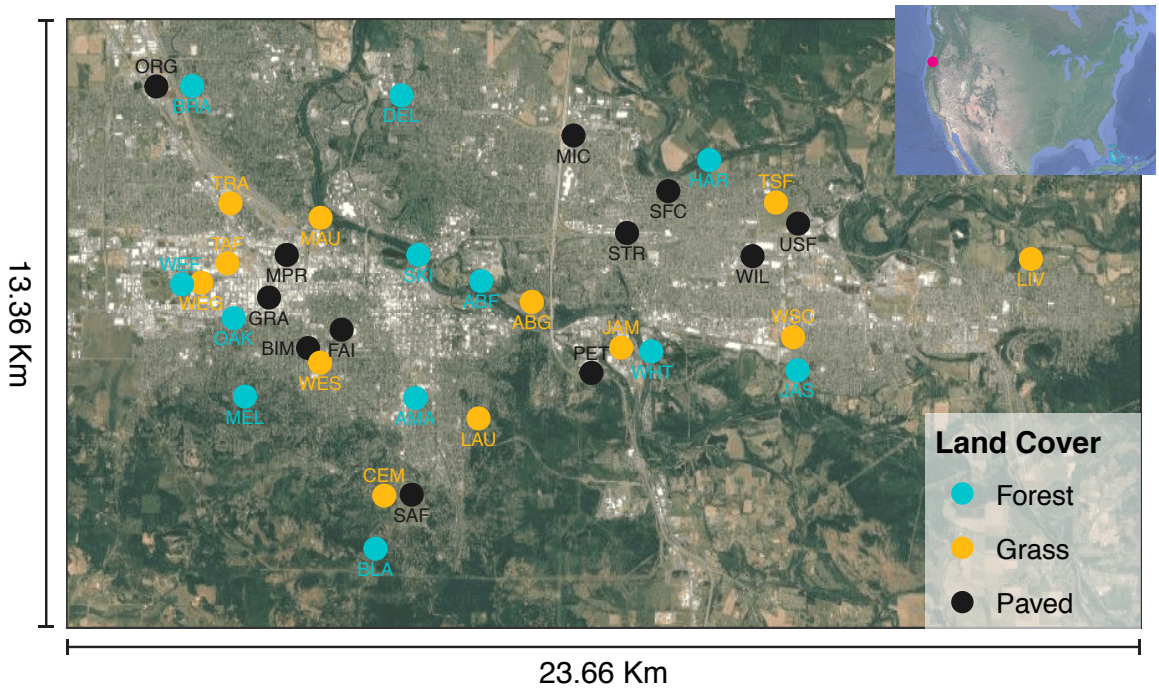


Figure 9. Sample selection criteria and site locations (only urban sites shown) in Eugene-Springfield, Oregon, USA.

2.2. Vegetation analysis

Vegetation analyses were performed using ArcGIS 10.3 (ESRI 2011) and FUSION (McGaughey 2016). Vegetation cover within 50, 100, 200, 400, and 800-meter circular buffer zones (i.e., radii from sampling point) around each site was measured by summing the proportions of trees and shrubs, green grass, and senesced grass within each buffer zone (Figure 10a). To do this, we differentiated land cover into six categories (trees and shrubs, green grass, senesced grass, dirt, and light-colored and dark-colored built surfaces) using supervised classification with maximum likelihood estimation of aerial 4-band orthoimagery (NAIP 2011). Water was lumped into the dark-colored built surface category.

Vegetation structural diversity within 25-meter buffers for forest and grassland sites only was measured using LiDAR (Light Detection And Ranging) point clouds for the study area (Figure 10b) obtained from the Oregon Department of Geology and Mineral Industries (DOGAMI). We extracted vertical cylinders from the point cloud and partitioned all the points into 1.5-meter height bins. The number of points within each height bin was used to calculate a

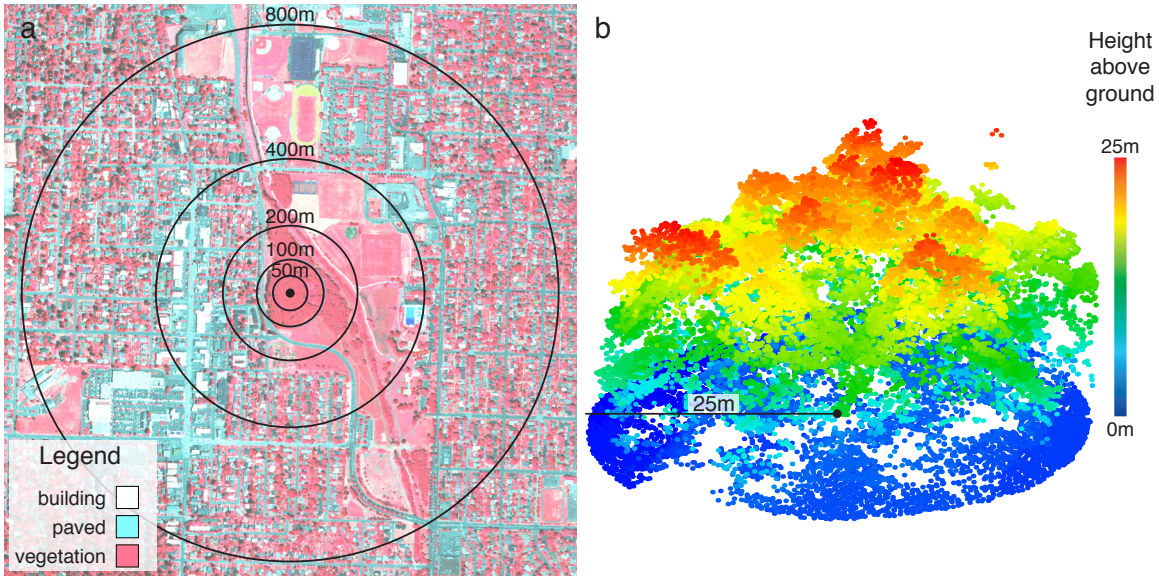


Figure 10. a) False color vegetation map (site AMA) with buffer zones; and b) LiDAR point cloud (site ABF).

Shannon-Weaver diversity index for each site.

2.3. Meteorological conditions

At each sampling location we measured air temperature at one-minute intervals for the entire sampling period using iButton dataloggers (Fondriest Environmental, #DS1921G) two meters above the ground. Prior to the sampling campaign iButtons were evaluated against a HOBO Datalogger (Onset, #U12) to ensure adequate accuracy. Results showed that all iButton measurements were within ±1 degree Celsius of the HOBO datalogger. Wind speed and direction data for the study area as a whole were obtained from the Eugene airport weather station after the sampling campaign ended, and the average, minimum, and maximum speeds, as well as dominant direction, were calculated for each sampling period.

2.4. Bacterial collection and analysis

We collected airborne bacteria at selected sites using passive settling dishes (sterile petris; Fisher Scientific, 100x15mm) for a 24-hour period on each sampling day. Passive settling dishes collect a similar sample of airborne bacteria to active methods, such as vacuum-powered

"button" samplers (Mhuireach et al. 2016; Leppanen et al. 2017). At each site, three settling dishes and their lids were installed two meters above the ground on portable sampling towers. Set-up occurred between the hours of 06:00 and 11:00 and take-down was during the same time period the following day. After collection, all samples were immediately placed on ice and transported to the lab for storage in a -80° F freezer.

The three petri dishes and lids from each sample site were individually swabbed with nylon-flocked swabs (COPAN FLOQSwabs, 502CS01) and then pooled before extraction. Samples were processed for sequencing following protocols used in Mhuireach et al. (2016). Manufacturer's instructions were used to extract DNA using MoBio PowerWater kits. Polymerase chain reaction (PCR) amplification was performed in triplicate using primers targeting the V4 region of the bacterial 16S rRNA gene (forward primer: 515; reverse primer: 806). PCR triplicates were pooled and cleaned, then sequenced at the University of Oregon on the Illumina NextSeq platform, (PE-150). All bioinformatic processing was performed in R, using the dada2 package (Callahan 2016). The dada2 workflow automates filtering, trimming, dereplication, inference of sequence variants, merging of paired-end reads, sequence table construction, chimera removal, and taxonomy assignment. After reviewing read quality plots, we truncated forward reads at 140bp and reverse reads at 135bp, which still allowed a minimum overlap of 20. Reads were also truncated when quality score dropped to 2 or lower. Reads with maximum expected error greater than 2 (calculated by EE = sum(10^(-Q/10))) were discarded. The dada2 package outputs amplicon sequence variants (SV), which represent individual taxa that are resolved down to single-nucleotide differences (Callahan et al. 2016). The Ribosomal Database Project (RDP) Classifier was used to assign taxonomy. SVs of interest were identified to putative species level by querying the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) with exact rRNA sequences when the RDP-assigned taxonomy was in question. Six samples (GRA-8.04, MIC-8.04, SFC-8.18, MAU-9.15, WEF-9.22, TEP-9.29), had fewer than our threshold of 10,000 sequence reads and were removed from downstream analyses, as were all the positive and negative controls.

Laboratory and reagent contaminations are well-documented problems that can plague low-biomass microbiome studies (Salter et al. 2014; Glassing et al. 2016). During sample processing prior to sequencing we attempted to reduce the impact of potential laboratory and reagent contaminants through the following protocols:

- Negative controls were included for all steps (collection, extraction, PCR);
- Positive controls were included for PCR;
- Sample extractions and PCR steps were all randomized so that different reagent kits would not be confounded with sampling date or other variables of interest; and
- All samples were placed on the same sequencing run.

In our post-sequencing bioinformatics procedure, potential contaminants were identified by listing all taxa observed in any of the negative controls (sterile petri dishes, extraction reagents, PCR reagents). Potentially consequential contaminants were determined by plotting the relative abundances of all taxa found in controls against their relative abundance in experiment samples (Figure 27, Appendix C). We found three taxa, identified as belonging to the genera *Vibrio*, *Methylobacterium*, and *Acinetobacter*, that were above our arbitrary 0.05 threshold for relative abundance in controls (all three were well under 0.05 relative abundance in collected samples) and removed them from downstream analyses.

2.5. Statistical analyses

We used R (version 3.5.0) for all statistical analyses, specifically the following packages: DESeq2, ggplot2, phyloseq, and vegan. All R analysis code is shown in Appendix D. Single- and double-tons were included for species richness calculations and removed prior to other analyses. We performed rarefaction analysis to compute species richness since our sample sizes were unequal, which can lead to inaccurate diversity index comparisons (Gihring et al. 2011). To accomplish this, we used the rarefy function in the vegan package, which we set to randomly subsample from all samples at the minimum sample size of 26,180 to estimate expected numbers of species. We then tested for significant differences in expected numbers of species using the

Kruskal-Wallis test, followed by pairwise Wilcoxon Rank Sum Test with Bonferroni correction for multiple testing.

For all beta diversity visualizations and analyses, we used the variance stabilizing transformation function in DESeq2, which adjusts for variation in dispersion due to differing sample sizes (Love et al. 2014), and the Morisita-Horn dissimilarity index. Overall temporal patterns in beta diversity were visualized using principal coordinates analysis (PCoA) ordination. We used direct gradient analysis in the form of a constrained PCoA to assess the amount of compositional variation attributable to land cover type. This is a useful technique for identifying real patterns in noisy data sets, since the ordination axes are required to be linearly related to the variable(s) of interest (Erb-Downward et al. 2012). To calculate statistically significant associations between bacterial composition and environmental variables of interest we used permutational multivariate analysis of variance (PERMANOVA; Anderson 2017), as implemented in the adonis function of the vegan package (Oksanen et al. 2018), where values were obtained using type III sums of squares with 9999 permutations of residuals under a reduced model. The PERMANOVA was performed first to test for the effect of land cover type while controlling for time (i.e., sampling date), followed by a post-hoc pairwise test also based on the adonis function (Martinez Arbizu 2017) to determine which land cover comparisons contributed most to observed variation. Site location, other vegetation metrics, and meteorological factors, were tested individually, since they were confounded in various ways. We performed PERMANOVA tests for land cover separately for each sampling date to evaluate effects that may have been obscured by the effect of time. Separately by date, we also executed a Mantel test based on Spearman's rank correlation to assess whether there was an effect of spatial correlation (i.e., samples close together in space being similar simply due to the fact of being close). Finally, generalized linear models (GLMs) based on the negative binomial distribution were executed using DESeq2 to determine which bacterial classes, families, and genera contributed to differences in community composition across land cover types. To perform this test, we split up the dataset into pairwise groups of forest versus paved, grass versus

forest, and grass versus paved sites. All statistical analyses used a significance level of $p < 0.05$ and were adjusted for multiple testing using Bonferroni correction.

3. Results

3.1. Vegetation analysis

Average vegetation cover proportion within 50 meters of each sampling site was 0.11 for paved sites, 0.90 for grass sites, 0.97 for forest sites (all vegetation analysis outputs shown in Table 8, Appendix C). Two of the forest sites were located within 50 meters of water bodies and thus have lower values for proportion of vegetation cover; one of the grass sites was later found to be artificial turf, which resulted in vegetation values similar to those of paved sites; and some of the paved sites were planted with individual landscaping planters, increasing their values to as high as 0.34. Proportion of vegetation cover within the 800-meter buffer ranged from 0.21 to 0.85, where the averages were 0.48 for paved sites, 0.53 for grass sites, and 0.60 for forest sites. Shannon-Weaver index values for structural diversity of forest sites ranged from 2.33–3.35 and values for grass sites ranged from 0.22–1.41, although two forest sites were outside the boundary of our LiDAR data and could not be assessed.

3.2. Meteorological conditions

Daily average temperatures across sampling dates ranged from 13.1–23.7° C (Table 9 in Appendix C). Paved sites tended to have the highest average temperatures and greatest fluctuation between the daily minimum and maximum, while forested sites tended to have the lowest temperatures and least fluctuation. Dates earlier in the season were generally warmer than later sampling dates; September 15 was the coolest of all dates. Average wind speeds for the study area ranged from 2.1 m/s on August 25 to 5.0 m/s on September 22 (Table 10 in Appendix C). The prevailing wind direction for most dates was from the north, except on August 14, September 1, and September 29, when it blew primarily from the south. There was

no precipitation recorded during sampling, however, it did rain heavily on the two days prior to the September 1 sampling date.

3.3. Bacterial community diversity and composition

After quality filtering there were 12,830,718 total sequence reads, including singletons. Single- and doubletons comprised 1,385 reads, which were removed for all analyses except calculating species richness. After removing single- and doubletons, the number of reads per sample ranged from 26,180–215,283. The total number of sequence variants observed was 34,909, comprising 30 different phyla. Across the entire study, the most abundant phyla were Proteobacteria (46%), Actinobacteria (26%), Firmicutes (10%), Bacteroidetes (9%), and Planctomycetes (3%). Within the Proteobacteria, the Alphaproteobacteria class comprised 29%, Gammaproteobacteria 9%, Betaproteobacteria 7%, and Deltaproteobacteria 2%. The species richness computed by rarefaction analysis (subsampled at 26,180 reads) ranged from 448–2,106 expected SVs and was significantly associated with sampling date ($F = 14.6$, $p = 4.3e\text{-}15$) and site location ($F = 1.9$, $p = 0.0071$), but not with land cover ($F = 1.8$, $p = 0.17$).

Investigating beta diversity among samples, direct gradient analysis indicated that land cover type explained a small but significant portion of variation among bacterial communities when assessed across all days (Figure 11). The greatest variability occurs along the first axis (CAP1), which accounted for the large majority of the variance explained. This result was substantiated by a PERMANOVA testing for the effect of land cover type while controlling for sampling date, which showed that land cover type was a significant but weak predictor of similarity ($R^2 = 0.06$, $p < 0.001$) while sampling date had a stronger impact ($R^2 = 0.24$, $p < 0.001$). Post-hoc pairwise testing of land cover types indicated that the largest difference was between forest and paved sites ($R^2 = 0.06$, $p < 0.001$), followed by forest and grass sites ($R^2 = 0.04$, $p < 0.001$), and the smallest difference was between grass and paved sites ($R^2 = 0.03$, $p < 0.001$).

We visualized the effect of sampling date in a principal coordinates analysis (PCoA) plot of Morisita-Horn dissimilarity distance on the variance-stabilizing transformed counts (Figure

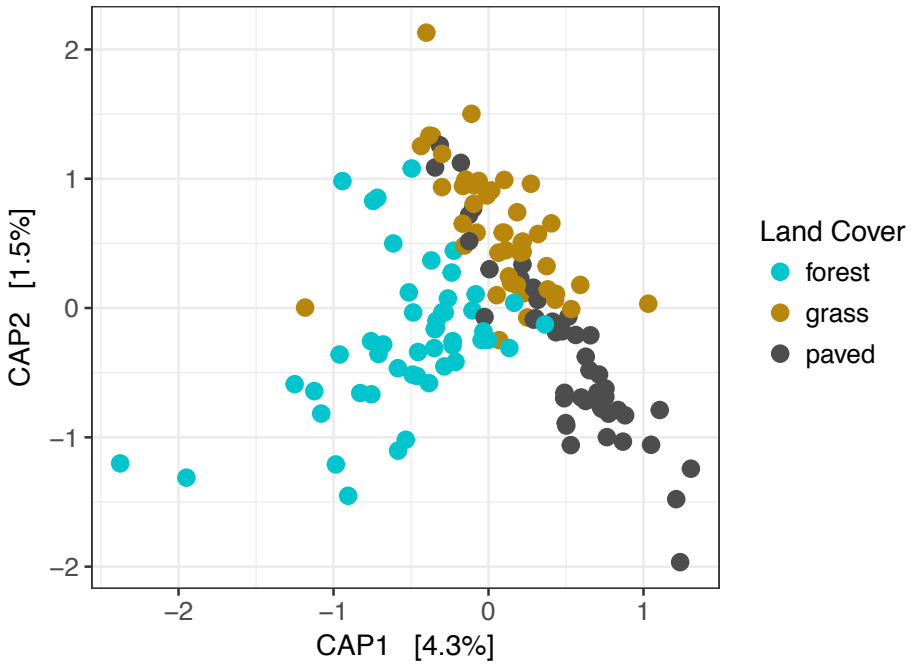


Figure 11. Constrained PCoA ordination plot for all samples, using Morisita-Horn dissimilarity distance on variance-stabilizing transformed counts, colored by land cover type.

12), where samples represented by points that are closer together have more similar composition than points further away. In particular, samples collected in July and early August form tight and distinct clusters, while those collected later are more dispersed. Within each cluster, paved sites tend to form a discrete subgroup from the vegetated sites, especially for the September sampling dates, implying that paved sites were more compositionally similar to each other than vegetated sites, particularly during periods when urban-scale variability was high.

Closer examination of bacterial compositional patterns revealed that the effect of sampling date was largely due to changing relative abundances of the dominant bacterial taxa over the course of the sampling campaign (Figure 13). On the first two sampling dates, for example, the communities were dominated by the family Sphingomonadaceae, primarily an SV identified as *Sphingomonas faeni*, which had equally high relative abundance across all sites. Later in the campaign, Acetobacteraceae, Enterobacteriaceae, and Oxalobacteraceae comprised larger relative components of the community, but tended to be unequally distributed across sites. For example, on one sampling date (September 1) we observed a large increase in the relative

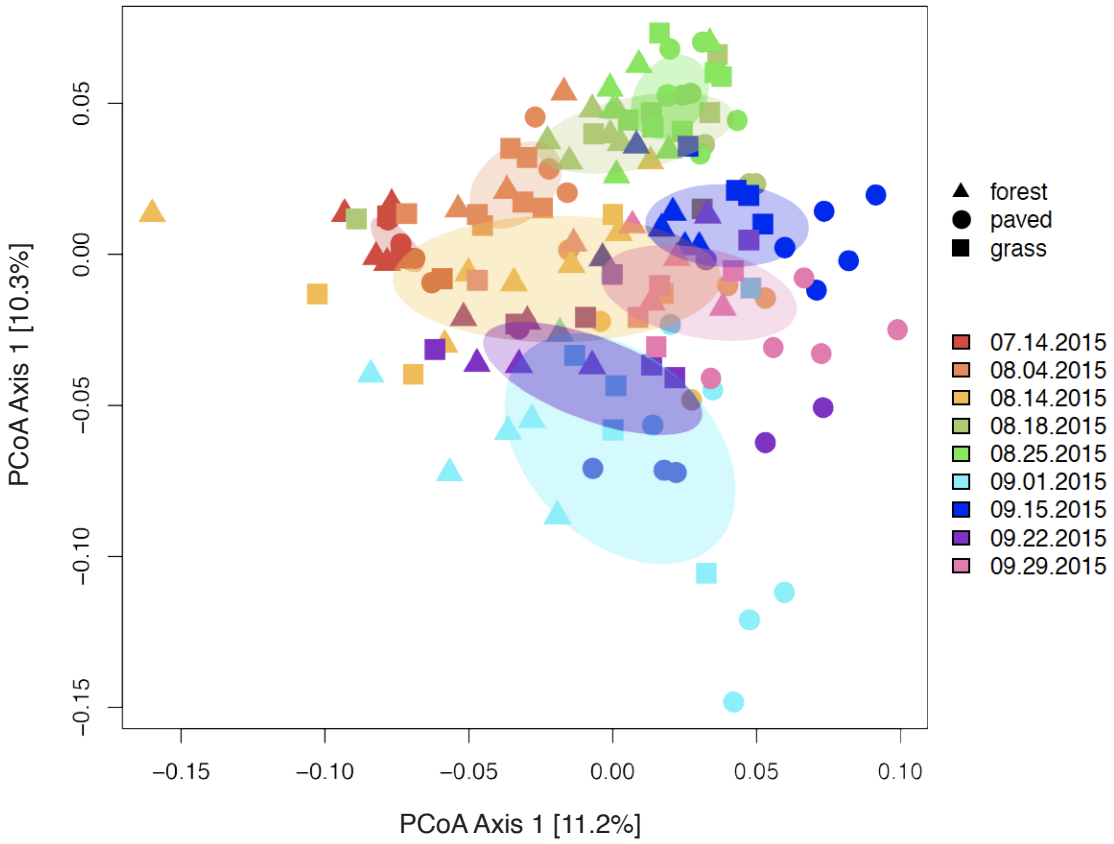


Figure 12. PCoA ordination plot for all samples, using Morisita-Horn dissimilarity distance on variance-stabilizing transformed counts, colored by sampling date. Ellipses represent 1 standard deviation from the mean.

abundance of *Acidiphilium multivorum* (family Acetobacteraceae), at two paved sites (SAF and BIM) where its relative abundance rose to over 50% of all sequences recovered from those sites (Figure 28 in Appendix C). Another notable feature was the high abundance of Enterobacteriaceae, especially *Erwinia billingiae*, which was found at highly maintained grassy sites (e.g., irrigated and mowed—cemetery CEM, park MAU, golf course LAU) across several dates.

When we performed individual PERMANOVA tests for the remaining variables of interest, site location was the dominant factor explaining compositional variation, and, in fact, was a stronger predictor than sampling date, while other vegetation and meteorological variables had significant but marginal associations (Table 1). Each of these factors was tested separately, since they were confounded with other factors (e.g., site location with average temperature). We also

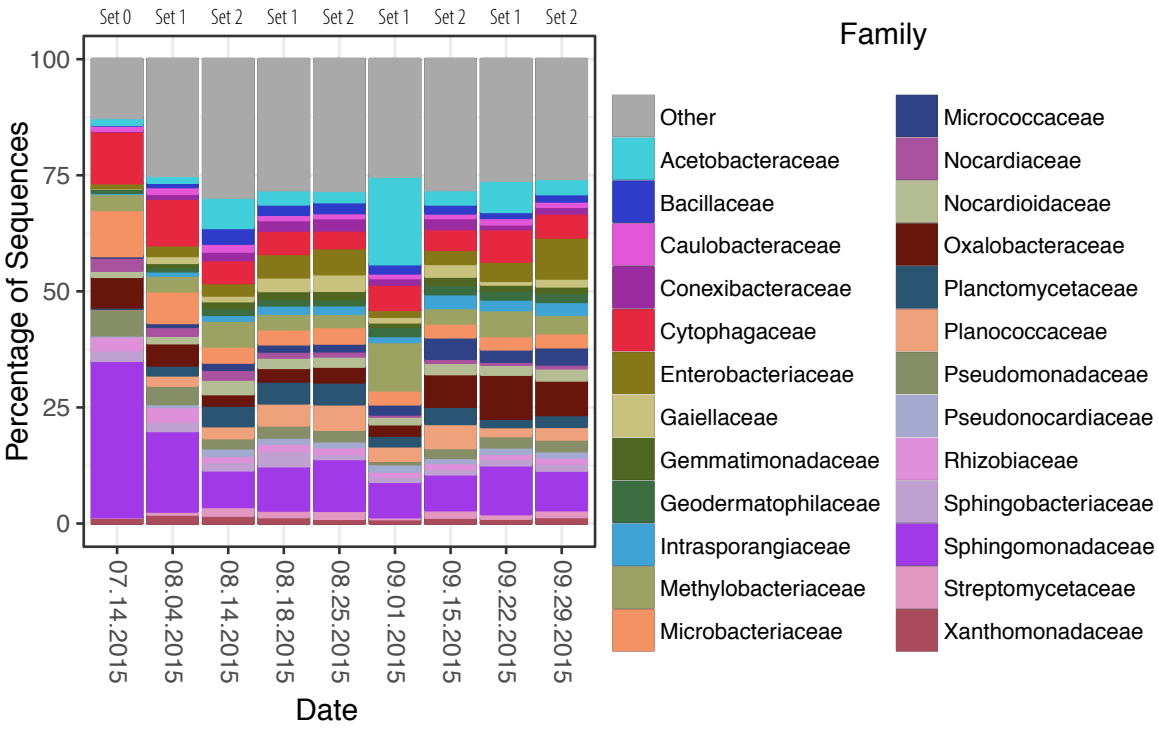


Figure 13. Community composition at the family level for all sites split by sampling date.

included sampling date and land cover type in these individual analyses, and found that the results were virtually identical to testing them in multivariate PERMANOVA.

We next analyzed samples separately by sampling date to see whether the strong effects of sampling date and site location were obscuring other important relationships. The results of this PERMANOVA test showed that the explanatory power of land cover type increased

Table 1. Results from individual PERMANOVA analyses on selected environmental factors, using Morisita-Horn dissimilarity distance on variance-stabilizing transformed counts. *P*-values with an asterisk remain significant after Bonferroni correction.

Factor	**R^2**	***p***
Site location	0.33	<0.001*
Sampling date	0.24	<0.001*
Land cover type	0.06	<0.001*
Dominant wind direction (entire study area)	0.05	<0.001*
Structural diversity (only vegetated sites)	0.04	<0.001*
Average temperature	0.04	<0.001*
Vegetation cover 50 m	0.03	<0.001*
Vegetation cover 800 m	0.02	<0.001*

dramatically once the obscuring effects of sampling date and site location were removed; six out of the nine dates were highly significant, while the remaining three dates fell just outside our chosen level of significance (Table 2 and Figure 29 in Appendix C). We also tested whether community similarity was associated with spatial proximity for each sampling date. Despite site location being a strong predictor when all samples were analyzed together, the Mantel test for spatial correlation showed little to no effect for most sampling dates (Table 2), although two dates (August 18 and 25) showed strong and significant effects.

Table 2. Results of PERMANOVA analyses on land cover type and Mantel spatial correlation test for each sampling date, using Morisita-Horn dissimilarity distance on variance-stabilizing transformed counts.

	Land cover type (PERMANOVA)		**Spatial correlation (Mantel)**	
Date	**R^2**	***p***	**statistic**	significance
07.14.2015	0.31	0.008*	-0.01	0.49
08.04.2015	0.18	0.053	0.20	0.09
08.14.2015	0.18	0.041	0.09	0.24
08.18.2015	0.18	0.000*	0.39	0.02*
08.25.2015	0.13	0.082	0.25	0.04*
09.01.2015	0.21	0.000*	-0.16	0.88
09.15.2015	0.18	0.000*	0.12	0.19
09.22.2015	0.21	0.004*	0.02	0.40
09.29.2015	0.16	0.001*	-0.08	0.70

Finally, we examined whether individual taxonomic groups contributed to the observed compositional variation across land cover types, using a negative binomial GLM to identify differentially abundant taxa. We first tested at the class level and found that the relative abundance of 29 out of the 53 bacterial classes varied significantly across forest, grass, and paved sites (Figures 14a–b), with 17 of these classes present in very low abundances (<1%). Only three classes, all abundant, differed among all three cover types. Nine classes differed for paved sites compared to the other two cover types, whereas four differed for grass and two

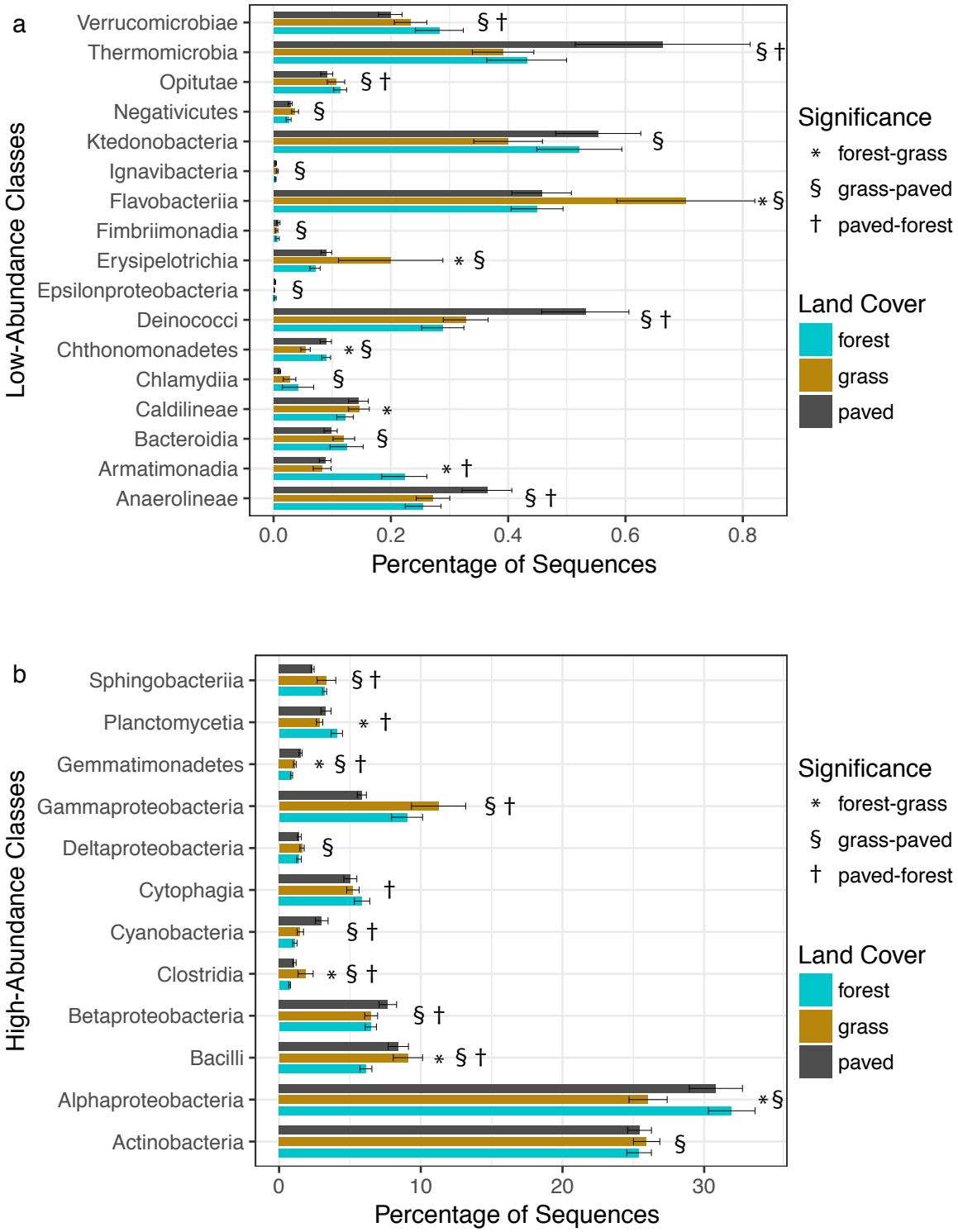


Figure 14. a) Low-abundance (< 1%) and b) high-abundance (> 1%) bacterial classes that were differentially abundant across land cover types, using a negative binomial GLM. Bars show mean relative abundance for each land cover type, error bars indicate standard error of the mean.

for forest. Nine classes differed between grass and paved sites, and one each for paved-forest and forest-grass. Gammaproteobacteria, a high-abundance class, was enriched at vegetated sites (both grass and forest), while Deinococci, Thermomicrobia, and Anaerolineae, all low-abundance taxa, were enriched at paved sites. We further tested whether generic diversity of the class Gammaproteobacteria was associated with land cover type because it has been suggested as a potential pathway for health benefits of greenspace (Hanski et al. 2012; Ruokolainen et al. 2014). The results showed that forest sites had higher generic diversity of Gammaproteobacteria than grass (ANOVA; t = -2.29, $p = 0.024$) or paved sites (ANOVA; t = -2.99, $p = 0.003$).

Because comparisons at higher taxonomic levels can mask important distinctions at lower taxonomic levels, we performed the same test for differential abundance at the family and genus levels. We chose these levels in part to facilitate comparison with results from other studies in the literature and because in many cases taxa were not resolved to the species level, since individual SVs frequently could not be confidently matched to single species in the NCBI databases. At these levels of taxonomic resolution, we noted that within the family Acetobacteraceae (class Alphaproteobacteria) one genus (*Granulibacter*) was highly indicative of forests, while other genera in the same family (*Gluconobacter*, *Rubritepida*, *Acidicaldus*) were indicative of paved areas (Figure 15; family level results reported in Figures 30–32, Appendix C). Likewise, several genera within the family Enterobacteriaceae (class Gammaproteobacteria) had contrasting abundance patterns. For example, when comparing forest with paved sites, *Arsenophonus* and *Erwinia* were enriched in forest sites and *Citrobacter* was enriched in paved sites, but *Citrobacter* was enriched in grass sites when compared with either forests or paved sites (Figures 15–18). Interestingly, the two top differentially abundant genera in paved sites compared with grass sites, *Frischella* and *Gilliamella* (Figures 17–18), are both members of family Orbaceae and exhibit resistance to the antibiotic oxytetracycline (Engel et al. 2013).

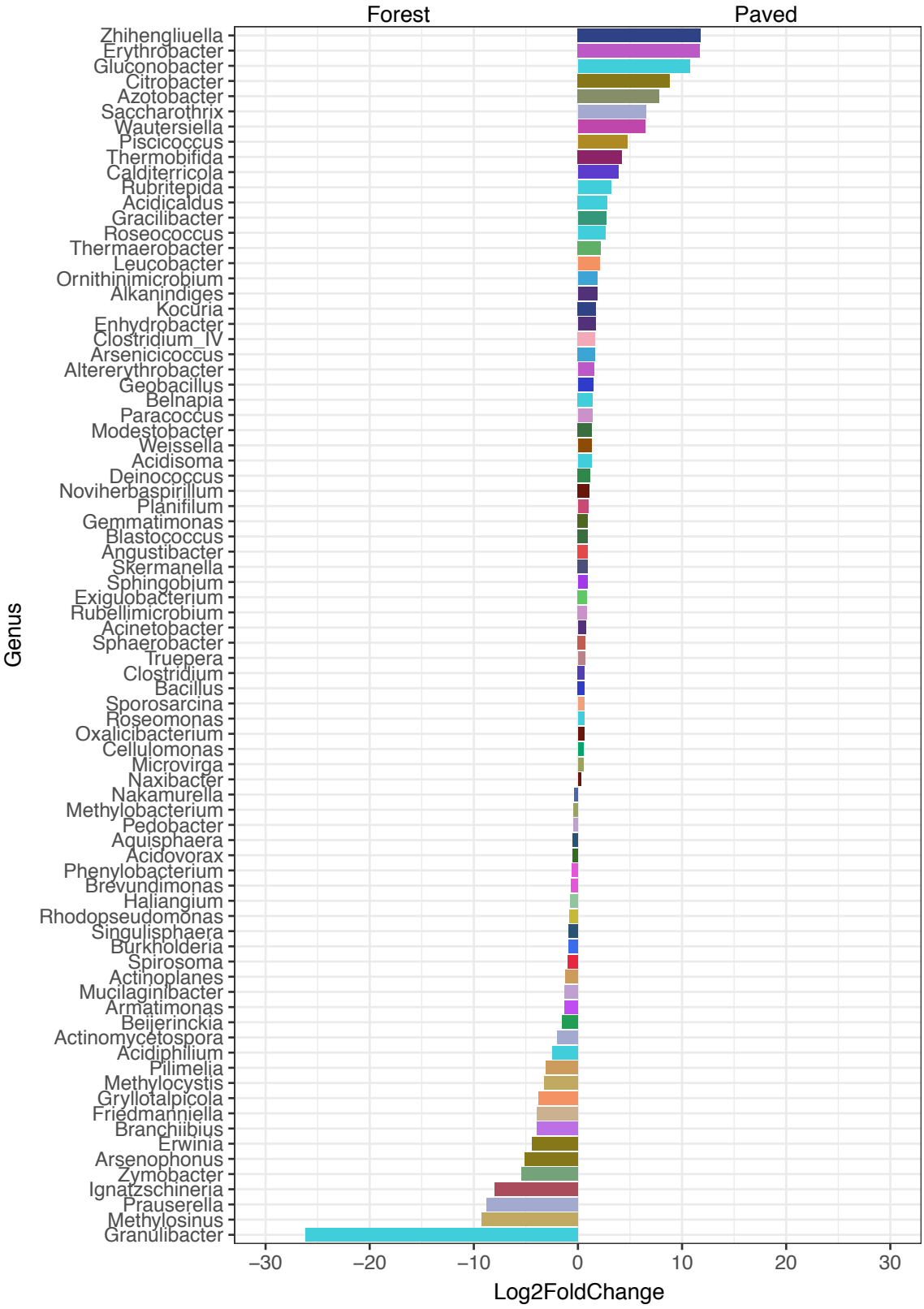


Figure 15. Differentially abundant genera in forest vs. paved sites.

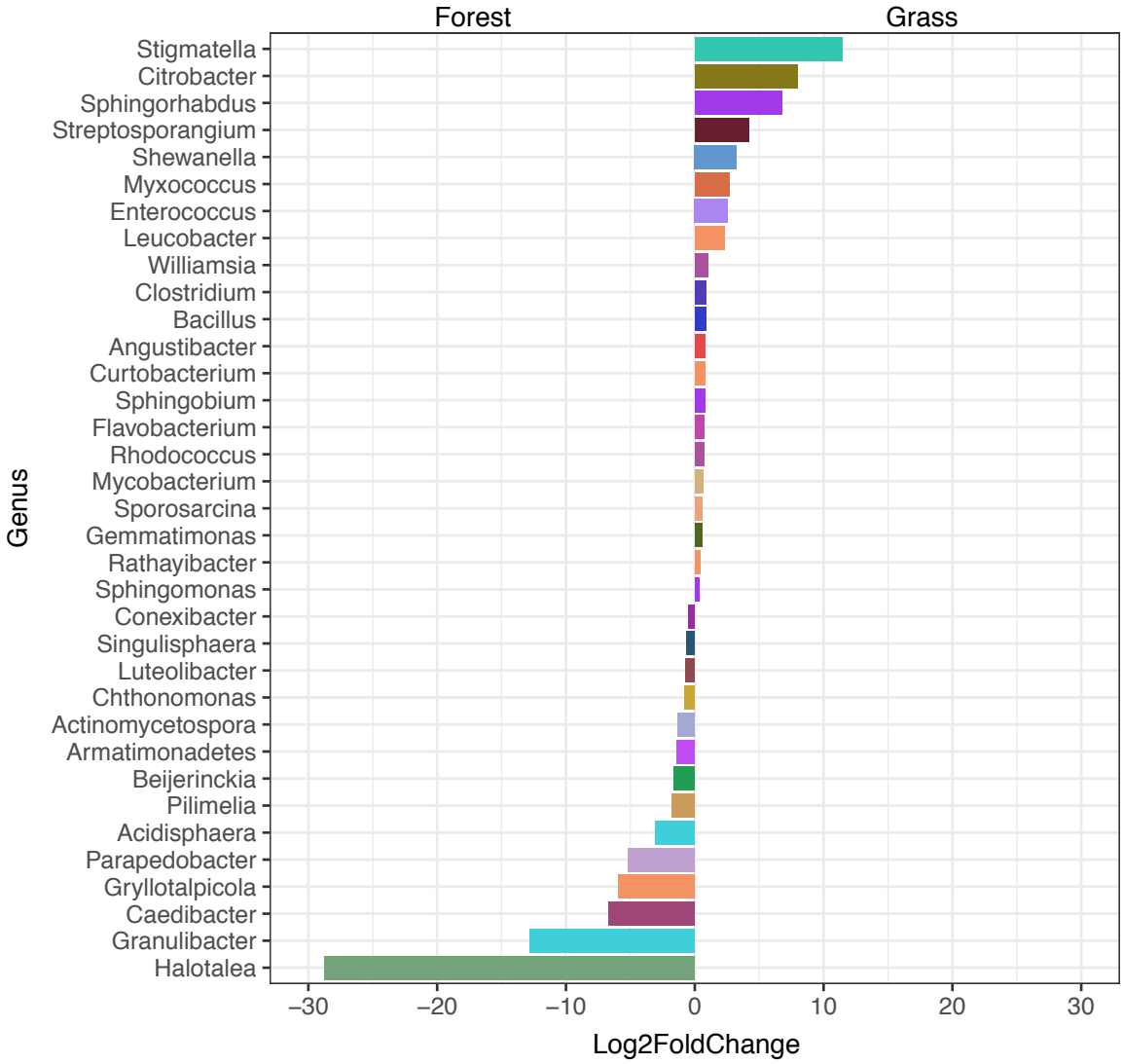


Figure 16. Differentially abundant genera in forest vs. grass sites.

4. Discussion

Urban land cover types exhibit differences in airborne bacterial community structure

We found that, while controlling for time and across all samples, land cover type had a significant but subtle association with community similarity, but not with species richness. An ordination of bacterial communities among across all days based on land cover type accounted for ~6% of the total variation in community composition (Figure 11). While this is a small amount of the total variation in communities it is notable in light of the large daily shifts in the

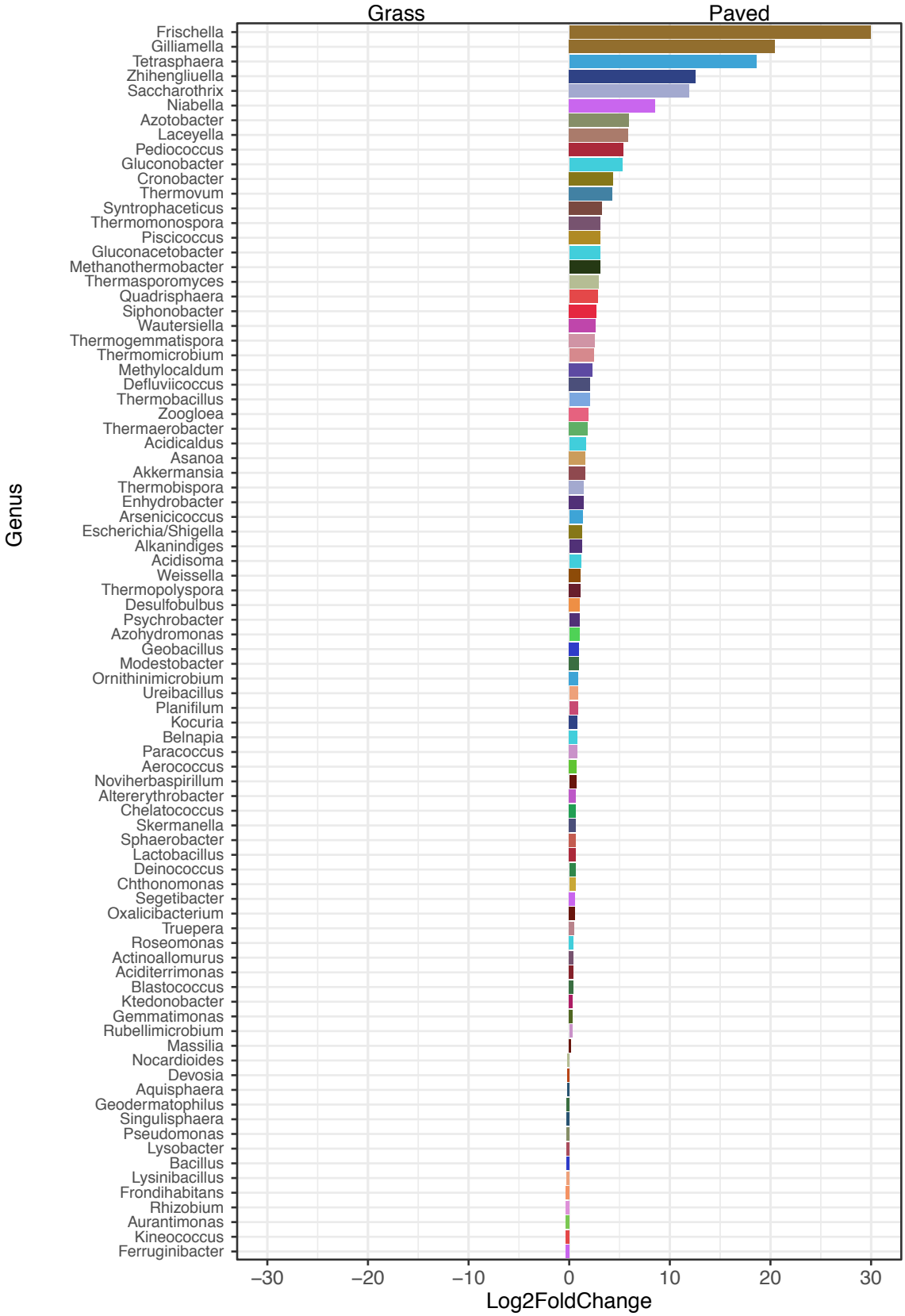


Figure 17. Differentially abundant genera in grass vs. paved sites.

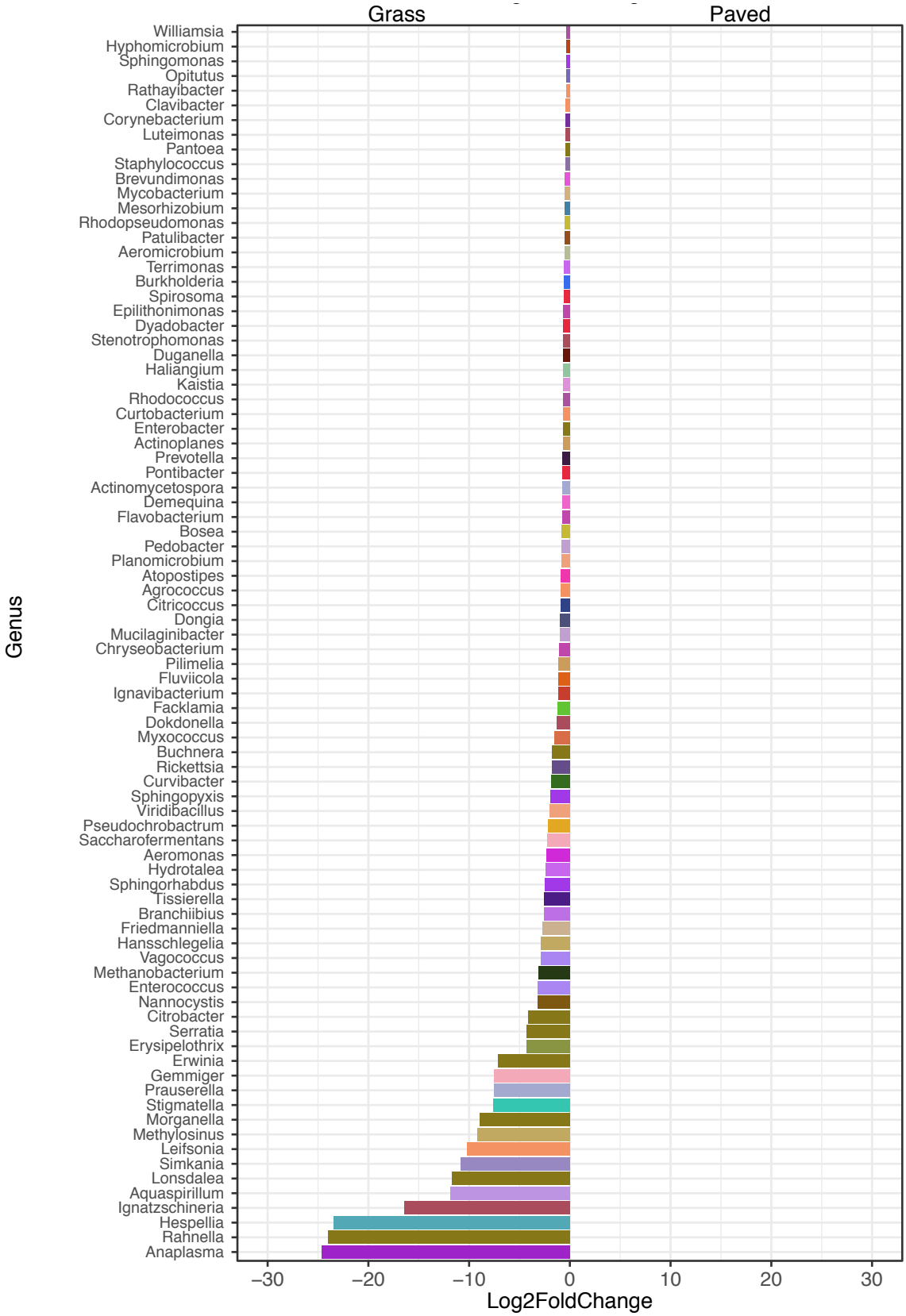


Figure 18. Differentially abundant genera in grass vs. paved sites (cont.).

aerobiome (represented by all sites on a given day in Figure 12), which were not accounted for in this ordination. Forests were the most distinctive and the most variable, exhibiting the largest dispersion in ordination space, and in particular the greatest variability along Axis 1, which accounted for the large majority of the variance explained. The tendency for each forested site to be relatively unique may be a function of a forest's dual roles as microbial source and mediator of air movement. As a microbial source, they have large amounts of vegetative surface area, which provides a substrate for plant-associated bacteria, and relatively high plant species diversity due to the presence of three structural layers (tree layer, shrub layer, ground layer) under low human management. In terms of the air movement, their greater height and structural diversity from the ground to the top of the canopy restricts air flow, which helps retain locally-emitted bacteria and reduces mixing with the regional air mass. Although we observed the hypothesized effect of land cover type on community composition, the association with sampling date was stronger by a factor of four. When samples were analyzed separately for each date, however, the association with land cover type increased from $R^2 = 0.06$ to an average of $R^2 = 0.19$ (Table 2), suggesting that it has an important effect on bacterial community variation that was obscured by broad compositional changes associated with time.

We conjecture that land cover type may serve to some degree as a "master controller" of the local microbial community through its influence on not only microbial sources (via total leaf surface area and plant species composition or the presence of non-vegetated surfaces), but also on microbial survival and growth (by modifying microclimatic factors such as humidity and temperature) and on the retention or dispersal of microbes (by its influence on airflow). Consistent with this idea, we note that, with one exception, bacterial communities were only significantly associated with vegetation cover and structural diversity on days when they were also associated with land cover type (Table 12 in Appendix C). In other words, when the effect of land cover type was significant, two of its more nuanced components were sometimes significant.

The influence of cover type also changed over time. Microbial communities were only

consistently associated with land cover type for the last four sampling dates of the campaign, starting on September 1 following the first major precipitation of the rainy season. Furthermore, temperature was either significant (three dates) or marginally significant (three dates) for the first six sampling dates of the campaign when temperatures were warm; for the last three dates, after temperatures dropped substantially, *p*-values for temperature rose sharply. This suggests a potential seasonal shift of controls from: a) site-scale variation during the drought and heat prior to the rainy season along with a strong regional land use signal when winds were predominantly from the north, to b) land cover type and other site-scale vegetation characteristics during the cooler, moister conditions after the rainy season began, herbaceous vegetation began to green up, and winds shifted to come from the forested hills to the south.

Environmental metrics that best captured relevant site features

When all samples across all days were analyzed together, the vegetation and meteorological variables tested, with the exception of site location, performed poorly, each explaining 5% or less of the variation among bacterial communities. Individual site location was the strongest explanatory variable in describing community similarity, even stronger than sampling date. This contrasts with other culture-independent aerobiome studies (e.g., Brodie et al. 2007; Fierer et al. 2008; Bowers et al. 2013) that found spatial location within an urban area to be insignificant compared to the effect of time. We conjecture that, in our study, site location was a proxy variable capturing a suite of features ranging in spatial scale from the urban block to the neighborhood that could affect bacterial sources, dispersal capabilities, and survival rates. These features might include vegetation characteristics from the local land cover type to the specific plant species found, microclimate conditions, and landscape management regimes (Figure 7). Surprisingly, proximity to other sites did not appear to affect compositional similarity for most sampling dates (Table 2), perhaps indicating that the larger urban air mass was well-mixed and distributed across the study area and that unique microbial taxa originating from individual sites became undetectable short distances away due to rapid dilution with distance. This idea is

supported by the diminishing importance of the proportion of vegetation cover with increasing buffer size around the sampling point (Figure 33, Appendix C).

Taxonomic groups contributing to variation in community composition across land cover types

We assessed differential abundance at several taxonomic levels—class, family, and genus—to facilitate comparisons and uncover patterns at different levels. Overall, at the genus level the grass-paved contrast showed more than twice as many differences as forest-paved, and the forest-grass contrast showed half again as many. Paved sites thus stand out as having the most unique bacterial composition at the genus level, and forest and grass the most similar. At the same time grass sites account for the 60% of "enriched" indicators in their contrasts with forest and paved sites, suggesting they may be the dominant sources for indicator taxa. The picture shifted, however when indicators were examined at the family level. At this level, the forest-paved contrast had the largest number of indicators and grass-paved the least, while grasslands still showed the greatest number of indicators for both its contrasts. At both the genera and family levels, paved sites were most enriched with indicators in the forest-paved contrast, suggesting that forests serve as sources for the smallest number of indicators. We note that these differential abundance tests were reported for all samples together, rather than separately by date, therefore the strong effect of sampling date may have concealed taxa that were differentially abundant only on certain dates.

It is also possible that forest sites as a whole show fewer indicator taxa, because individual forest sites are characterized by different sets of indicator taxa. In support of this, we note that: a) different vegetation species host distinct bacterial communities; b) trees are large sources of bacteria, due to their large surface area; and c) forests alter microclimate (e.g., air flow, temperature, radiation) such that more bacteria from local sources may survive and be retained than at other land cover types. If this is the case, then testing for differential abundance by aggregating all forest sites may have resulted in fewer indicator taxa than grassland or paved sites. This possibility is consistent with our finding that individual site location is a stronger

predictor of bacterial community similarity than land cover type when all samples are analyzed together. Similarly, paved sites may exhibit greater numbers of indicator taxa due to the unique biophysical environment, which is similar across all parking lots (e.g., asphalt substrate, high radiation, presence of heavy metals and petroleum-based fluids, little to no vegetation). In an earlier study, Mhuireach et al. (2016) found that bacterial communities collected from parking lots tended to be similar to each other, while those from grassy parks were different both from other parks and from parking lots.

Indicator taxa for paved sites generally had characteristics that suggested they might originate from the primary surface type found there—asphalt pavement. Specific bacterial classes that were enriched at paved sites included Deinococci, Thermomicrobia, Anaerolineae, Betaproteobacteria, and Cyanobacteria. Many of these taxa, especially Deinococci, are known for their ability to survive very hostile conditions, such as intense radiation, desiccation, and heavy metal contamination, all of which are likely to characterize paved parking lots. At the species level, *Acidiphilium multivorum* (family Acetobacteraceae) was primarily found in paved locations across several different dates and was particularly abundant on September 1 after two days of heavy rain, the first major precipitation after several months of summer drought. *A. multivorum* has been previously isolated from pyritic acid mine drainage; it utilizes organic compounds, such as methanol, for growth (Wakao et al. 1994) and is highly tolerant to heavy metals (Mahapatra and Banerjee 1996). We speculate that this taxa may thrive on leaked automobile fluids present on paved surfaces, especially following rains.

On the other hand, samples obtained from vegetated sites tended to have more taxa associated with soil and leaf surfaces. At the family level, Halomonadaceae, Armatimonadaceae, Beijerinckiaceae, Pseudonocardiaceae, and Planctomycetaceae were observed in higher abundance at forest sites, while Sphingobacteriaceae and Methylocystaceae, were abundant in vegetated environments generally (Figures 30–32, Appendix C). Other researchers have shown these families to be common phyllosphere inhabitants (Lymperopoulou et al. 2016; Laforest-Lapointe et al. 2016; Smets et al. 2016). Our study thus provides additional evidence that

nearby plants are measurable sources of bacteria to the air.

Vegetation management was also implicated as a factor influencing bacterial community composition. At grassland sites, we observed greater abundance of class Flavobacteriia and families Enterobacteriaceae, Enterococcaceae, Xanthomonadaceae, Flavobacteriaceae, Nocardiaceae, and Mycobacteriaceae (Figures 30 and 32, Appendix C). The high relative abundance of Enterobacteriaceae may have been primarily driven by a particular species, *Erwinia billingiae*, at grass sites with human management regimes—a cemetery, a golf course, and a public park. We speculate that this bacterium may have been released in high quantities at these sites during mowing or irrigation events that coincided with our sampling dates.

We also specifically investigated the generic diversity of class Gammaproteobacteria, since high generic diversity of Gammaproteobacteria on the skin of healthy individuals has been associated with a protective effect against atopy (Hanski et al. 2012). Results showed that abundance was enriched in both forest and grass sites compared with paved sites (Figure 14b), but had greater generic diversity at forest sites, which aligns with previous work (e.g., Bowers et al. 2010; Hanski et al. 2012) and may have important health implications.

Temporal patterns of bacterial community change

Even the brief three-month duration of our study revealed strong temporal patterns and fluctuations from one sampling date to the next, which tended to mask relationships between bacterial community structure and fine-scale vegetation metrics, while suggesting key interactions across larger spatial and temporal scales. As indicated by our conceptual model (Figure 6), the time scale of our study encompassed several processes and events occurring within the range of days to months that could be expected to impact the aerobiome. In particular, seasonal biotic processes (e.g., vegetation senescence), changes in weather and dominant airflow patterns, and regional-scale human activities (e.g., agriculture) appeared to play roles in the observed shifts in dominant taxonomic groups, as has been noted in other studies (Gandolfi et al. 2015; Lee et al. 2017). For instance, we found that July samples for all

sites were dominated by *Sphingomonas faeni*. This taxa was distributed relatively equally across all sites, suggesting a regional rather than local source, and generally decreased over the course of the sampling campaign. *S. faeni* has been previously isolated from air in agricultural barns where bales of hay were being broken open (Andersson et al. 1999). Because our study area was located downwind of a large grass seed production region and July is a prime harvesting month, we posit that *S. faeni* originated from seasonal agricultural activities, which corroborates previous work (Lighthart 1984; Mhuireach 2016). Other taxonomic groups also altered abundance over the course of our campaign, for example, Cytophagaceae and Microbacteriaceae generally decreased, while Acetobacteraceae, Enterobacteriaceae, and Oxalobacteraceae increased (Figure 13). These changes may reflect other seasonal changes associated with late summer, such as lowering temperatures and precipitation events.

The study had several limitations whose resolution could help strengthen future studies of the urban aerobiome. First, finding sampling sites that met our a priori criteria of <25% vegetation within 50 meters was challenging. Eugene-Springfield is a small metropolitan area with an abundance of vegetation and therefore may not have provided as wide a range of variation in cover type as other cities. In addition, many of the paved sites had substantial landscaping, in part due to city codes requiring 10% or more of the total area in parking lots to be landscaped with vegetation. The resulting bioswales and planter islands may have contributed plant- and soil-associated bacteria to the nearby air. We also stress the importance of using wind dataloggers at individual sites in future studies to help disentangle the dual roles of vegetation as microbial source and mediator of air movement. We expect that air movement at the local scale of individual sites may be influential on airborne microbial community dynamics and that nearby vegetation, especially structurally diverse forests, alters patterns of air movement. Therefore this would be a fruitful avenue for future research.

5. Conclusions

Understanding key factors that shape urban aerobiome assembly may be vital to designing

healthier neighborhoods and cities. This study contributes to that goal by indicating that human decisions about the configuration and management of the urban landscape, as well as the surrounding landscape may influence the composition of urban airborne bacterial communities at varying scales of space and time. Interplay among different spatial and temporal drivers of bacterial community assembly can begun to be understood by assessing features that range from fine scale (e.g., land cover type, vegetation species composition) to coarse scale (e.g., agricultural activities, seasonal vegetation changes), including periodic events (e.g., precipitation, irrigation, and mowing). Our work suggests that different land cover types may have potential to shape the composition of the airborne bacterial communities to which people are exposed on a daily basis. The specific mechanisms through which land cover type influences bacterial communities may include fine-scale features, such as irrigation and mowing regimes, presence of asphalt surfaces and automobile use, and variation in vegetation species composition. Importantly, we have begun to lay a foundation for understanding how spatiotemporal factors interact across scales to shape the urban aerobiome, which, ultimately, may give rise to tools that help designers and planners manage it.

To secure the potential health benefits we must, however, establish practical design guidelines, such as how much and what type of vegetation and at what distance from residents' homes, and we must supply concrete evidence linking variation in aerobiome composition with human health. Although our work revealed that certain bacterial groups varied in abundance and diversity across land cover types, further research is needed to confirm whether exposure to these airborne bacteria can impact health. Improved knowledge of the spatial and temporal dynamics of the urban aerobiome could benefit long-term green infrastructure plans for healthier and more equitable cities.

Bridge to Chapter IV

In Chapter III, I demonstrated how urban airborne bacterial communities can be influenced by land cover, vegetation, and management activities—factors that can be controlled by

landowners and policies. I also explored the ways in which these and other factors may interact across different scales of space and time, suggesting that interventions ranging in scale from site design to regional planning could be relevant to shaping the aerobiome. Specifically, I confirmed that land cover type (forest, grass, paved) had a significant effect on the composition of urban airborne bacterial communities, although the effect was overwhelmed by the effects of sampling date and individual site location. I also found that forests, in particular, may have higher abundance and generic diversity of the bacterial class Gammaproteobacteria, which has been associated with reduced prevalence of atopic sensitization. The goal of the final chapter is to connect scientific findings from Chapters II and III, as well as other published research, with specific ideas that could be useful to designers and planners. The motivation behind this chapter is to facilitate the application of evidence-based design to test emerging evidence that microbial exposures in the urban landscape may benefit human health, through. I begin by briefly reviewing how vegetation is related to human health, how vegetation influences airborne microbial communities, and how microbial exposures may impact health. Then, I construct a conceptual framework to translate the scientific understanding of microbial ecology into language and processes that are more familiar to designers. Finally, I propose example design hypotheses and interventions at different spatial scales, as a foundation for future research.

CHAPTER IV

VARIATION IN THE URBAN MICROBIOME ACROSS SCALES: AN INVISIBLE COMPONENT OF LANDSCAPE DESIGN FOR HEALTH

The visible is set in the invisible; and in the end what is unseen decides what happens in the seen; the tangible rests precariously upon the untouched and ungrasped. The contrast and the potential maladjustment of the immediate, the conspicuous and focal phase of things, with those indirect and hidden factors which determine the origin and career of what is present, are indestructible features of any and every experience.

— Dewey 1958, 43-44

Introduction

We are in the midst of a health paradigm shift due to emerging recognition of the importance and complexity of human-microbe relationships. For the majority of the 20th century it has been generally accepted that exposure to microbes, such as bacteria, fungi, and viruses, equates to significant risk of contracting an infectious disease. However, we now understand that not only do we host a vast assemblage of microbiota in and on our bodies, but we also encounter innumerable environmental microbes, including pathogens, in our day-to-day life. These exposures can have radically different outcomes, from health benefits to disease contraction (Rook et al. 2017; Gilbert and Tauber 2016).

The idea that some microbial exposures may provide health benefits is generally recognized to have begun with the formulation of the Hygiene Hypothesis, which posited that early life infections in large households might have a protective effect against allergic rhinitis, or "hay fever" (Strachan 1989). Graham Rook (2008) later proposed that the unavoidable exposure throughout most of our evolutionary history to a group of organisms dubbed 'Old Friends,' which include commensal microorganisms and helminthic parasites, led to remodeling of the immune system to tolerate, and even require, that exposure. Similarly, the Biodiversity Hypothesis posited that the microbiodiversity (diversity of microbial life) of a given environment is dependent to a large extent on its macrobiodiversity (diversity of plants and animals) and, furthermore, that human health is interrelated with both (von Hertzen

et al. 2011). Building upon these ideas, the more recent Microbiome Rewilding Hypothesis proposed that restoring urban biodiversity may provide immune system training as an ecosystem service (Mills et al. 2017). Together, these theories begin to establish a compelling rationale for recognizing that the environments in which we spend our daily lives can have major health implications and, perhaps, that incorporating evolutionary theory in urban landscape design could have substantial health benefits (Milne 2015). In this paper, we argue that vegetation represents a potential linkage between health, environment, and microbes, since the health benefits of vegetation are increasingly acknowledged, although the mechanisms have, to a large degree, remained poorly understood.

Goals and organization of paper

The overarching goal of this paper is to lay a foundation for evidence-based urban landscape design that recognizes microbial exposures as a potential pathway for promoting health. To achieve this goal, we first take a brief look at historical lessons in designing the physical fabric of cities to improve health and reduce exposure to pathogenic microbes at the population scale. Second, we discuss the current state of evidence linking vegetation with health and the potential pathways mediating that linkage, including a newly emerging 'immune enhancement' pathway. Third, we review evidence that environmental microbial exposure can impact human health and may represent a specific mechanism through which immune enhancement occurs. Fourth, we describe factors that determine how urban microbial communities assemble, limiting the discussion to airborne bacterial communities. Lastly, we construct a framework to translate current understanding of drivers and constraints on microbial ecological dynamics into concepts useful for evidence-based landscape design for health.

Historical lessons in designing cities for health

For centuries, good urban design was perceived as an important way to promote human health. Hippocrates (circa 400 BC) may have been among the first to document how

environmental conditions, such as air and water quality, could impact people's well-being. He suggested that to understand health conditions it was important to consider the surrounding landscape, ". . . whether it be naked and deficient in water, or wooded and well watered, and whether it lies in a hollow, confined situation, or is elevated and cold," (Hippocrates n.d.). Several centuries later, these ideas were recapitulated by Roman civil engineer and architect Vitruvius, who suggested that, "a mild, thick air . . . strengthens and restores," and the design of urban streets should be ". . . laid down on the lines of division between the quarters of two winds," (Vitruvius n.d.). In this era of history, people believed in miasma theory, or the idea that illness was caused by noxious odors emitted by decaying organic matter. Later, during periods of industrialization in Europe and the US, interest in environmental quality rose to the forefront of public health efforts. Early industrializing cities were population sinks with shorter life expectancies and poorer overall health than rural areas, largely due to crowding and inadequate infrastructure (e.g., waste removal, clean water supply) to support the massive influx of people (Haines 2001; Kearns 1988; Reher 2001). As a result of high human mortality rates, several new approaches to combat disease gained force throughout the 19th century, including public sanitation, hygiene reform, and the parks movement (Porter 1999; Eisenman 2013). Historians largely agree that these public health efforts aimed at creating healthier living environments, in addition to rising standards of living, were responsible for reducing mortality rates, whereas advances in medical technology contributed little (Harris 2004; Haines 2001; Kearns 1988; Szreter 1988).

The parks movement is particularly germane, as public urban parks were proposed as a way to clean out "bad air" and provide places for mental respite, thus improving the well-being of residents (Ward Thompson 2011). In 1879, landscape architect and prior secretary of the US Sanitary Commission Frederick Law Olmsted designed Boston's Back Bay Fens, whose "leading and only justifying purpose" was "the abatement of a complicated nuisance, threatening soon to be a deadly peril to the whole city as a propagating and breeding-ground of pestilential epidemics" (Boston Dept. of Parks 1881, 27). According to Martensen (2009), Olmsted was

one of the two most influential figures during this era in terms of healthy urban design, the other was John Rauch, a Chicago medical physician whose report, "Public Parks: Their Effect upon the Moral, Physical and Sanitary Conditions of the Inhabitants of Large cities; with special reference to the City of Chicago," facilitated the addition of about 2,500 acres of parkland and one million trees in Chicago.

Despite the tight bond between living environment and health, the ascendancy of germ theory near the turn of the 20th century eventually led to a disconnect between the urban planning and public health disciplines. Poor health came to be seen as a problem that could be solved through medical treatment with vaccines and antibiotics, while the planning profession embraced zoning as a way of reducing the ill effects of industrial pollutants (Corburn 2004). The education and treatment of individuals based on bacteriology was considered to be a cheaper and more "efficient" method for attaining better public health than improving environmental conditions for the entire population (Fairchild et al. 2010). A key turning point was the publication of *The New Public Health* by Hibbert Hill (1913, 10) stating:

> The old public health was concerned with the environment, the new is concerned with the individual. The old sought the sources of infectious disease in the surroundings of man; the new finds them in man himself. The old public health sought these sources in the air, in the water, in the earth, in the climate and topography of localities, in the temperature of soils at four and six feet deep, in the rise and fall of ground-waters ; it failed because it sought them, very painstakingly and exhaustively, it is true, in every place and in every thing where they were not.

Notwithstanding advances in medical knowledge, chronic and autoimmune disorders have continued to increase in step with modernization and urbanization (Sorci et al. 2016), although the causes are unclear. Since it is now apparent that treatment of the individual does not ensure good health, medical and health professionals are again looking towards environmental solutions (Afifi and Breslow 1994). Likewise, planning professionals are reconsidering the effects of urban zoning ordinances, which, although originally intended to protect urban residents from toxic pollutants emitted by factories, have decreased walkability and forced

people into automobiles, simultaneously reducing physical activity and increasing pollution. As the separation of health and planning has resulted in unexpected effects, researchers and practitioners from both disciplines have come full circle to recognize once again that external environmental conditions remain a crucial aspect of health promotion. They issue joint calls to reunite public health with urban design and planning by creating more walkable neighborhoods and more greenspace (Frumkin 2003).

A new "settings-based" approach to health strives to reinvigorate the historic emphasis on how environmental conditions impact health and lifestyle choices (Dooris 2009). Its emergence reveals a fundamental shift in the conception of health, indicating a new emphasis on salutogenic, or health-promoting, environments and recognizing the complex interactions of personal, social and biophysical factors that together determine individual health. This approach recalls famed microbiologist René Dubos, who stated, "In most situations, design could certainly be improved by a better knowledge of man's nature and of the effects that the environment exerts on his physical and mental being," (Dubos 1968, 131). To a large degree however, the link between health and environment is still a "black box of places" (Macintyre et al. 2002), where associations between health and environment are known but the mechanisms underlying those associations are not.

The vegetation-health link

Abundant research demonstrates that vegetation and natural environments are key features of landscape design for health (e.g., Fong et al. 2018; Nieuwenhuijsen et al. 2017; Ward-Thompson and de Oliviera 2016; Kabisch et al. 2015; Hartig et al. 2014; Lee and Maheswaran 2010; Tzoulas et al. 2007). In urban areas, residents who have more vegetation near their homes or visit natural areas tend to experience less stress and depression (Kuo 2001; Grahn and Stigsdotter 2003; Stigsdotter et al. 2010; Groenewegen et al. 2012; Ward Thompson et al. 2012; McEachan et al. 2015; Cox et al. 2017), more happiness and calm feelings (McMahan and Estes 2015; MacKerron and Mourato 2013; White et al. 2013; Park et al. 2007), lower rates of

autism (Wu and Jackson 2017), better overall health (Ulmer et al. 2016; Maas 2006; Maas et al. 2009; Mitchell and Popham 2007), less mortality from all causes and from cardiovascular and respiratory disease and cancer in particular (Donovan et al. 2013; Villenueve et al. 2012; Li et al. 2008), better birth outcomes (Dadvand et al. 2012; Donovan et al. 2011; Hystad et al. 2014), improved childhood well-being (Feng and Astell-Burt 2017), lower rates of allergies and asthma (Alcock et al. 2017; Hanski et al. 2012; Ruokolainen et al. 2014), reduced risk of diabetes (Dalton et al. 2016; Dendup et al. 2018), and less obesity (Dadvand et al. 2014). Several of these studies report that exposure to environmental microbes may be a mediating factor of the benefits.

Despite myriad health benefits, the amount of vegetation in many cities is declining due to population growth and urban densification (Chen et al. 2017; Gan et al. 2014; Dallimer et al. 2011). In the US, for example, out of 20 large cities, 16 experienced increases in impervious surfaces over a 5-year period, while 17 had significant decreases in tree canopy cover, translating to losses of about 4 million urban trees annually (Nowak and Greenfield 2012). Part of the problem may be that, at least in the US, there are no national regulations for how much green space should be provided for urban residents or how close it should be to their homes.

Since there is a clear relationship between greenspace and health, it is troubling that its distribution in many urban areas is inequitable. Researchers have found that disadvantaged neighborhoods tend to have less vegetation, poorer access to green space, and are more likely to be located near sources of air pollution (Harlan et al. 2006; García and White 2006; The Coalition for a Livable Future 2007; Fitzpatrick and LaGory 2000; Bell and Ebisu 2012; Dai 2011; Jennings et al. 2012; Landry and Chakraborty 2009), potentially exacerbating the health burden of low-income and minority populations. Additionally, in the process of urban densification, disproportionate vegetation loss has been observed in vulnerable and disadvantaged neighborhoods (Casey et al. 2017).

As vegetation is an important factor in creating wildlife habitat, it may be unsurprising that urban biodiversity can also vary with socioeconomic status. In several studies, neighborhood

plant and bird species diversity were positively associated with household income (Strohbach et al. 2009; Kinzig et al. 2005; Hope et al. 2003). This effect appears to be controlled by both "bottom-up" and "top-down" processes, including residential yard design and maintenance, and city park and open-space policies, respectively (Kinzig et al. 2005), suggesting that people have the ability to influence their exposures to macrobiodiversity at both individual and policymaker levels.

Biodiverse living environments provide important health benefits, therefore ensuring adequate vegetation in residential neighborhoods has been recommended as a relatively low-cost strategy for improving urban public health and combating "sick-city syndrome" (Milne 2017; Liddicoat et al. 2016). However, there is a dearth of empirical evidence describing how much vegetation is needed, of what type or configuration, and within what distance of peoples' homes. In response, scholars, designers, planners and policymakers have called for a stronger evidence base and better understanding of the mechanisms underlying that linkage (e.g., Frumkin et al. 2017; Markevych et al. 2017; Shanahan et al. 2015). Most studies to date have used observational and cross-sectional design, which suffers from the risks of bias and confounding factors and does not allow conclusions about causality to be drawn (Frumkin et al. 2017). If effective design solutions are to be conceived and reliably implemented, it is critical to bolster our knowledge with longitudinal and case-control intervention-based studies *in the settings where people actually live* to begin unraveling causal mechanisms. Deeper understanding of what underlies the vegetation-health linkage could lead to more effective, less costly and more sustainable public health strategies than interventions that are focused on treating the individual.

Potential pathways through which vegetation may affect health

Four primary pathways have been hypothesized to explain vegetation-health relationships: 1) promotion of physical activity; 2) air quality improvement; 3) mental restoration and stress reduction; and 4) fostering social cohesion (Hartig et al. 2014; Markevych et al. 2017). It is

likely that all are important pathways and that health benefits accrue to varying degrees through each of them (Kuo 2015). However, many of the associated physical and mental health benefits have been linked to more than one pathway. For instance, having vegetation near one's home has been repeatedly associated with lower levels of stress and depression (Roe et al. 2013; Stigsdotter et al. 2010; de Vries et al. 2013; Wood et al. 2017). Stress and depression, meanwhile, have been separately linked with physical activity and social cohesion, so it is unclear whether vegetation, physical activity, and social cohesion each have separate, additive effects on mental well-being, or if one of more of these pathways can be subsumed within the others.

Furthermore, these four pathways also do not consistently or fully explain the beneficial effects of vegetation. Different studies have found different pathways to be in play or not, with widely varying effect sizes and large unexplained variance (e.g. de Vries et al. 2013; Richardson et al. 2013; Dadvand et al. 2016; Hystad 2014; James et al. 2016; Vienneau et al. 2017; Dadvand et al. 2015; Triguero-Mas et al. 2017). If all these pathways are relevant but do not appear to singly or collectively explain the measured benefits, some other mechanism(s) must be in play. One candidate is immune system enhancement from exposure to beneficial microbes associated with vegetation.

Immune enhancement as a fifth, and possibly central, pathway

Spending time in and around vegetated areas may lead to enhanced immune function, which could explain the majority of observed health effects, mental and physical, across all four pathways (Kuo 2015). However, little is known about exactly how vegetation promotes immune function. Evidence has shown that immune function is affected by the other four vegetation-health pathways, i.e., physical activity, air quality, mental stress and social factors (Pedersen and Hoffman-Goetz 2000; Calderón-Garcidueñas et al. 2008; Padgett and Glaser 2003; Uchino 2006), so it could accrue through one or more of them. On the other hand, as pointed out by Kuo (2015), recent studies have suggested that walking in forests may increase immune function through exposure to plant-produced compounds known as 'phytoncides,'

including α-pinene, ß-pinene, and ∂-limonene (Li 2010; Mao et al. 2012), while exposure to environmentally-derived microbes (e.g., those associated with soil, plants, water, and animals) may represent another under-appreciated mechanism through which natural settings foster well-being (Liddicoat et al. 2018; Liddicoat et al. 2016; Rook 2013; von Hertzen and Haahtela 2006) (Figure 19).

If conclusively demonstrated, a microbe-immune system pathway could begin to give measurable answers to questions of how much vegetation, of what kind, and where, is needed to most effectively support human health.

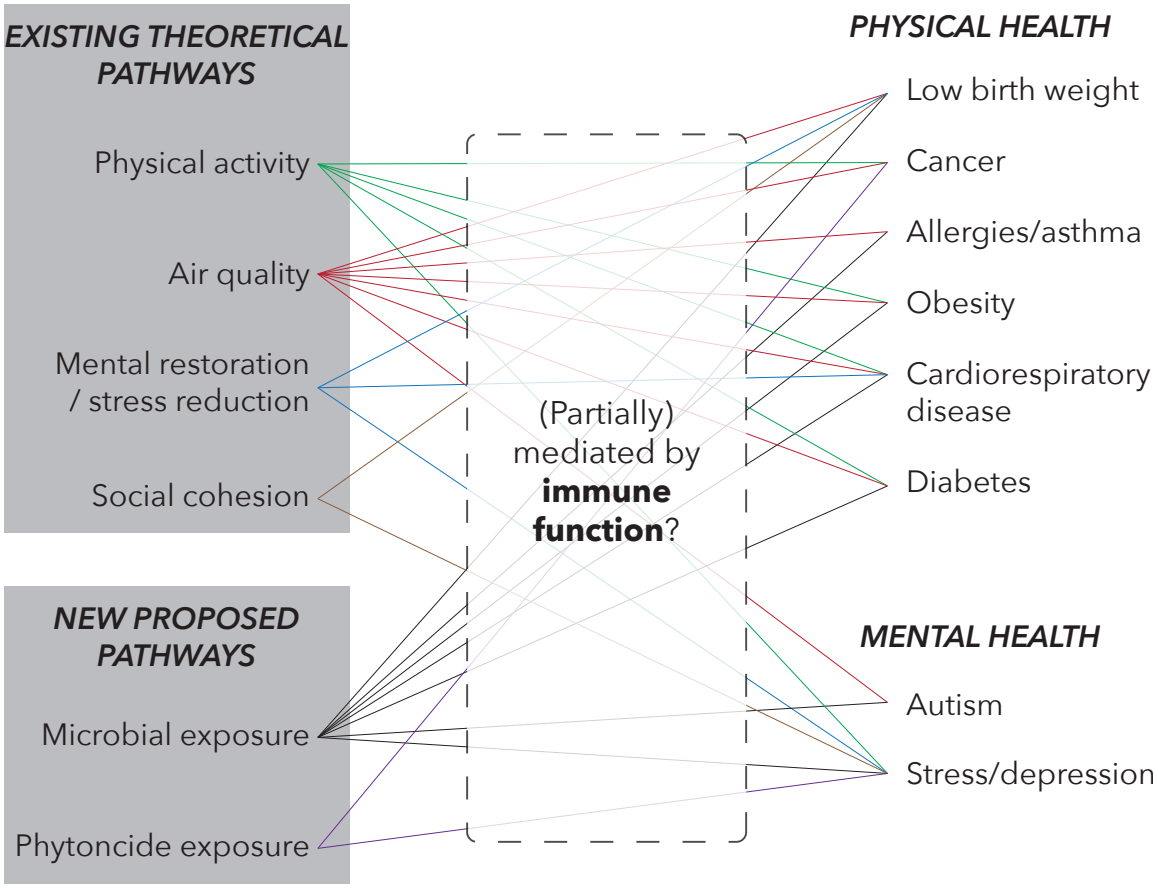


Figure 19. Pathways through which vegetation may impact health, all of which are supported in the literature.

Evidence that environmental microbial exposures impact health

According to microbial ecologists, "we live in a microbial world" (Whitman et al. 1998). People are continuously exposed to microbial communities in ways that influence their health, for better or for worse, in however subtle or conspicuous a manner. In the past, our knowledge of these invisible cohabitants of the environment was restricted to what we could learn through culturing in petri dishes, which is effective for only a small fraction of microbial life. New technologies, such as high-throughput DNA sequencing, have enabled us to better study the larger microbial community that, until recently, has been relatively unknown and remains poorly characterized.

Although a DNA-based scientific understanding of microbial exposures may be relatively new, the exposures themselves are not new to our physiology. The human body has developed a number of ways to respond to this immersion. Skin forms a barrier, mucous membranes capture and immobilize intruding microbes, and the immune system identifies and battles against recognized enemies, but for the most part we tolerate them. This relationship with the microbial world has been the status quo for our entire evolutionary history. Without these countless interactions, particularly early in life, the human immune system does not learn toleration and may become biased towards inflammatory responses, not only to pathogens but also to environmental substances like pollen and food, and even to self, possibly leading to inflammatory disorders like asthma, depression and Crohn's disease (Rook 2013). A vast number of microbes have been tolerated and co-evolved with for so long that they have become part of our own self, our commensal microbiome, without which we would almost certainly lead poorer lives (Gilbert and Neufeld 2014).

Dysbiosis, an imbalance of our commensal microbes, is now being recognized as a contributing factor to a host of chronic and autoimmune disorders, both mental and physical (Logan et al. 2016). Many of these disorders are the same as those implicated in vegetation-health studies, including depression, autism, schizophrenia, allergies and asthma, and low birth weight. As one example among many, experimental evidence points to an important

bidirectional relationship between mental health and gut commensal microorganisms, termed the 'gut-brain axis,' wherein mental symptoms can be produced and/or ameliorated by altering the composition of the gut microbiome (Cryan and Dinan 2015; Matthews and Jenks 2013). The commensal gut microbiome appears to play a role in the relationships among stress, immune function and both mental and physical disorders, such as post-traumatic stress syndrome and inflammatory bowel disorder (Bharwani et al. 2016). The commensal microbiome has even been identified as a target for therapeutic intervention. For instance, Gur et al. (2015) state that "the ability of the microbiome to impact the developing CNS [central nervous system], and participate in the effect of psychosocial stress is an exciting concept, because it is susceptible to targeting with pre- and probiotics, which has vast implications for both neurodevelopment and other health outcomes."

Despite the enthusiasm and accumulating support for the health-promoting potential of microbial therapy, it remains a goal for the future. Before we can develop firmly-supported therapeutic microbial products, additional clinical and experimental evidence is needed to identify causal direction, internal biological mechanisms, and other potential co-factors that may be in play. Some researchers have proposed that the prioritization of urban greenspaces, like parks and community gardens, could represent a low-risk strategy for delivering small doses of probiotics on a community-wide scale *if* there is adequate evidence that humans can acquire microbes from their surroundings *and* that vegetated environments have greater abundance of beneficial microbes than other land use types (Rook 2013; Liddicoat 2016). Next, we explore the current state of knowledge with regard to these two requirements.

Interactions between humans and environmental microbes

The first and most important exposures to environmental microbes occur in utero and early infancy. This time window sets the trajectory of physiological systems such as the immune system (McDade 2005). For instance, during the prenatal period, maternal exposure to a diversity of farm animals has been linked with the long-term immune function of the

developing infant and is likely to be mediated by microbial interactions related to animal husbandry (von Mutius 2016). After birth and while still under the protection of maternal antibodies from breastmilk, infants begin learning about their physical setting almost immediately after exiting the womb, largely via the senses of touch and taste. Infants explore their surrounds by touching and placing hands and objects in their mouth; some researchers suggest that this helps calibrate the immune system to a best-fit trajectory suited to the pathogen load in the local environment (Fessler and Abrams 2004; McDade 2005). By ingesting potential commensal microbes this behavior may also influence the long-term composition of our individual microbiomes. In this early period of microbial assembly, priority effects (meaning that the first microbes to establish may impact the ability of subsequent taxa to colonize) can have lasting effects on host metabolism, immune function, and mental development (Sprockett et al. 2018).

Although microbiota compositions at various body sites (e.g., skin, intranasal and lung, gut) are relatively stable over time (Faith et al. 2013; Oh et al. 2016; Frank et al. 2010), the adult commensal microbiome may be influenced by environmental exposures. For the purposes of this discussion, we are primarily interested in exposure to airborne microbes, which occurs primarily via skin contact, inhalation, and ingestion. There are typically 10^4–10^8 bacterial cells in a cubic meter of air (Bowers et al. 2010) and we inhale 8–20 liters of air per minute under normal daily activities (Adams 1993), which translates to somewhere between 115,000–2,880,000,000 bacterial cells inhaled per day, to say nothing of fungi, viruses and other microbes. Many of the larger particles are cleared out of the respiratory system by mucociliary action and enter the gastrointestinal tract (Kish et al. 2014). Some of these exposures can lead to colonization by environmental microbes, as has been shown for humans and other mammals (Mulder et al. 2009; Seedorf et al. 2014; Brooks et al. 2014; Kraemer et al. 2018; Camarinha-Silva et al. 2014). For example, there are rural-urban differences in skin microbial community (Ying et al. 2015; Lehtimaki et al. 2017), and transient contact with soil-associated microbes can leave an imprint on the skin microbiome for at least 24 hours afterward, even after washing and bathing

(Fahimipour et al. in prep.). Additionally, commensal microbiota of the human nose differ strongly between pig farmers and cattle farmers, indicating that airborne microbes associated with certain environments can colonize human mucosal surfaces (Kraemer et al. 2018).

The environmental microbial composition to which humans are exposed can vary substantially across different land uses. Initial evidence suggests that some land use types can impact human health, particularly asthma and allergies, by interacting with the commensal microbiome (Jatzlauk et al. 2017). One example is the work done by Ege et al. (2011), showing that children living on farms were exposed to greater microbial diversity and had less allergies and asthma than urban children. In a follow-up study, the risk of having asthma, atopy, and/or hay fever was inversely related to exposure to certain bacterial genera, including *Acinetobacter* and *Lactobacillus* (Ege et al. 2012). Other studies have indicated that growing up near farms and forests exerts a protective effect against childhood atopy, and may act by influencing skin bacterial composition, in particular *Acinetobacter lwoffii* (Hanski et al. 2012; Ruokolainen et al. 2015; Fyhrquist et al. 2014).

Several lines of experimental research suggest that variation in exposure to environmental microbes may affect not only physical, but also mental and emotional state. For instance, exposure to *Mycobacterium vaccae*, a soil-associated bacterium, has been shown to reduce stress and anxiety-related symptoms in mice (Lowry et al. 2007; Matthews and Jenks 2013) by increasing serotonin (a "happiness" brain chemical) levels. Although considerably fewer investigations have focused on the skin microbiome than on the gut, initial evidence suggests that commensal skin microbes affect the localized immune system (Naik et al. 2012). This may occur through several pathways: 1) skin cells respond to exposures to new, potentially commensal, microbes by accumulating greater numbers of cytokine-producing T cells; 2) commensal microbes respond to encounters with potentially pathogenic microbes by producing targeted antibiotics to inhibit pathogen survival; 3) commensal microbes signal skin cells to mount immune defense against pathogens; 4) commensal microbes promote accelerated wound healing; and 5) commensal microbes regulate expression of genes that control development and

differentiation of skin tissue (Naik et al. 2012; Eyerich et al. 2018; Linehan et al. 2018; Meisel et al. 2018). It has even been speculated that responses of human skin cells to environmental cues, possibly including microbes, can affect our emotional states (Prescott et al. 2017).

This brief review of the literature suggests that exposure to environmental microbes, particularly associated with farms, forests, and soil, may alter the commensal microbiome and affect both psychological and physiological processes, including immune function. With this foundation in mind, we next examine design-relevant factors that may control the taxonomic composition of microbes we encounter in our surroundings.

Factors that determine urban airborne microbial community assembly

Landscape configuration and composition, including vegetation and built infrastructure, may affect aerobiome assembly at several scales of space and time (Chapter III). In some ways, investigation of the aerobiome may have parallels with biogeographical studies in ocean environments, as both types of media are dynamic and inevitably three-dimensional (Womack et al. 2010). As described next, we posit that airborne microbial composition may be influenced by landscape design through three fundamental means: 1) by determining the types and distribution of different sources of airborne microbes, which are contributed from virtually all types of surfaces, including vegetation, soil, water, and built surfaces; 2) by controlling factors that influence the survival and growth rates of microbial taxa; and 3) by shaping microbial dispersal patterns, largely through air flow through neighborhoods and across the urban environment as a whole (Burrows et al. 2009; Chapter III).

Microbial sources

Most airborne microbes originate from surfaces that provide the conditions for persistence, growth, and reproduction (e.g., temperature, relative humidity, irradiation, nutrient availability). Therefore, aerobiome composition is likely related to the amount and configuration of nearby biotic and abiotic surfaces. Plants and soil have been noted as especially important sources of

microbes to the air above terrestrial environments (Lindemann and Upper 1985; Lindemann et al. 1982; Bowers et al. 2011; Lymperopoulou et al. 2016); here, we describe some characteristics of vegetation as a source and recognize that where there is vegetation there is also often unsealed (i.e., not paved over or built upon) soil. Soil comprises an enormously complex and influential source, which we will not address but refer to recent reviews (see, for example, Fierer 2017) for additional information. Other surface types, such as water, asphalt, and stone, are likely to act as substrates for different microbial communities as well.

Leaf surfaces comprise an estimated billion or more square kilometers of area—the largest biological surface type on the planet—and may contain between 10^6 and 10^7 bacteria alone per square centimeter (Lindow and Brandl 2003; Vorholt 2012). Not all leaf surfaces are alike from the standpoint of microbes, however. Bacterial community composition of leaf surfaces is influenced both by site conditions and plant species (Knief et al. 2010; Izhaki et al. 2013; Delmotte et al. 2009; Kembel et al. 2014; Laforest-Lapointe et al. 2016). For example, a study investigating tree leaf bacterial communities along an urban gradient found that bacterial composition varied as a function of urban intensity (i.e., a composite measure including human population density, roadways, built-up area, land use, and other infrastructure), tree species, and degree of tree isolation (e.g., individual street trees) (Laforest-Lapointe et al. 2017). In addition, plant leaf surfaces have been observed to host different communities depending on whether they are growing in urban or rural locations (Smets et al. 2016). These leaf-inhabiting microorganisms become airborne during plant processes like evapotranspiration, as well as by meteorological processes, such as rain splash, wind gusts and thermal plumes (Lighthart et al. 2009; Whipps et al. 2008; Kinkel 1997).

Microbial growth and survival

There is great variability in the resilience of microbial taxa to environmental conditions, and the effects of being airborne on their ability to proliferate are poorly understood, although recent studies indicate that atmospheric organic compounds may support microbial metabolism

and that microbes may be metabolically active while airborne (Fröhlich-Nowoisky et al. 2016; Klein et al. 2016). For airborne bacteria, most taxa decline in growth and survival at temperatures above 24° C (75.2° F), while the response to relative humidity appears to be bimodal, with highest survival at very low and very high relative humidity (Tang et al. 2009), and solar radiation generally reduces survival rates (Tong and Lighthart 1997). Nutrient availability, pH, and salinity are also important factors controlling growth and survival, though a vast range of substrates, including carbon-, sulfur-, phosphorus- and nitrogen-based compounds, can be utilized by different microbes. Even in the face of adverse conditions many microbes exhibit strategies that allow them to persist, such as the ability to enter a non-growth, dormant, or spore phase.

Microbial dispersal

There is constant interplay among microbial communities transported from distant continents via tropospheric air flow, those coming from large regional sources, and those that are emitted locally (Seifried et al. 2015; Innocente et al. 2017; Smith et al. 2012; Lighthart 1984; Mhuireach et al. 2016). The background flow of air, known as the "synoptic flow" contains a mixture of microbes and particles, which may have been transported from relatively short or very long distances away. For example, the Pacific Northwest receives particles from Asia (Timonen et al. 2013; Smith et al. 2012) and Saharan dust travels to the Caribbean (Monteil 2007).

Since airborne microbes fall within the same size range as particulate matter (Liu and Liptak 1997) and are assumed to follow the same principles of aerosol physics, we hypothesize that in open areas, microbial composition will be largely determined by conditions upwind because the constant air exchange tends to both dilute locally-emitted microbes with those from the regional air mass and to carry them downwind. Supporting this, Mhuireach et al. (2016) found that the overall composition of airborne microbial communities collected from open grass parks and parking lots was very similar when focusing on the most abundant taxa,

which they conjectured were part of the regional air mass originating upwind. The size and intensity of the local source, however, is also important in determining its contribution to site air composition and dispersal capability. For example, a recent study revealed that microbial abundance directly above vegetated areas and 50 meters downwind is 2–10 times greater than in upwind non-vegetated areas (Lymperopoulou et al. 2016).

The point of this brief consideration of microbial sources, growth and survival, and dispersal is that it appears plausible that surrounding land uses, vegetation types, and management activities from the regional to the city block scale can all impact airborne microbial community composition (Chapter III, Figure 7) and, in turn, that many of these factors can be influenced by urban design and planning decisions. Figure 20 illustrates some of the linkages between landscape features and microbes and between microbes and health outcomes across a range of spatial scales that have been put forth in the literature. See the review by Hoisington et al. (2015) for additional microbe and health linkages. In the next section, we propose a framework of how design and planning of landscape features can structure microbial assemblages through their influences on microbial sources, growth and survival, and dispersal. We then harness this framework as a new building block for evidence-based urban landscape design.

A framework for evidence-based urban landscape design for health

Although there is far from a complete picture of how microbes assemble in the urban environment or how environmental microbes may affect human health, we can begin building a framework to facilitate translation of scientific knowledge into design-relevant language. We hope that this will aid the development of design hypotheses and interventions that advance our understanding of these relationships. We propose such a framework, which consists of three components: 1) a conceptual model that illustrates how common landscape features and processes may influence airborne microbial sources, growth and survival, and dispersal; 2) a design process to guide landscape architects in assessing the opportunities and constraints presented by a given site for structuring local airborne microbial communities; and 3) a design

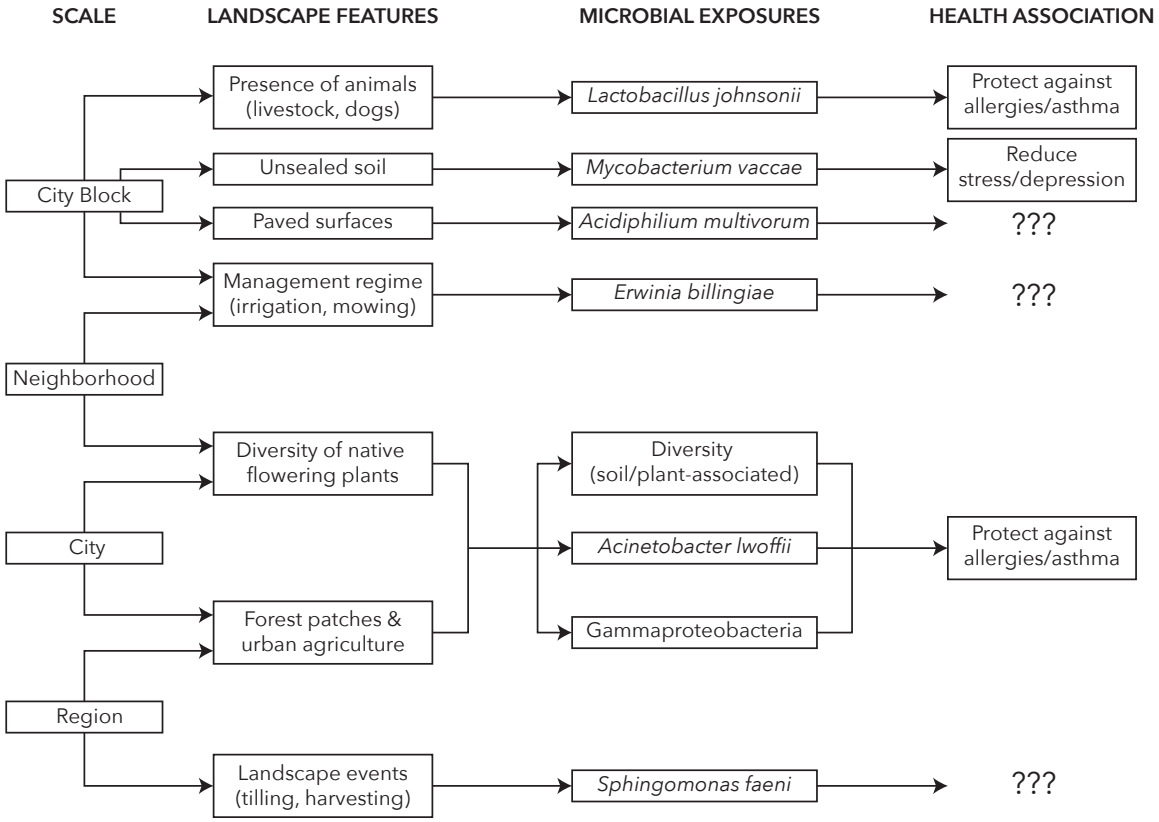


Figure 20. Specific landscape features associated with microbial exposures and potential health impacts.

hypothesis-intervention approach to producing new knowledge about the effects of landscape design on microbial communities.

Initially, the framework may be most useful for introducing the aerobiome as a novel facet of landscape ecology that can be shaped by designers and for developing experimental interventions to test hypotheses that link design, the environmental microbiome, and public health. As the scientific evidence base expands, we can use the same framework to manage our exposures through landscape design. Although protecting human health from airborne pathogens in indoor settings has been considered at length (e.g. Kowalski and Bahnfleth 1998; Leung and Chan 2006), designing to promote exposure to beneficial outdoor microbial communities has not. To our knowledge, the sole contribution is that of Mills et al. (2017), which proposes an "urban microbiome rewilding" program. They suggest that environmental

restoration projects to increase urban biodiversity may represent a means to improve human immune health. With our conceptual framework we seek to extend and clarify the idea of promoting health through microbial exposures associated with vegetated environments. First, we identify landscape features that are familiar to designers and planners and that exert influence on microbial sources, survival and growth, and dispersal.

Landscape features that influence microbial sources, growth and survival, and dispersal

In the outdoor environment there are likely thousands of different microbial taxa present in varying abundances, depending on configuration of, and interactions among, landscape features. The specific composition of an urban microbial community, we hypothesize, is a function of three interrelated features: 1) land cover, which determines the type, amount, and composition of biotic and abiotic surfaces that act as microbial sources and can influence growth and survival; 2) microclimatic conditions, which may impact microbial growth and survival and dispersal, and are influenced by land cover type, as well as topography, hydrology, and other factors; and 3) intermittent or cyclical processes and events that may affect sources, growth and survival, and dispersal of microbes (Figure 21). Each of these three features may be influenced by designers, planners, and/or policymakers, therefore, understanding how and why they are related to aerobiome composition could help with decision-making.

Land cover

At the coarsest scale, broad patterns of land cover and land use, in aggregate, determines regional microbial emissions (Burrows et al. 2009). For instance, microbial communities vary by major land uses, such as forested, agricultural, and urban areas (Bowers et al. 2010; Shaffer and Lighthart 1997), and recent studies have indicated that vegetated land cover types at the fine scale (e.g., city block scale) have significantly different bacterial community composition than non-vegetated land cover types (Chapter III).

Urban forests may be of particular interest to designers of microbial landscapes, for several

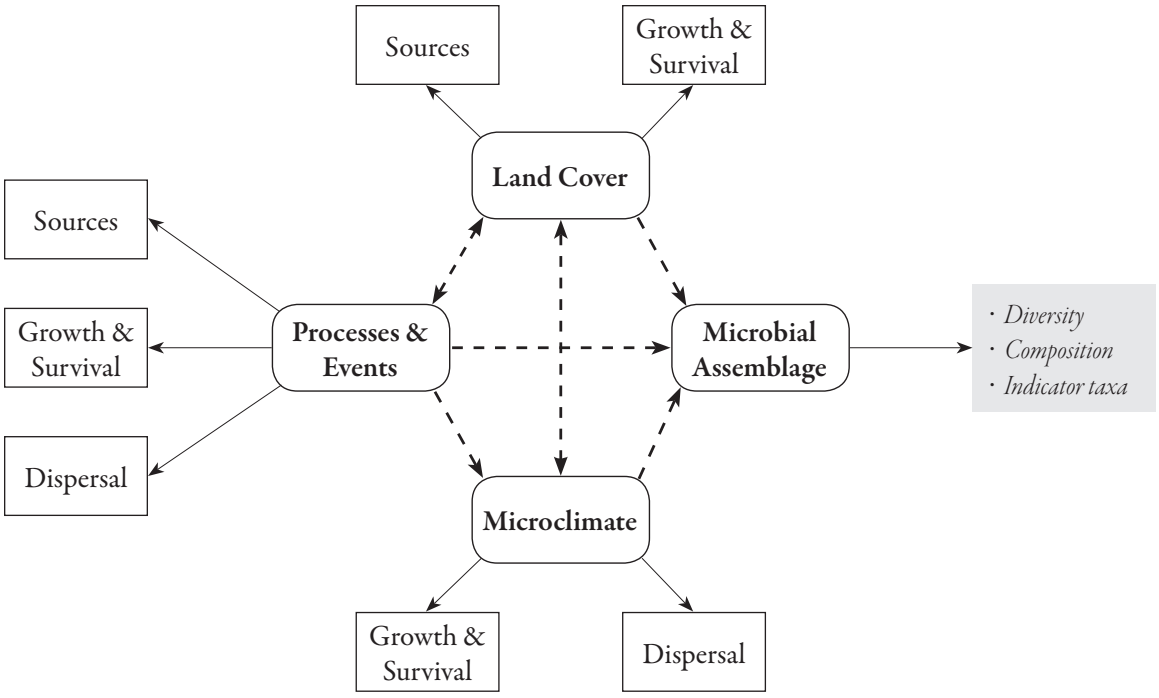


Figure 21. Conceptual model of landscape features that can structure microbial assemblages through their influences on microbial sources, survival and growth, and dispersal.

reasons. First, researchers in Colorado have found that airborne microbial communities within a forest were more similar to soil communities than other site types (Bowers 2010). Second, studies suggest that forests have higher abundance of Gammaproteobacteria than built areas (Hanski et al. 2012) and greater generic diversity of Gammaproteobacteria than grass areas (Chapter III), both of which have been linked with lower incidence of allergic sensitization (Hanski et al. 2012). Finally, forested areas create a different microclimate below the canopy than open areas, since they tend to provide shelter from sun, wind and precipitation. Not only would the modulation of microclimate influence which microorganisms might thrive in that environment, but it would also reduce the rate of air "turnover," which, in turn, may better retain the microbes from local sources.

Another example of an urban land use type that may be a source of distinct microbial taxa is the parking lot. Paved parking lots are characterized by temperature extremes that are generally outside those experienced in vegetated sites, surfaces that are desiccated and exposed to harsh radiation, and the presence of organic compounds (e.g., automobile fuels and fluids)

and heavy metals that are not typically found in vegetated environments. Many of the microbes found in these environments (e.g., *Acidiphilium multivorum*) are those associated with extreme environments, such as acid mine drainages, sewage sludge, and sulfur hot springs (Chapter III). It is currently unknown whether exposure to these "extremophiles" has implications for human health, although many of them may be useful as bioremediation agents (Auld et al. 2013).

Microclimate

Structures in the landscape, like trees and buildings, can impact microbial growth and survival and dispersal by creating sun and shade patterns, influencing humidity, and altering air flow in their immediate locale. This diversity of microclimates can make a given location more or less hospitable for microbes (and humans). These effects can also change over the course of a day or year, depending on sun angle, vegetation leaf condition, wind direction and precipitation.

At scales larger than a city block (i.e., neighborhood, city, region), the form and organization of buildings, vegetation, and topography act in aggregate to affect air speed and turbulence, thus determining the degree of mixing among microbial assemblages from regional and local sources (Wuyts et al. 2008; Grimmond and Oke 1998). Specifically, the height, volume and "packing density" of buildings and trees act together to influence air movement patterns. When structures are low and sparse, turbulence is moderate and the air is generally well-mixed among the buildings. As the packing density grows tighter, the buildings begin to act in aggregate as one object forcing air flow up (i.e., "skimming flow"), reducing turbulence and air mixing among buildings. Around taller structures, on the other hand, turbulence is great and the air is thoroughly mixed (Britter and Hanna 2003).

Buccolieri et al. (2010) introduce the idea of city "breathability," or ability to remove and dilute pollutants or heat, similar to the idea of air exchange rate in buildings, which would also apply to microbes. According to their simulation study, the "mean age of air" increases with packing density and with distance downwind from the approaching air flow, suggesting that the urban aerobiome may be more influenced by nearby local sources when packing density is great.

Processes and events

The composition of airborne microbes may change seasonally due to the effects of vegetation phenology, human activities, and weather patterns (Franzetti et al. 2010; Bowers et al. 2013) on microbial sources, growth and survival rates, and dispersal. For instance, an annually-repeating pattern of microbial succession during the growing season has been observed on deciduous plant leaves (Knief et al. 2010), with microbial composition clustering into early, mid-, and late season groups (Redford and Fierer 2009). Human management activities can be important at several spatial scales. At the regional scale, activities such as agricultural harvesting can contribute enormous amounts of biological material, including microbes, into the air, with subsequent dispersal to downwind areas (Lighthart 1984; Mhuireach et al. 2016; Chapter III). Since particular taxa may preferentially inhabit certain types of agricultural crops, the relative abundance of these crop-associated microbes may be greatly increased during harvesting season. One example is the high relative abundance of *Sphingomonas faeni*, a bacterium associated with hay dust (Andersson et al. 1999), in the air downwind of Linn County, Oregon, colloquially termed "the grass seed capital of the world," in the prime harvesting month of July (Mhuireach et al. 2016; Chapter III). At a smaller scale, vegetated sites that were intensively maintained, like cemeteries, golf courses, and parks, tended to be associated with bacterial taxa (e.g., *Erwinia billingiae*) that were not prevalent in other sites (Chapter III). Such taxa may be contributed to the air through management activities, including irrigating and mowing.

Design process to shape airborne microbial communities

To develop design interventions intended to shape airborne microbial communities, it is first necessary to assess existing landscape features that may influence how these communities assemble (Figure 21) across relevant scales. In particular, identifying important regional and site-scale sources, primary air flow patterns and microclimate conditions, and local management activities will provide a foundation for developing design goals and hypotheses. Following an initial assessment, the designer can interpret opportunities and constraints in light of explicit

design-based hypotheses and develop specific interventions to address them (Figure 22). For example, being downwind of a large waste remediation operation might be seen as a constraint to be blocked or diverted, whereas being downwind of a biodiverse forest might be seen as an opportunity to encourage the flow of potentially beneficial air. One caveat to this process is that we still lack adequate knowledge about which microbial assemblages might be beneficial or detrimental to human health. We address this challenge by proposing a design hypothesis-intervention approach to building the scientific evidence base for what constitutes a beneficial urban microbiome.

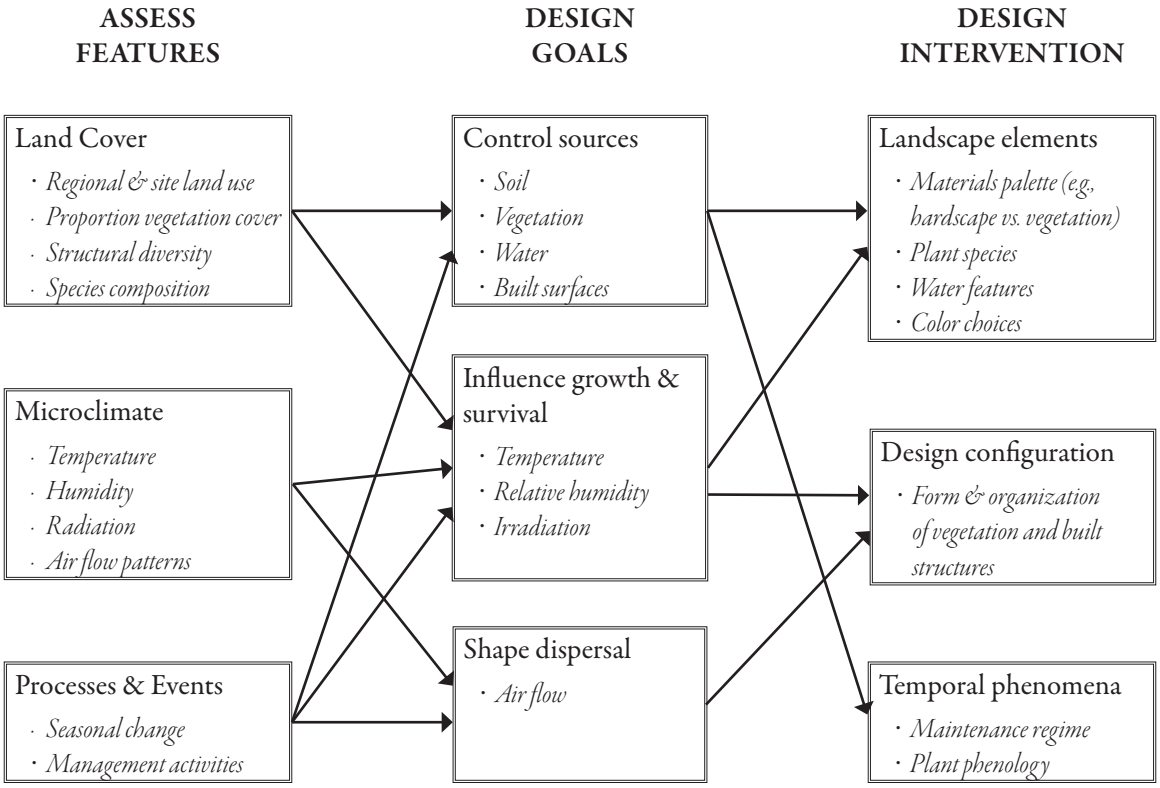


Figure 22. Design process for assessing the opportunities and constraints presented by landscape features and develop design interventions to influence airborne microbial communities.

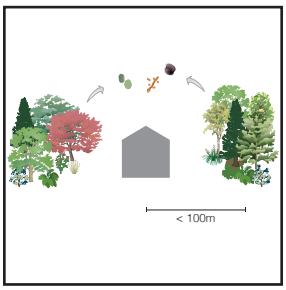
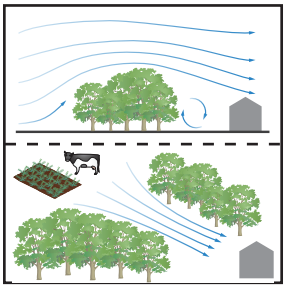
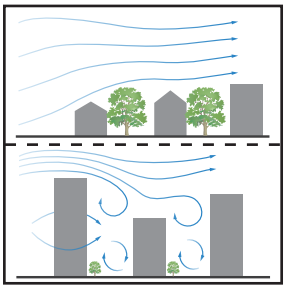
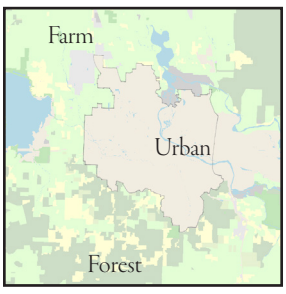
Hypothesis-intervention approach to knowledge production

Studying health determinants in the settings where people live, work, and play means that clinical and case-control methods are often difficult or impossible to apply. This explains why

most vegetation-health studies are observational and cross-sectional, rather than mechanistic. If we hope to purposefully create salutogenic landscapes by managing the urban microbiome, we must understand the causal mechanisms that mediate between landscape features and human health via microbial exposures. Since urban construction projects are occurring all the time and all over the world, one approach would be to use new projects as opportunities to measure outcomes related to both microbial communities (e.g., diversity, composition, indicator taxa) and human health (e.g., stress, depress, autoimmune disorders). This type of post-occupancy evaluation (POE) is already common in design fields as a means to assess how well the design achieves stated goals. Using this approach, each action taken to alter the urban environment is viewed as an experiment from which valuable health and microbiome data could be collected. POEs are critical to elucidating cause and effect and should include well-validated measures of well-being, ranging from subjective surveys and questionnaires to objective measures, such as biological markers of stress and inflammation or skin microbial communities.

We propose several preliminary and diagrammatic examples of design hypotheses and interventions that attempt to influence microbial community structure and dynamics (Table 3). These range in scale from the city block to the region, and are each based on findings reported in the previous sections. To test such hypotheses and interventions, one might use study designs ranging from before/after at a particular site to paired new developments with contrasting interventions. In addition to these examples, we note that a vast range of hypothesis-intervention opportunities exist. For instance, while studies have indicated that urban land use types (e.g., forest, grassland, paved) differ in their microbial communities, there can be substantial variation in vegetation cover, species composition, and management regimes across a single land cover type and it is unknown whether this within-group variation influences microbial community structure. A fruitful research avenue might be to compare microbial communities from different grassland types (e.g., mowed monoculture lawn vs. diverse meadow) or parking lot configurations (e.g., those with bioswales and tree canopy cover vs. those without). In addition to collecting data about how these "designed experiments" might

Table 3. Design hypotheses and interventions across spatial scales for certain design goals.

Spatial Scale & Design Goal	Hypothesis & Intervention	Visualization
Block: Determine sources	*H*: The concentration of microbes from local soil and plants is rapidly diluted more than 100 m from their source. *I*: Establish >80% vegetation cover within 100 m of homes to promote soil- and plant-associated microbes rather than pavement-associated microbes.	
Neighborhood: Shape dispersal	*H*: Large upwind microbial sources can affect the composition of airborne microbes at the neighborhood scale. *I*: Create windbreaks or wind funnels to impede or encourage air flow, thereby decreasing or increasing the regional microbial signal relative to the local signal.	
Neighborhood: Determine sources	*H*: Forests and areas with high plant diversity harbor greater abundance and diversity of Gammaproteobacteria. *I*: In residential zones, establish biodiverse public parks & forests to increase abundance and diversity of class Gammaproteobacteria (may be protective against allergy).	
City: Shape dispersal	*H*: Upwind regional and tropospheric air masses influence the composition of the urban airborne microbiome. *I*: Design gradual height transitions with vegetation or building structures to divert air flow above and reduce mixing; alternatively, use random height variation to encourage mixing with regional air mass.	
Region: Determine sources	*H*: Different land use types represent aggregate sources of particular microbial community compositions. *I*: Evaluate regional land use for health implications; farms/forests may promote microbial diversity and Gammaproteobacteria, in particular.	

impact the local airborne microbial communities, assessing relationships with human health outcomes or markers would also be essential. Critically, we must realize that these studies are needed to first provide a baseline of data so that we can begin to set standards of "healthy" versus "unhealthy" landscape microbial communities. Once we understand what constitutes the baseline communities, then we can evaluate how design interventions change them.

Results from such intervention-based studies would be incorporated into the body of scientific knowledge to revise hypotheses, develop theories, and eventually improve subsequent design interventions (Figure 23). Moving iteratively from science to design to implementation and back to science in this way has been argued as an important pathway to ensure that science is informed by what works and does not work in practical application and that design is founded in rigorous knowledge (Nassauer and Opdam 2008; Felson and Pickett 2005; Cook et al. 2004)). Evidence-based design is often seen as a 'one-way street,' where designers can use scientifically-produced knowledge to create better designs, however, these types of projects are not always evaluated as experiments and even more rarely are the evaluations fed back into the

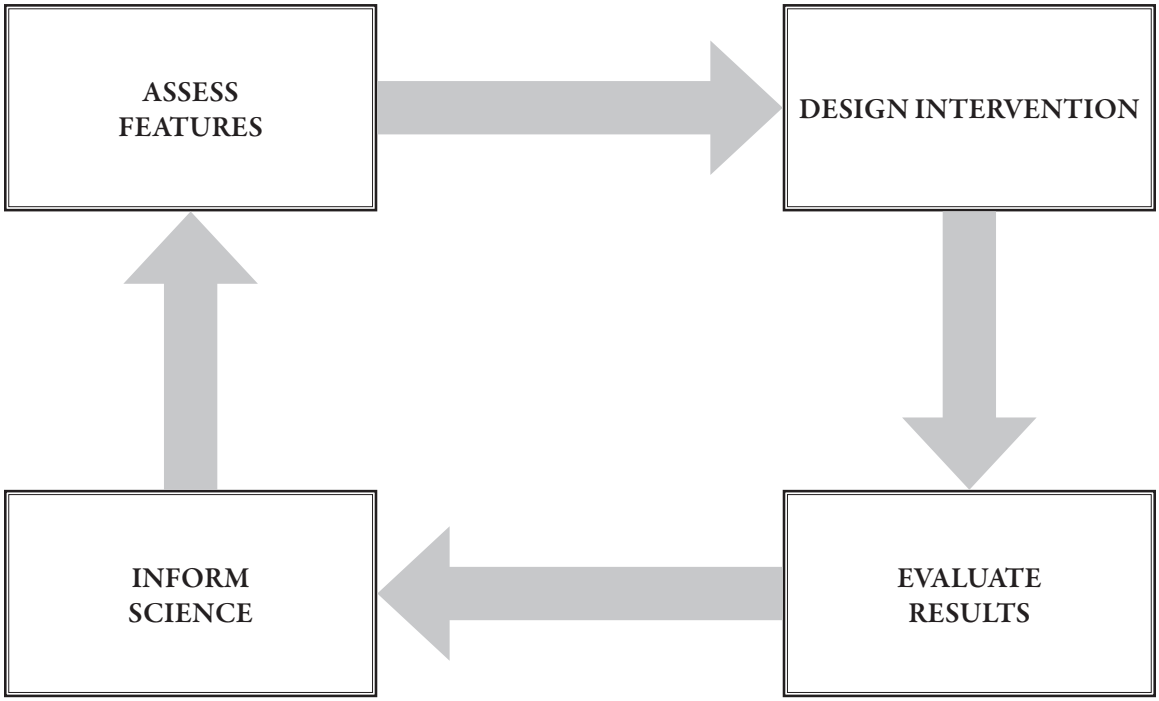


Figure 23. Iterative science-design-science process.

scientific knowledge base through peer-reviewed journal articles. We believe that both designers and scientists could benefit by collaborating to improve our understanding of linkages among design of the built environment, microbial exposures, and human health. This iterative science-design-science concept is well-aligned with the field of implementation science, which is used by many disciplines and includes a number of related strategies, such as evidence-based practice and practice-based evidence (medical and health fields, education, architecture) and designed experiments and adaptive management (ecology, forestry, fisheries). At its core, the goal of implementation science is to foster the flow of scientific knowledge into routine use and ensure that laboratory findings translate successfully to practical application. Equally importantly, this approach can help validate the accuracy of clinical findings and refine research questions based on feasibility and applicability in non-laboratory settings. Such studies can also generate valuable tools, including models for prediction and decision-making (Frumkin et al. 2017).

Public health and urban planning are in the midst of reuniting, although now using contemporary cutting-edge tools and knowledge, such as DNA sequencing and the Biodiversity Hypothesis, rather than the historic miasma theory. Still, much remains to be discovered. We must identify which microbial community types and taxa are "good" versus "bad" for human health, possibly differing across human life stages (e.g., immune-training during infancy, stress-relieving during adulthood, anti-cancer during old age), and their dominant means of interaction (skin contact, inhalation, or ingestion). Collaboration among scientists, designers, and health professionals is increasingly needed to further the development of landscape intervention strategies that could influence urban microbial assemblages by intentionally specifying elements that influence their growth and survival, or that alter the degree of mixing that occurs between locally- and regionally-sourced microbes.

Conclusion

An estimated 2 billion people will be born over the next 30 years, most of whom will reside in urban areas (U.N. 2017; U.N. 2014). Urban infrastructure (transportation, sewer and water,

electricity, communication systems) is difficult and costly to alter once built, and the cities and neighborhoods we design now will be the places where our children and grandchildren live, work and play. It is thus of utmost importance to determine principles and guidelines for creating healthy cities. According to René Dubos (1966) it has been "established beyond doubt that early environmental influences affect the whole of human life—even more profoundly and lastingly than Hippocrates had anticipated." Ample evidence suggests that urban vegetation has major health implications and should be viewed as a public health necessity rather than an amenity or luxury; burgeoning theories speculate that some of the health benefits may be due to immune enhancement properties of natural environments. If it can be rigorously documented that the immune enhancement pathway is mediated by microbial exposures, it could support measurable standards for assessing and prescribing urban vegetation and ensuring its equitable distribution as cities grow and become more densely populated.

The primary contribution of this chapter was the construction of a conceptual framework linking the science of microbial ecology with design practice, which can be used to inform future scientific investigations as well as to apply the knowledge gained to design. Although we do not fully understand the drivers of airborne microbial assembly in the urban environment, there is sufficient evidence to develop initial hypotheses for shaping the urban aerobiome through landscape design choices that could be implemented and tested in real projects.

Although well over 90% of our time is spent inside buildings and cars (Klepeis et al. 2001), it should be recognized that the ultimate source of all indoor air is the immediate outdoor air. Outdoor microbial diversity associated with natural elements is important not only when we spend time outdoors, but also because it determines in large part the diversity of indoor microbiota (Parajuli et al. 2018). Thus, the management of microbial sources, the microclimatic conditions they experience in situ and as they disperse, and the ways in which air movement affects their retention or dispersal have the potential to become the primary tools and building blocks through which we can shape a salutogenic airborne urban microbiome from the outdoors to the indoors.

We end with a quote from landscape architect Anne Spirn (1988, 110):

> The dialogue between the human organism and the natural environment takes place on both an unconscious and a conscious level. Before humans built towns and cities, our habitat was ordered primarily by nature's processes . . . The most intimate rhythms of the human body are still conditioned by the natural world outside ourselves: the daily path of the sun, alternating light with dark; the monthly phases of the moon that tug the tides; and the annual passage of the seasons.

CHAPTER V

CONCLUSION

The overarching goal of this dissertation was to demonstrate the influence of vegetation on the urban aerobiome and its potential as a mediating factor linking vegetation and human health. Although the health benefits of vegetation have been widely documented, designers and planners continue to struggle with vague and 'woolly' practical implications of that research. Specifically, there are still no guidelines for what the best type or composition of vegetation is, how much of it there should be, or how far it should be from residents' homes. Although my research did not answer these questions (nor did it intend to), it did show that the urban aerobiome may be influenced by nearby vegetation type and abundance and by human management activities.

A key contribution of this dissertation was that I investigated urban microbial ecology from the perspective of a landscape architect, and I worked closely with designers, biologists, and social scientists to accomplish the research. I believe that this type of transdisciplinary collaboration is key to solving most of the complex problems we face due to increasing population growth, urbanization, and environmental uncertainty.

Summary of Results

Chapter II

While airborne microorganisms are ubiquitous in urban areas, the influence of nearby vegetation on the urban aerobiome remains poorly understood. In this chapter, I established that there is a significant compositional difference in airborne bacterial communities at fine spatial scales by examining two urban land cover types with the highest contrast in vegetation—urban parks and parking lots. I also demonstrated that passive collection of airborne microbes via settling dishes provided similar results to active collection via vacuum pumps and was a

defensible method to use in the large-scale collection campaign. This work sets a foundation for understanding how urban vegetation may impact the aerobiome, with potential implications for designing neighborhoods and open space systems that foster better human health.

Chapter III

Disentangling the complex interactions that drive the assemblage of the urban aerobiome at different scales of space and time could benefit long-term green infrastructure plans for healthier and more equitable cities of the future. In our study, site location and sample collection date were strongly related to both richness and compositional variation of airborne bacterial communities. Community composition was influenced by land cover type but the signal was overwhelmed by compositional shifts associated with date. Certain taxa appeared to be indicative of land cover type (e.g., class Gammaproteobacteria) while others tended to signal temporal changes (e.g., *Sphingomonas faeni*).

Overall, we concluded that the interplay among spatiotemporal controls on microbial assemblage can be understood by looking at site-scale features (e.g., land cover type, vegetation species composition), regional-scale drivers (e.g., agricultural activities, seasonal vegetation changes), and periodic events (e.g., precipitation, irrigation, and mowing). Our key message was that human land use decisions and management activities can and do influence the composition of airborne bacterial communities in ways that could potentially affect urban health, thereby warranting further investigation for urban design implications.

Chapter IV

Public health is (re)entering a phase where the external environment is recognized as important to human well-being. The new "settings-based" approach to health (i.e. considering individual health through a lens that includes the day-to-day living environment) strives to reinvigorate the historic emphasis upon environmental conditions that may impact health and lifestyle choices. Because urban infrastructure (transportation, sewer and water, electricity,

communication systems) is difficult and costly to alter once built, the cities and neighborhoods we are designing now will be the places where our children and grandchildren live, work and play. It is thus of utmost importance to determine principles or guidelines for creating healthy cities.

Increasing urban vegetation constitutes a public health intervention with enormous potential benefits; it is associated with numerous immune-mediated and inflammatory chronic conditions, which may be at least partly mediated by microbial exposures. Although there is far from a complete picture of how microbes assemble in the urban environment, we have sufficient evidence to begin developing hypotheses for shaping the urban aerobiome through landscape design choices, implementing the hypotheses in real projects, and evaluating the resulting data to inform subsequent hypotheses. This iterative incorporation of scientific findings into experimental design practice and then evaluating whether those designed experiments provide the anticipated benefits is an important pathway to ensure that science is informed by what works and does not work in non-laboratory settings. The primary contribution of this chapter to the discipline was the construction of a conceptual framework linking the science of microbial ecology with design practice, which can be used to inform future scientific investigations as well as to apply the knowledge gained to design.

Future Directions

Promoting human well-being in urban areas is a grand challenge in the 21st century, which has been so far characterized by unprecedented changes in our environments and lifestyles. Although humans are extraordinarily adaptable, our underlying genetic structures and requirements for optimal health have not appreciably changed over the millenia. Vegetation, and biodiversity in general, is a vital component of our environmental life support system; it nourishes us, keeps us sane, awes us with its exquisite forms and, perhaps, hosts the microbes that educate our immune systems.

I believe that the most consequential direction for future research is further elucidation of

how microbes affect human health. We have demonstrated that design and management can alter urban aerobiome assemblage, but we don't yet know which microbes are "good" or "bad" for people's health, or whether it is more important to consider the diversity and composition of the entire community we encounter. It would be valuable to identify other health-relevant individual taxa, or particular immune-training assemblages that are associated with certain land uses or activities. Once we ascertain which taxa and/or assemblages are beneficial, then we can conceive design strategies to foster the beneficial and inhibit the detrimental. In this dissertation, I focused on bacteria, but fungi, viruses, and other microorganisms can also impact health. Future work could investigate how urban vegetation influences these other groups.

Since my dissertation demonstrated that the compositional differences in the urban aerobiome are measurable and significant across land cover types, future research might examine urban land cover types that were not studied here. For instance, exploring microbial communities across the gradient of compact downtown core to single-family residential zones to dispersed peri-urban neighborhoods and linking these exposures to human health outcomes would generate important knowledge about the places where people actually live. In particular, exploring the connections between therapeutic green environments, like gardens, and microbial exposures could be important. It would also be useful to evaluate the potential consequences of environmental microbial exposures for different demographic groups, particularly vulnerable populations such as the elderly, children, and those with mental disorders.

Furthermore, a deeper exploration of the conceptual model illustrated by Figure 7 (Chapter IV) is warranted to advance our understanding of how interactions among spatial and temporal features across different scales drive aerobiome assembly. In particular, studies intended to elucidate whether urban design decisions at the local scale or policy decisions at the regional scale affect microbial communities and studies that describe microbial community change across the course of a year as related to landscape features would be of value.

APPENDIX A

SUPPLEMENTARY INFORMATION FOR CHAPTER II

Table 4. Proportion of vegetated area calculated at varying buffer sizes.

Site	Proportion of Vegetated Area (within buffer zone radii)				
	50 m	100 m	200 m	400 m	800 m
park-ALT	0.95	0.93	0.83	0.67	0.5
park-AMA	0.97	0.93	0.77	0.59	0.46
park-MAU	0.98	0.83	0.63	0.51	0.39
park-WES	1	0.83	0.64	0.58	0.5
park-WEW	0.96	0.97	0.7	0.56	0.36
lot-ALB	0.01	0.15	0.35	0.51	0.5
lot-LOW	0.03	0.02	0.05	0.22	0.43
lot-MOC	0.01	0.04	0.09	0.29	0.52
lot-PPN	0.02	0.08	0.11	0.37	0.49
lot-VRC	0	0.17	0.26	0.38	0.41

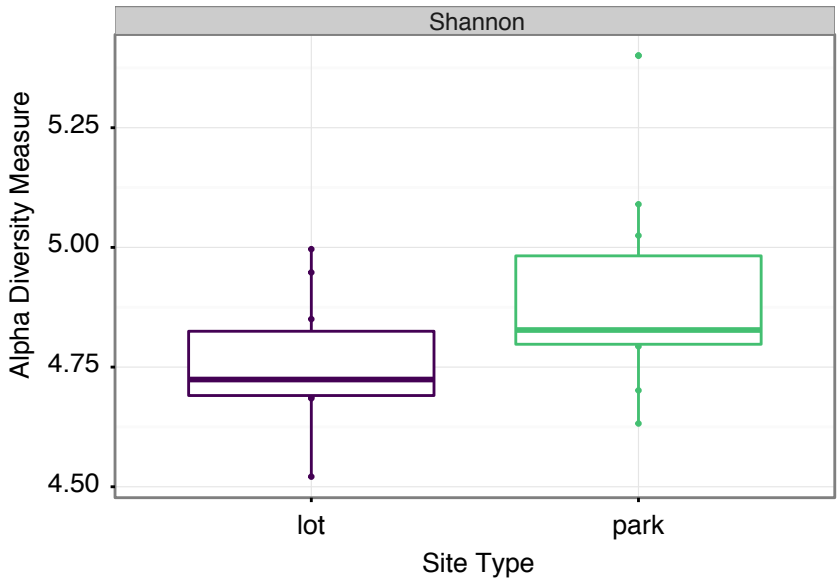


Figure 24. Comparison of alpha diversity for parking lots and parks.

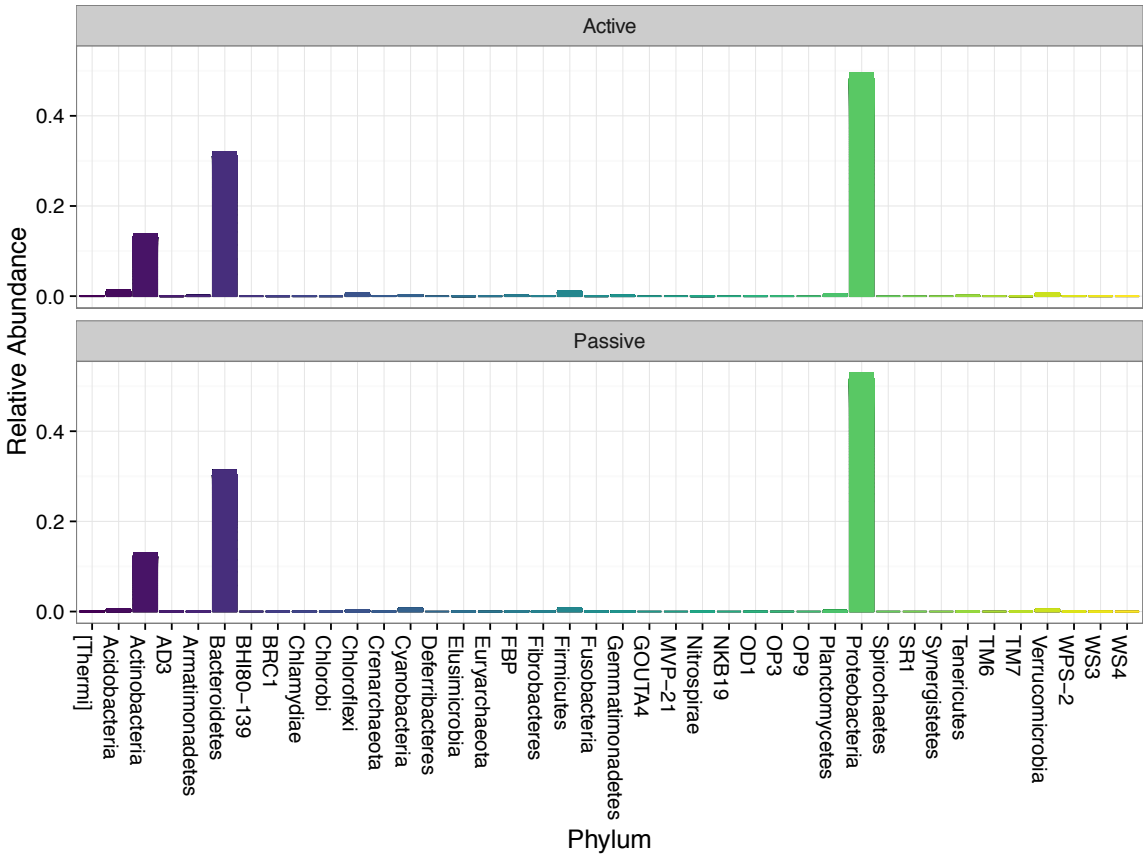


Figure 25. Airborne bacterial community composition at the phylum level for active (top) versus passive (bottom) sampling methods

Table 5. Change in statistical prediction value and significance of PERMANOVA model at varying buffer zone radii.

Buffer Zone Radii	**R^2**	***p*-value**
50 m	0.150	0.026
100 m	0.143	0.060
200 m	0.151	0.068
400 m	0.127	0.240
800 m	0.108	0.420

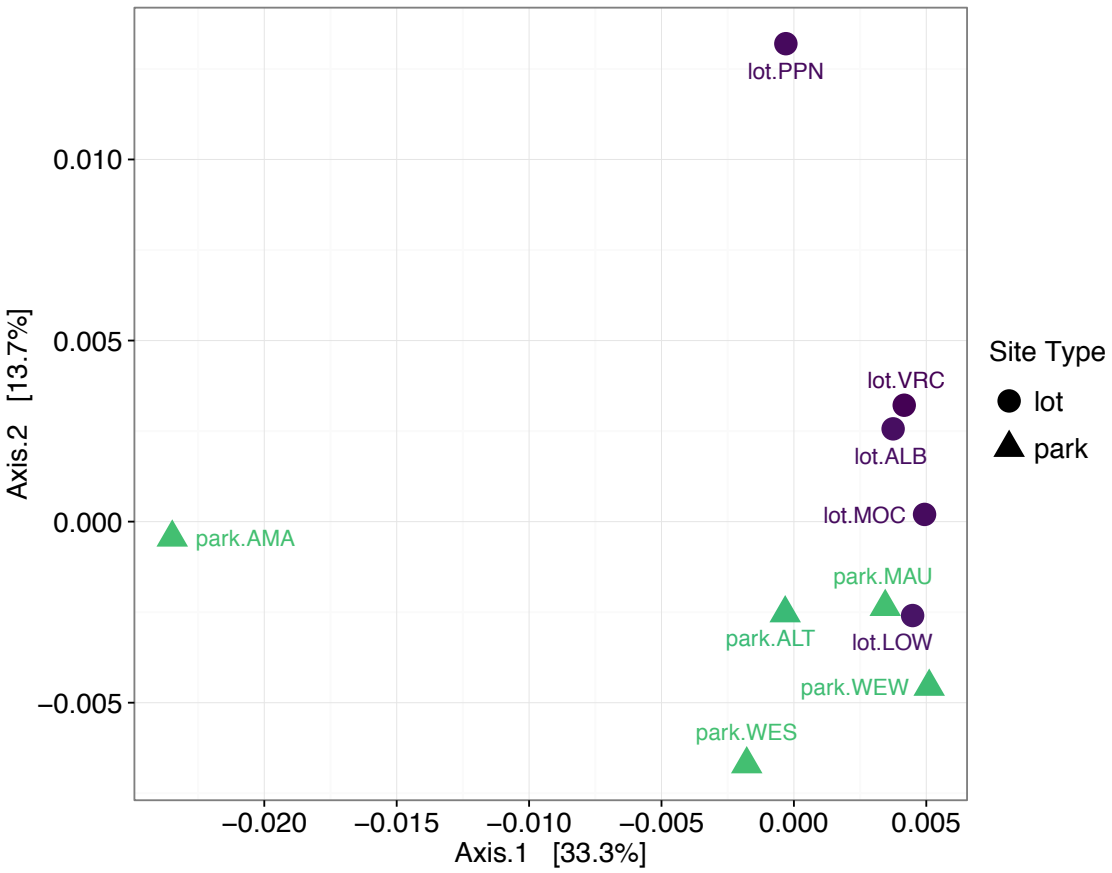


Figure 26. Unconstrained PCoA ordination of airborne bacterial communities from parks (green triangles) and parking lots (purple circles).

Table 6. Closest NCBI matches, accession numbers and match % for all OTUs that were identified as differentially abundant in parks or parking lots.

Closest NCBI Match	Accession Number	Match %
Aerosakkonema funiforme	NR_114306.1	93
Aridibacter kavangonensis	NR_133698.1	87
Acidisoma sibiricum	NR_042706.1	95
Lactobacillus jensenii	NR_117072.1	81
Nostoc punctiforme	NR_074317.1	99
Calothrix desertica	NR_114995.1	94
Hymenobacter ginsengisoli	NR_109449.1	96
Sediminicoccus rosea	NR_132670.1	95
Calothrix sp. PCC 7507	NR_102891.1	98
Spirosoma arcticum	NR_134186.1	96
Gemmatirosa kalamazoonesis	NR_132675.1	89
Acidisoma sibiricum	NR_042706.1	94
Acidisoma tundrae	NR_042705.1	94
Frondihabitans peucedani	NR_116933.1	94
Georgenia soli	NR_116959.1	97
Oxynema thaianum	NR_125585.1	93
Geodermatophilus nigrescens	NR_109505.1	97
Acidisoma tundrae	NR_042705.1	94
Frondihabitans peucedani	NR_116933.1	94
Georgenia soli	NR_116959.1	97
Oxynema thaianum	NR_125585.1	93
Geodermatophilus nigrescens	NR_109505.1	97
Sphingomonas jaspsi	NR_114034.1	98
Komagataeibacter oboediens	NR_113397.1	95
Telmatobacter bradus	NR_115074.1	97
Roseiarcus fermentans	NR_134158.1	98
Mucilaginibacter frigoritolerans	NR_116979.1	100
Granulicella paludicola	NR_115072.1	96
Steroidobacter denitrificans	NR_044309.1	94
Burkholderia sordidicola	NR_041916.1	100
Micromonospora cremea	NR_108478.1	99
Rhodopila globiformis	NR_037120.1	96
Burkholderia phytofirmans	NR_102845.1	100
Granulicella sapmiensis	NR_118023.1	99
Blastococcus jejuensis	NR_043633.1	96

Table 6. continued

Closest NCBI Match	Accession Number	Match %
Beijerinckia mobilis	NR_042180.1	96
Granulicella pectinivorans	NR_115071.1	97
Novosphingobium taihuense	NR_042934.1	99
Burkholderia phenoliruptrix	NR_102849.1	97
Silvibacterium bohemicum	NR_135209.1	97
Rubritalea halochordaticola	NR_113049.1	88
Acidicapsa borealis	NR_117182.1	98
Granulicella pectinivorans	NR_115071.1	97
Acidicaldus organivorans	NR_042752.1	94
Nguyenibacter vanlangensis	NR_125459.1	95
Granulibacter bethesdensis	NR_074276.1	95
Candidatus Solibacter usitatus	NR_074351.1	93
Silvibacterium bohemicum	NR_135209.1	98
Granulicella tundricola	NR_074295.1	98
Silvibacterium bohemicum	NR_135209.1	97
Rhizobium sullae	NR_029330.1	99
Conexibacter woesei	NR_074830.1	91
Rhodopila globiformis	NR_037120.1	96
Rhodococcus coprophilus	NR_118607.1	100
Conexibacter woesei	NR_074830.1	98
Blastochloris gulmargensis	NR_115056.1	96
Conexibacter woesei	NR_074830.1	96
Silvibacterium bohemicum	NR_135209.1	94
Rhodopila globiformis	NR_037120.1	95
Granulicella cerasi	NR_134047.1	97
Acidocella aluminiidurans	NR_114266.1	94
Granulicella pectinivorans	NR_115071.1	98
Erwinia billingiae	NR_102820.1	100
Acidobacterium capsulatum	NR_074106.1	97
Acidipila rosea	NR_113179.1	99
Brevundimonas olei	NR_117268.1	96
Enterococcus hirae	NR_114783.2	100

APPENDIX B

R ANALYSIS CODE FOR CHAPTER II

```
`r opts_chunk$set(cache=FALSE, tidy=FALSE, comment='')`

# Airborne Microbiome of Parks vs. Parking Lots: Pilot Study

#### G. Mhuireach (gwynhwyfer.mhuireach at gmail dot com)

---
title: "ParksLots_petris_SotTE"
author: "G. Mhuireach"
date: "February 1, 2016"
output: html_document
---

This document provides the code used to generate all analyses and
figures for our pilot study of airborne microbial communities in parks
and parking lots in Eugene, Oregon. A manuscript has been submitted for
publication and is currently undergoing peer review.

First, prepare the workspace by setting parameters, defining the working
directory, and loading required packages.

```{r initialSetup, echo=FALSE}
set.seed(2) # set seed for reproducible results
options(scipen=7) # curtail scientific notation
options(digits=5) # number of digits to print on output
# The required package list:
reqpkg <- c("phyloseq", "ggplot2", "DESeq2", "ape", "RColorBrewer",
"vegan",
      "rgdal", "maptools", "spdep", "gstat", "VennDiagram", "reshape2")
# Load all required packages and show version
for (i in reqpkg) {
  print(i)
  print(packageVersion(i))
  library(i, quietly = TRUE, verbose = FALSE, warn.conflicts = FALSE,
character.only = TRUE)
}

# set ggplot2 theme
theme_set(theme_bw(base_size = 15))

## load extra functions
# function to extract OTU table from phyloseq and ensure samples are
rows
getTab <- function(physeq) {
```
```

```
 require("vegan")
 OTU = otu_table(physeq)
 if (taxa_are_rows(OTU)) {
  OTU = t(OTU)
 }
 return(as(OTU, "matrix"))
}

```

Load data (contains buttons & petris combined). Then clean up the data: identify potential contaminants from lab processing and remove them, remove extraction controls, remove taxa that occur only 1-2 times, identify and remove failed samples.

```{r importData, echo=FALSE}
# load data
load("~/Downloads/ParksLots.RData")

# identify potential lab contaminants
PL_ec1 <- prune_samples(sample_names(ParksLots) == "ec.1", ParksLots)
PL_ec2 <- prune_samples(sample_names(ParksLots) == "ec.2", ParksLots)
PL_ec3 <- prune_samples(sample_names(ParksLots) == "ec.3", ParksLots)
PL_ec4 <- prune_samples(sample_names(ParksLots) == "ec.4", ParksLots)
PL_contams <- merge_phyloseq(PL_ec1, PL_ec2)
PL_contams <- merge_phyloseq(PL_contams, PL_ec3)
PL_contams <- merge_phyloseq(PL_contams, PL_ec4)
PL_contams <- prune_taxa(taxa_sums(PL_contams) > 0, PL_contams)
contamAbund <- data.frame(tax_table(PL_contams))
contamAbund$Count <- taxa_sums(PL_contams)
#write.csv(contamAbund, file="contamAbund.csv")
# what are the 3 most abundant contaminants?
top3cont <- sort(taxa_sums(PL_contams), TRUE)[1:3]
top3cont <- prune_taxa(names(top3cont), PL_contams)
otu_table(top3cont)
tax_table(top3cont)

# remove extraction control samples
PL <- prune_samples(sample_names(ParksLots) != "ec.1", ParksLots)
PL <- prune_samples(sample_names(PL) != "ec.2", PL)
PL <- prune_samples(sample_names(PL) != "ec.3", PL)
PL <- prune_samples(sample_names(PL) != "ec.4", PL)

# see if any samples failed
sort(sample_sums(PL))
# 3 samples have very low counts - remove them
PL <- prune_samples(sample_sums(PL) >= 25000, PL)

# remove plant and mitochondrial sequences
get_taxa_unique(PL, "Order")
```

```
PL_nocont <- subset_taxa(PL, Order!="Streptophyta")
get_taxa_unique(PL, "Family")
PL_nocont <- subset_taxa(PL_nocont, Family!="mitochondria")

# remove top 3 potential contaminants, as well as Halomonas and
Shewanella spp.
taxa_sums(PL_nocont)["573035"]
PL_nocont <- prune_taxa(taxa_names(PL_nocont) != "573035", PL_nocont)
taxa_sums(PL_nocont)["3505032"]
PL_nocont <- prune_taxa(taxa_names(PL_nocont) != "3505032", PL_nocont)
taxa_sums(PL_nocont)["811074"]
PL_nocont <- prune_taxa(taxa_names(PL_nocont) != "811074", PL_nocont)
PL_nocont <- subset_taxa(PL_nocont, Genus!="Halomonas")
PL_nocont <- subset_taxa(PL_nocont, Genus!="Shewanella")

# remove taxa seen fewer than 3 times
PL_prune <- prune_taxa(taxa_sums(PL_nocont) > 3, PL_nocont)

```

Preprocess data and apply transformations for downstream analyses. There has been discussion about the respective benefits of "rarefying" (i.e., selecting a random subsample from each sample at the level of the lowest sample count to adjust for differences in sample sizes) versus data transformation, for example, the "variance stabilizing transformation" (VST) method in the DESeq2 package for R (cite). We report the results of the VST method in this paper.

```{r preProcess, echo=FALSE}

# make new variable to merge by sampling method for each location
var1 = as.character(get_variable(PL_prune, "sample_type"))
var2 = as.character(get_variable(PL_prune, "location"))
sample_data(PL_prune)$mergeVar <- mapply(paste0, var1, var2, collapse =
"_")
allMerge <- merge_samples(PL_prune, "mergeVar")
# repair values for location & site_type
sample_data(allMerge)$location <- levels(sample_data(PL_
prune)$location)
sample_data(allMerge)$site_type[sample_data(allMerge)$site_type==1] <-
"lot"
sample_data(allMerge)$site_type[sample_data(allMerge)$site_type==2] <-
"park"
sample_data(allMerge)$sample_type[sample_data(allMerge)$sample_type==1]
<- "button"
sample_data(allMerge)$sample_type[sample_data(allMerge)$sample_type==2]
<- "petri"
sample_data(allMerge) # check to make sure it worked correctly

# subset for petris (passive)
```

```
PL_petri <- subset_samples(PL_prune, sample_type == “petri”)
# are there any taxa not observed in any lot samples?
any(taxa_sums(PL_petri) == 0)
sum(taxa_sums(PL_petri) == 0) # how many?
PL_petri <- prune_taxa(taxa_sums(PL_petri) > 0, PL_petri) # remove them

# merge all 3 petri samples for each site
petMerge <- merge_samples(PL_petri, “location”)
# repair values for location & site_type
sample_data(petMerge)$location <- levels(sample_data(PL_
petri)$location)
sample_data(petMerge)$site_type[sample_data(petMerge)$site_type==1] <-
“lot”
sample_data(petMerge)$site_type[sample_data(petMerge)$site_type==2] <-
“park”
sample_data(petMerge) # check to make sure it worked correctly

# add a column for site labels
sample_data(petMerge)$site_label <-
 c(“lot.ALB”, “park.ALT”, “park.AMA”, “lot.LOW”, “park.MAU”, “lot.MOC”,
  “lot.PPN”, “lot.VRC”, “park.WES”, “park.WEW”)

# transform merged petri counts for downstream analysis using vst
petMerge_des2 <- phyloseq_to_deseq2(petMerge, ~ site_type)
petMerge_des2 <- estimateSizeFactors(petMerge_des2)
petMerge_des2 <- estimateDispersions(petMerge_des2, fitType=”local”,
maxit=260)
# perform DESeq2 variance stabilization instead of rarefying
petMerge_vst <- getVarianceStabilizedData(petMerge_des2)
# Save the untransformed data as a separate variable so you can go back
to it
petMerge0 <- petMerge
otu_table(petMerge) <- otu_table(petMerge_vst, taxa_are_rows=TRUE)
petMerge_vst <- petMerge
petMerge <- petMerge0
rm(petMerge0)
# Set values below zero, to zero
otu_table(petMerge_vst)[otu_table(petMerge_vst) < 0.0] <- 0
sort(taxa_sums(petMerge))[1:10]

# transform merged petri counts to relative abundance by sample
petMergeRel <- transform_sample_counts(petMerge, function(x) 100 * x/
sum(x))

# get raw OTU and sample data tables out of phyloseq, samples as rows
petTab <- getTab(petMerge)
petTab[1:5, 1:5] # check

```

Access basic information, perform initial visualization of
untransformed data.

```{r basicInfo, echo=FALSE}
sum(sample_sums(allMerge)) # total number of OTUs
ntaxa(allMerge) # total number of taxa
length(get_taxa_unique(allMerge, "Phylum")) # number of unique Phyla
siteSums <- sort(sample_sums(allMerge)) # numbers of OTUs found in each
sample
sum(siteSums["buttonmauriejacobs"], siteSums["petrimauriejacobs"])
sum(siteSums["buttonwestmoreland"], siteSums["petriwestmoreland"])
sum(siteSums["buttonamazon"], siteSums["petriamazon"])
sum(siteSums["buttonppnw"], siteSums["petrippnw"])
sum(siteSums["buttonaltonbaker"], siteSums["petrialtonbaker"])
sum(siteSums["buttonalbertsons"], siteSums["petrialbertsons"])
sum(siteSums["buttonlowes"], siteSums["petrimauriejacobs"])
sum(siteSums["buttonmoc"], siteSums["petrimoc"])
sum(siteSums["buttonwew"], siteSums["petriwew"])
sum(siteSums["buttonlowes"], siteSums["petrivrc"])
rev(sort(taxa_sums(allMerge)))[1:10] # range of top ten OTU abundances
rev(sort(taxa_sums(allMerge)))[1]/sum(sample_sums(allMerge)) #
proportion top 1

# top 10 taxa
pl10 <- sort(taxa_sums(allMerge), TRUE)[1:10]
pl10 <- prune_taxa(names(pl10), allMerge)
pl10_sums <- data.frame(taxa_sums(pl10))

# get taxonomic Families of top 10
pl10tax <- data.frame(tax_table(pl10Rel)[, c("Phylum", "Class",
"Order", "Family", "Genus",
               "Species")])
pl10tax <- cbind(pl10tax, pl10_sums)
names(pl10tax)[names(pl10tax) == "taxa_sums.pl10."] <- "Abundance"
pl10matrix$RelAbund <- 100 * pl10matrix$Abundance/sum(taxa_
sums(allMerge))
pl10matrix <- pl10matrix[order(-pl10matrix[,8]), ]

# compare shared taxa across site type
allLot <- subset_samples(allMerge, site_type == "lot")
# are there any taxa not observed in any lot samples?
any(taxa_sums(allLot) == 0)
sum(taxa_sums(allLot) == 0) # how many?
allLot <- prune_taxa(taxa_sums(allLot) > 0, allLot) # remove them

# Subset petri-parks
allPark <- subset_samples(allMerge, site_type == "park")
# are there any taxa not observed in any park samples?
any(taxa_sums(allPark) == 0)
sum(taxa_sums(allPark) == 0) # how many?
```

```
allPark <- prune_taxa(taxa_sums(allPark) > 0, allPark) # remove them

# Venn diagrams for lots vs. parks (use to calculate number of shared
taxa)
parkNames <- as.vector(taxa_names(allPark))
lotNames <- as.vector(taxa_names(allLot))
venn.diagram(x=list("Parks"=parkNames, "Lots"=lotNames),
       filename="Venn-allpl.tiff", col="transparent",
       fill=c("seagreen", "gray"), alpha=0.50,
       cex=1.5, fontfamily="serif", fontface="bold",
       cat.col=c("seagreen4", "gray3"), cat.cex=1.5,
       cat.pos=c(270, 90), cat.dist=0.2, cat.fontfamily="serif",
       rotation.degree=0, margin=0.2)

# plot alpha diversity index
plot_richness(allMerge, x="site_type", color="site_type",
measures=c("Shannon"),
       title="Alpha Diversity") +
       geom_boxplot() +
       scale_color_viridis(begin=0, end=0.7, discrete=TRUE) +
 theme(axis.text.x=element_text(size=16), axis.text.y=element_
text(size=16),
    axis.title=element_text(size=18), legend.position = "none"
    )

# test diversity across sites
# first check that the data are normally distributed
sample_data(allMerge)$Shannon <- estimate_richness(allMerge)[,
"Shannon"]
hist(sample_data(allMerge)$Shannon)
t.test(sample_data(allMerge)$Shannon ~ sample_data(allMerge)$site_type)

```

Compare passive and active collection methods in terms of species richness, overall composition, and relative abundances of the taxa comprising the top X percent of sequences.

```{r passAct, echo=FALSE}

plot_richness(allMerge, x="sample_type", color="sample_type",
       measures=c("Shannon"), title="") +
       geom_boxplot() +
       scale_color_viridis(begin=0, end=0.7, discrete=TRUE,
guide=FALSE)

# test richness of petri vs. button samples
# first check that the data are normally distributed
sample_data(allMerge)$Shannon <- estimate_richness(allMerge)[,
"Shannon"]
```

```
hist(sample_data(allMerge)$Shannon)
t.test(sample_data(allMerge)$Shannon ~ sample_data(allMerge)$sample_
type)

# barplots to compare proportional composition for each sample type by
phylum
sampMerge <- merge_samples(PL_prune, “sample_type”)
# repair values for location & site_type
sample_data(sampMerge)$sample_type <- levels(sample_data(PL_
prune)$sample_type)

sampMergeRel <- transform_sample_counts(sampMerge, function(x) x/
sum(x))
otu_table(sampMergeRel)[1:2, 1:10]
TaxobarAll_Phylum <- plot_bar(sampMergeRel, x=”Family”, fill=”Family”)
TaxobarAll_Phylum +
 geom_bar(aes(color=Family, fill=Family, legend=””), stat=”identity”) +
 facet_wrap(~ sample_type, ncol=1) +
 scale_color_viridis(discrete=TRUE, guide=FALSE) +
 scale_fill_viridis(discrete=TRUE, guide=FALSE)

# get relative abundance for top 50 taxa
sum(sort(taxa_sums(allMerge),TRUE)[1:50])/sum(taxa_sums(allMerge))
top50 <- sort(taxa_sums(allMerge), TRUE)[1:50]
top50 <- prune_taxa(names(top50), allMerge)
top50Tab <- getTab(top50)
top50Env <- data.frame(sample_data(top50))
top50res <- adonis(top50Tab ~ sample_type, data=top50Env, perm=99999,
method=”horn”)
top50res
Taxobar_top50 <- plot_bar(top50, x=”Genus”, fill=”Genus”)
Taxobar_top50 +
 geom_bar(aes(fill=Genus), stat=”identity”) +
 facet_wrap(~ sample_type, ncol=1) +
 scale_fill_viridis(discrete=TRUE)
```

Explore and visualize data:

```{r dataExplore, echo=FALSE}
## ordinations
# using variance-stabilizing transformed data and Horn-Morisita
distance metric
petPcoa <- ordinate(petMerge_vst, method=”PCoA”, distance=”horn”)
plot_scree(petPcoa)
require(ggrepel)
plot_ordination(petMerge_vst, petPcoa, type=”samples”, shape=”site_
type”,
        color=”Green_50m”) +
```

```
 geom_point(size=6) +
 geom_text_repel(aes(label=sample_data(petMerge_vst)$site_label,
size=18),
         box.padding=unit(0.6, “lines”)) +
 scale_color_viridis(begin=0, end=.7) +
 theme(legend.text=element_text(size=16),
    legend.title=element_text(size=16, vjust=1),
    axis.text=element_text(size=16),
    axis.title=element_text(size=16)
    )

# PERMANOVA for site type, veg. proportion within 50m, 100m, 200m,
400m, 800m
petMergeTab <- getTab(petMerge_vst)
petMergeEnv <- data.frame(sample_data(petMerge_vst))

varlist <- names(petMergeEnv)[c(6,13,15,17,19,21)]
results <- list()
for (i in varlist){
      form <- as.formula(paste(“petMergeTab”, i, sep=”~”))
      results[[i]] <- adonis(form, data=petMergeEnv, perm=99999,
method=”horn”)
      }
sink(“adonisResults.txt”)
results
sink()

# show relationship between buffer size and adonis results
buffDist <- c(50, 100, 200, 400, 800)
resR2 <- c(results$Green_50m$aov.tab$R2[1], results$Green_100m$aov.
tab$R2[1],
      results$Green_200m$aov.tab$R2[1], results$Green_400m$aov.
tab$R2[1],
      results$Green_800m$aov.tab$R2[1])
resPr <- c(results$Green_50m$aov.tab$Pr[1], results$Green_100m$aov.
tab$Pr[1],
      results$Green_200m$aov.tab$Pr[1], results$Green_400m$aov.
tab$Pr[1],
      results$Green_800m$aov.tab$Pr[1])
buffAdon <- data.frame(cbind(buffDist, resR2, resPr))
p <- ggplot(buffAdon, aes(x=buffDist, y=resR2)) + geom_point()
p + geom_smooth(method = “lm”, se = FALSE)

p <- ggplot(buffAdon, aes(x=buffDist, y=resPr)) + geom_point()
p + geom_smooth(method = “lm”, se = FALSE)

buffR2 <- lm(buffDist ~ resR2)
summary(buffR2)
buffPr <- lm(buffDist ~ resPr)
summary(buffPr)
```

```
# test spatial autocorrelation with Mantel test
pet.spat <- as.matrix(dist(cbind(petMergeEnv$Longitude,
petMergeEnv$Latitude)))
pet.spat
pet.horn <- as.matrix(vegdist(petMergeTab, method="horn"))
petMan <- mantel(pet.horn, pet.spat, method="spearman",
permutations=99999)
petMan

# test hypothesis with direct gradient analysis with constrained PCoA
ordcap <- ordinate(petMerge_vst, "CAP", "horn", ~ Green_50m)
plot_ordination(petMerge_vst, ordcap, type="samples", shape="site_
type",
  color="Green_50m") + geom_point(size=6) +
  geom_text_repel(aes(label=sample_data(petMerge_vst)$site_label,
size=18),
          box.padding=unit(0.6, "lines")) +
  scale_color_viridis(begin=0, end=0.7) +
  theme(
   legend.text=element_text(size=16),
   legend.title=element_text(size=16, vjust=1),
   axis.text=element_text(size=16),
   axis.title=element_text(size=16)
   )

```

Investigate abundant vs. common vs. rare taxa, as in "rare biosphere"
papers
(Logares et al. 2014). We used the raw counts, transformed to
proportional abundance by sample.

```{r rareAbund, echo=FALSE}
# write.csv(taxa_sums(petMerge), file = "taxa_sums.csv")
# define rare (< 0.1%) taxa, per Pedros Alio (2006)
petRareRel <- filter_taxa(petMergeRel, function(x) mean(x) < 0.1, TRUE)
# select rare taxa from untransformed OTU table
petRare <- prune_taxa(taxa_names(petRareRel), petMerge)
rev(sort(taxa_sums(petRare)))[1:10]
ntaxa(petRare)

# define abundant (> 0.1%) taxa
petAbundRel <- filter_taxa(petMergeRel, function(x) mean(x) > 0.1, TRUE)
# select abundant taxa from untransformed OTU table
petAbund <- prune_taxa(taxa_names(petAbundRel), petMerge)
ntaxa(petAbund)

# transform rare counts for downstream analysis using vst
```

```
petRare_des2 <- phyloseq_to_deseq2(petRare, ~ site_type)
# calculate geometric means prior to estimate size factors
gm_mean <- function(x, na.rm=TRUE){
 exp(sum(log(x[x > 0]), na.rm=na.rm) / length(x))
 }
geoMeans <- apply(counts(petRare_des2), 1, gm_mean)
petRare_des2 <- estimateSizeFactors(petRare_des2, geoMeans=geoMeans)
petRare_des2 <- estimateDispersions(petRare_des2, fitType="local",
maxit=260)
# perform DESeq2 variance stabilization instead of rarefying
petRare_vst <- getVarianceStabilizedData(petRare_des2)
# Save the untransformed data as a separate variable so you can go back
to it
petRare0 <- petRare
otu_table(petRare) <- otu_table(petRare_vst, taxa_are_rows=TRUE)
petRare_vst <- petRare
petRare <- petRare0
rm(petRare0)
# Set values below zero, to zero
otu_table(petRare_vst)[otu_table(petRare_vst) < 0.0] <- 0

# ordinate rare taxa
petRareOrd <- ordinate(petRare_vst, "PCoA", "horn")
ordcapRare <- ordinate(petRare, "CAP", "horn", ~ Green_50m)
plot_ordination(petRare, petRareOrd,
  type="samples", shape="site_type", color="Green_50m") +
  geom_point(size=5) + geom_text(label=sample_data(petRare)$location,
size=4, vjust=1.3) +
  scale_colour_gradientn(colours=rainbow(10, s=0.4, v=0.75, start=0,
end=.5, alpha=.8))

# test for statistical differences for rare taxa
petRareTab <- getTab(petRare)
petRareEnv <- data.frame(sample_data(petRare))
adonisRareSite <- adonis(petRareTab ~ site_type, petRareEnv,
perm=99999, method="horn")
adonisRareSite
adonisRareVeg <- adonis(petRareTab ~ Green_50m, petRareEnv, perm=99999,
method="horn")
adonisRareVeg

# plot abundant taxa
# transform petri counts for downstream analysis using vst
petAbund_des2 <- phyloseq_to_deseq2(petAbund, ~ site_type)
petAbund_des2 <- estimateSizeFactors(petAbund_des2)
petAbund_des2 <- estimateDispersions(petAbund_des2, fitType="local",
maxit=260)
# perform DESeq2 variance stabilization instead of rarefying
petAbund_vst <- getVarianceStabilizedData(petAbund_des2)
# Save the untransformed data as a separate variable so you can go back
```

```
to it
petAbund0 <- petAbund
otu_table(petAbund) <- otu_table(petAbund_vst, taxa_are_rows=TRUE)
petAbund_vst <- petAbund
petAbund <- petAbund0
rm(petAbund0)
# Set values below zero, to zero
otu_table(petAbund_vst)[otu_table(petAbund_vst) < 0.0] <- 0

petAbundOrd <- ordinate(petAbund, "PCoA", "horn")
plot_ordination(petAbund, petAbundOrd,
  type="samples", shape="site_type", color="Green_200m") +
  geom_point(size=5) + geom_text(label=sample_data(petAbund)$location,
size=4, vjust=1.3) +
  scale_colour_gradientn(colours=rainbow(10, s=0.4, v=0.75, start=0,
end=.5, alpha=.8))

petAbundTab <- getTab(petAbund_vst)
petAbundEnv <- data.frame(sample_data(petAbund_vst))
adonisAbuSite <- adonis(petAbundTab ~ site_type, petAbundEnv,
perm=99999, method="horn")
adonisAbuSite
adonisAbuVeg <- adonis(petAbundTab ~ Green_50m, petAbundEnv,
perm=99999, method="horn")
adonisAbuVeg

```
```

Do hypothesis testing

```{r hypTest, echo=FALSE}
# multiple test using negative binomial GLM fitting and Wald statistics
- petris
# levels are alphabetical (i.e., lots, parks) & "untreated" is first
petMerge_glm <- DESeq(petMerge_des2, test="Wald", fitType="parametric")
# Investigate test results table
res <- results(petMerge_glm, cooksCutoff=FALSE)
alpha <- 0.05
# The following omits the NA p-values.
sigtab <- res[which(res$padj < alpha), ]
# This line then reinstates the previous NA padj values if their
# log2FoldChange was greater than for p-value threshold.
#sigtab <- res[abs(res$log2FoldChange) >
min(abs(sigtab$log2FoldChange)), ]
sigtab <- cbind(as(sigtab, "data.frame"),
        as(tax_table(petMerge)[rownames(sigtab),], "matrix"),
        as(t(otu_table(petMergeRel)[,rownames(sigtab)]), "matrix"),
        as(taxa_sums(petMergeRel)[rownames(sigtab)], "matrix")
        )
colnames(sigtab)[24] <- c("RelAbund")
```
```

```
head(sigtab)
mcols(res, use.names=TRUE)
dim(sigtab)
sum(sigtab$RelAbund)
rev(sort(sigtab$RelAbund))[1:10]
# just the OTUs that were significantly higher in the parks
parksigtab <- sigtab[sigtab[, “log2FoldChange”] > 0, ]
parksigtab <- parksigtab[order(parksigtab[,6]),]
write.csv(as.data.frame(parksigtab), file=”pet_SigParks.csv”)
# just the OTUs that were significantly higher in the lots
lotsigtab <- sigtab[sigtab[, “log2FoldChange”] < 0, ]
lotsigtab <- lotsigtab[order(lotsigtab[,6]),]
head(lotsigtab)
write.csv(as.data.frame(lotsigtab), file=”pet_SigLots.csv”)

parkabundtaxa <- data.frame(row.names(parksigtab), parksigtab$Family,
parksigtab$wew,
              parksigtab$mauriejacobs, parksigtab$altonbaker,
              parksigtab$amazon, parksigtab$westmoreland,
              parksigtab$lowes, parksigtab$vrc, parksigtab$ppnw,
              parksigtab$moc, parksigtab$albertsons
              )
names(parkabundtaxa) <- c(“OTU”, “Family”, “park-WEW”, “park-MAU”,
“park-ALT”,
             “park-AMA”, “park-WES”, “lot-LOW”, “lot-VRC”, “lot-PPN”,
             “lot-MOC”, “lot-ALB”
             )
head(parkabundtaxa)
write.csv(as.data.frame(parkabundtaxa), file=”parkabundtaxa.csv”)
lotabundtaxa <- data.frame(row.names(lotsigtab), lotsigtab$Family,
lotsigtab$wew,
              lotsigtab$mauriejacobs, lotsigtab$altonbaker,
              lotsigtab$amazon, lotsigtab$westmoreland,
              lotsigtab$lowes, lotsigtab$vrc, lotsigtab$ppnw,
              lotsigtab$moc, lotsigtab$albertsons
              )
names(lotabundtaxa) <- c(“OTU”, “Family”, “park-WEW”, “park-MAU”,
“park-ALT”,
             “park-AMA”, “park-WES”, “lot-LOW”, “lot-VRC”, “lot-PPN”,
             “lot-MOC”, “lot-ALB”
             )
write.csv(as.data.frame(lotabundtaxa), file=”lotabundtaxa.csv”)

parkAbu <- prune_taxa(as.character(parkabundtaxa$OTU), petMergeRel)
parkAbu.df <- as.data.frame(t(otu_table(parkAbu)))
parkAbu.df$SiteAbu <- “park”
head(parkAbu.df)
parkBLAST <- read.csv(“~/Downloads/SigParksBLAST.csv”, header=TRUE)
head(parkBLAST)
parkAbu.df$BLAST <- parkBLAST[match(row.names(parkAbu.df),
```

```
parkBLAST$OTU), 6]
head(parkAbu.df)
length(parkAbu.df$BLAST)
write.csv(as.data.frame(parkAbu.df), file="parkAbu.df.csv")


lotAbu <- prune_taxa(as.character(lotabundtaxa$OTU), petMergeRel)
lotAbu.df <- as.data.frame(t(otu_table(lotAbu)))
lotAbu.df$SiteAbu <- "lot"
lotBLAST <- read.csv("~/Downloads/SigLotsBLAST.csv", header=TRUE)
lotAbu.df$BLAST <- lotBLAST[match(row.names(lotAbu.df), lotBLAST$OTU),
6]
head(lotAbu.df)
length(lotAbu.df$BLAST)
write.csv(as.data.frame(lotAbu.df), file="lotAbu.df.csv")

difAbund <- rbind(parkAbu.df, lotAbu.df)
difAbund$parkAbund <- difAbund$altonbaker + difAbund$amazon +
 difAbund$mauriejacobs + difAbund$westmoreland + difAbund$wew
difAbund$lotAbund <- difAbund$albertsons + difAbund$lowes +
 difAbund$moc + difAbund$ppnw + difAbund$vrc
difAbund$log2FoldChange <- sigplot[match(row.names(difAbund), row.
names(sigplot)), 2]
difAbund <- difAbund[order(-difAbund[,15]),]
head(difAbund)
tail(difAbund)

# create positive-negative barplot of differentially abundant taxa
sigplot <- data.frame(row.names(sigtab), sigtab$log2FoldChange,
sigtab$Family)
sigplot$SiteType <- "park"
sigplot$SiteType[sigplot$log2FoldChange < 0] <- "lot"
names(sigplot) <- c("OTU", "log2FoldChange", "Family", "SiteType")
sigplot$BLAST <- difAbund[match(row.names(sigplot), row.
names(difAbund)), 12]
sigplot$parkAbund <- difAbund[match(row.names(sigplot), row.
names(difAbund)), 13]
sigplot$lotAbund <- -difAbund[match(row.names(sigplot), row.
names(difAbund)), 14]
head(sigplot)
#write.csv(as.data.frame(sigplot), file="sigplot.csv")
sigplot <- sigplot[order(-sigplot[,2]),]
str(sigplot)
head(sigplot)
tail(sigplot)
write.csv(as.data.frame(sigplot), file="sigplot_reorder.csv")

pieColors <- colors()[c(384,83,656,655,652,  497,494,496, #park colors
            86,89,259,47,51,514,517,139,448,258,256,
            576,574,472,10,635,429,109,71,68,114,121,125,617,131,
```

```
            565,491,597,435,471,468,548,99,463,452,459,

            372,376,33,133,555,553,630,61,569,506, #lot colors
            57,503,587,585,54,56,621,92,94,90,
            498,76,78)]
ggplot(sigplot, aes(x = reorder(OTU, -log2FoldChange), y =
log2FoldChange,
          fill = log2FoldChange, legend="")) +
    geom_bar(stat = "identity",
    fill=pieColors, position = "identity") +
    coord_flip() +
    scale_x_discrete(labels=sigplot$BLAST) +
  theme(legend.position = "none", axis.text=element_text(size=10),
   axis.title=element_text(size=16)
   )

# create abundance plot for differentially abundant taxa
ggplot(data=sigplot, aes(fill=log2FoldChange)) +
  geom_bar(data=sigplot, aes(x=reorder(OTU, -log2FoldChange),
y=parkAbund, fill=pieColors, legend=""), stat="identity") +
  geom_bar(data=sigplot, aes(x=reorder(OTU, -log2FoldChange),
y=lotAbund, fill=pieColors, legend=""), stat="identity") +
  coord_flip() +
  scale_x_discrete(labels=sigplot$BLAST) +
  theme(legend.position = "none", axis.text=element_text(size=10),
   axis.title=element_text(size=16)
   )


# make pie charts for each site that include only the differentially
abundant taxa
# without taxa labels
pie(difAbund$altonbaker, col=pieColors, labels="")
pie(difAbund$albertsons, col=pieColors, labels="")
pie(difAbund$amazon, col=pieColors, labels="")
pie(difAbund$lowes, col=pieColors, labels="")
pie(difAbund$mauriejacobs, col=pieColors, labels="")
pie(difAbund$moc, col=pieColors, labels="")
pie(difAbund$ppnw, col=pieColors, labels="")
pie(difAbund$vrc, col=pieColors, labels="")
pie(difAbund$westmoreland, col=pieColors, labels="")
pie(difAbund$wew, col=pieColors, labels="")


# legend only
png(filename="pieLeg.png", width=4, height=10, units="in", pointsize=10,
   bg="white", res=150)
par(mar=c(0,0,0,0))
plot.new()
legend("center", as.character(difAbund$BLAST), cex=0.8,
```

```
fill=rainbow(length(difAbund$altonbaker)))
```

```
```

APPENDIX C

SUPPLEMENTARY INFORMATION FOR CHAPTER III

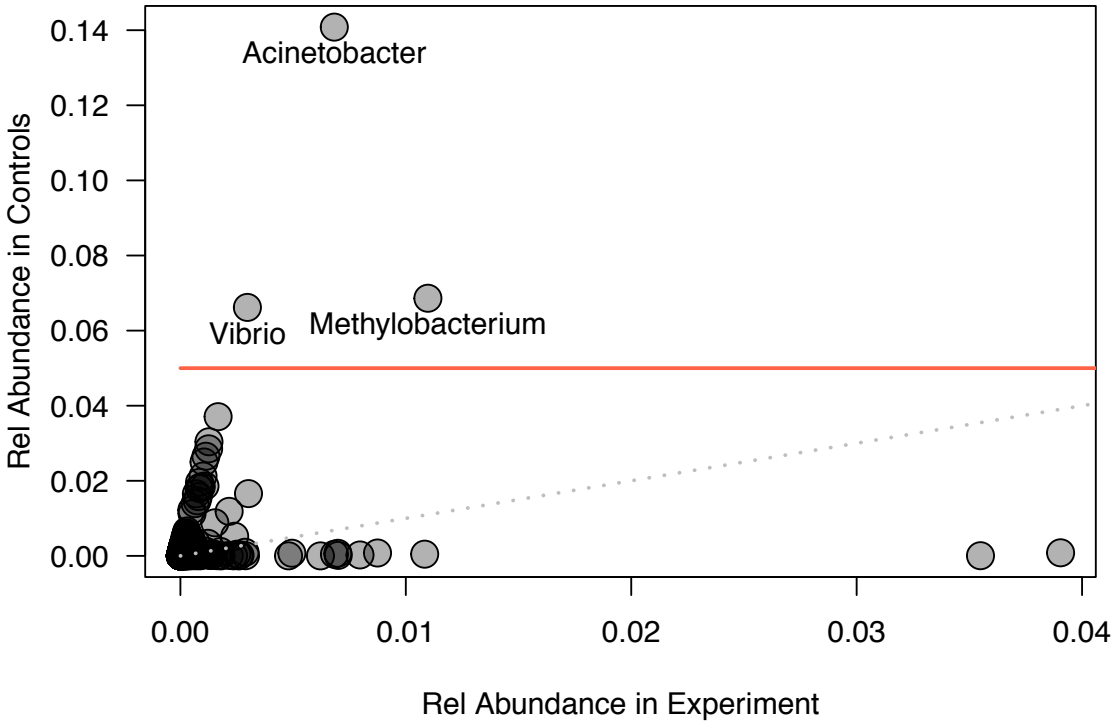


Figure 27. Abundance of potential contaminants in experiment samples (x-axis) versus control samples (y-axis).

Table 7. Breakdown of sampling groups by site location, land cover type, and date.

Land Cover Type	Group 1 Sites			Group 2 Sites		
	Sampling dates: 8.04, 8.18, 9.01, 9.22			Sampling dates: 8.14, 8.25, 9.15, 9.29		
	Code	Lat.	Long.	Code	Lat.	Long.
Forest	ABF*	44.053	-123.07	AMA	44.030	-123.09
	BRA*	44.092	-123.15	BLA	44.000	-123.10
	JAS	44.036	-122.98	DEL	44.090	-123.09
	MEL*	44.031	-123.13	HAR	44.077	-123.01
	SKI*	44.059	-123.09	OAK	44.046	-123.14
	WEF	44.053	-123.15	WHT	44.039	-123.02
Grass	LAU*[1]	44.026	-123.07	ABG	44.049	-123.05
	TAF*	44.057	-123.14	CEM[2]	44.011	-123.10
	TRA	44.069	-123.14	JAM	44.040	-123.03
	TSF	44.069	-122.99	LIV	44.058	-122.92
	WES*	44.037	-123.11	MAU[3]	44.066	-123.11
	WSC	44.042	-122.98	WEG	44.053	-123.15
Paved	BIM*	44.040	-123.12	FAI	44.044	-123.11
	GRA*	44.050	-123.13	MPR	44.059	-123.12
	MIC*	44.082	-123.04	PET	44.035	-123.04
	ORG	44.092	-123.16	STR	44.063	-123.03
	SAF*	44.011	-123.09	USF	44.065	-122.98
	SFC	44.071	-123.02	WIL	44.058	-122.99

* Samples also collected 7.14 for pilot testing

[1] Golf course (highly maintained—mowed, fertilized, irrigated)

[2] Cemetery (highly maintained—mowed, irrigated)

[3] Public park (highly maintained—mowed, irrigated)

Table 8. Land cover type, vegetation cover proportion, and structural diversity for each site.

Cover Type	Site Location	Cover (50 m)	Cover (100 m)	Cover (200 m)	Cover (400 m)	Cover (800 m)	Structural Diversity
Forest	ABF	0.98	0.98	0.93	0.68	0.58	2.59
	AMA	1.00	0.96	0.81	0.65	0.58	2.33
	BLA	1.00	1.00	0.96	0.87	0.85	2.79
	BRA	1.00	0.79	0.71	0.65	0.53	2.86
	DEL	1.00	0.83	0.57	0.53	0.54	2.47
	HAR	1.00	0.93	0.77	0.86	0.79	2.94
	JAS	0.82	0.82	0.73	0.79	0.66	2.61
	MEL	1.00	1.00	0.74	0.69	0.69	3.16
	OAK	1.00	0.89	0.57	0.48	0.40	2.55
	SKI	0.80	0.64	0.52	0.49	0.53	3.35
	WEF	1.00	0.98	0.90	0.78	0.53	2.35
	WHT	1.00	0.92	0.83	0.69	0.59	2.74
Grass	ABG	0.99	0.99	0.98	0.77	0.60	0.58
	CEM	0.97	0.92	0.90	0.80	0.73	0.51
	JAM	0.93	0.78	0.56	0.47	0.51	1.28
	LAU	0.97	0.98	0.93	0.83	0.76	1.41
	LIV	1.00	0.99	0.94	0.90	0.82	0.32
	MAU	1.00	0.94	0.74	0.53	0.43	0.58
	TAF	1.00	0.93	0.73	0.36	0.30	0.69
	TRA	0.92	0.78	0.53	0.38	0.37	0.69
	TSF	1.00	0.79	0.71	0.65	0.53	0.22
	WEG	0.87	0.90	0.74	0.55	0.40	1.03
	WES	0.86	0.82	0.73	0.61	0.56	0.70
	WSC	0.31	0.19	0.23	0.36	0.40	0.69
Paved	BIM	0.02	0.16	0.40	0.55	0.55	N/A
	FAI	0.00	0.09	0.20	0.47	0.54	N/A
	GRA	0.02	0.01	0.03	0.06	0.24	N/A
	MIC	0.11	0.10	0.18	0.25	0.39	N/A
	MPR	0.12	0.10	0.18	0.19	0.21	N/A
	ORG	0.04	0.04	0.16	0.29	0.44	N/A
	PET	0.17	0.26	0.48	0.59	0.68	N/A
	SAF	0.07	0.26	0.50	0.64	0.70	N/A
	SFC	0.34	0.34	0.43	0.47	0.58	N/A
	STR	0.12	0.09	0.20	0.38	0.47	N/A
	USF	0.27	0.26	0.34	0.46	0.53	N/A
	WIL	0.09	0.22	0.25	0.26	0.38	N/A

Table 9. Temperature measurements by sampling date and land cover type.

Date	Forest			Grass			Paved		
	Avg. Temp. (°C)	Min. Temp. (°C)	Max. Temp. (°C)	Avg. Temp. (°C)	Min. Temp. (°C)	Max. Temp. (°C)	Avg. Temp. (°C)	Min. Temp. (°C)	Max. Temp. (°C)
7.14.2015	21.0	13.0	34.0	22.5	10.5	36.0	23.4	12.5	37.5
8.04.2015	21.8	14.0	32.0	23.1	9.0	36.0	23.7	12.5	41.5
8.14.2015	20.7	13.5	37.5	21.8	12.0	38.5	21.9	14.0	36.0
8.18.2015	21.8	11.5	37.5	22.4	9.0	35.5	23.6	11.0	41.0
8.25.2015	19.3	9.0	34.5	19.9	5.5	33.0	20.9	8.0	38.0
9.01.2015	19.6	12.5	31.5	20.8	11.5	31.0	21.9	12.5	36.0
9.15.2015	13.3	5.5	26.5	13.1	4.0	27.5	13.8	5.0	26.0
9.22.2015	15.7	7.0	29.5	16.2	4.5	27.5	17.9	6.0	36.5
9.29.2015	14.8	6.5	26.0	14.5	3.5	30.5	16.4	6.0	36.5
Average	18.7	10.3	32.1	19.4	7.7	32.8	20.4	9.7	36.6

Table 10. Wind measurements by sampling date.

Date	Avg. Wind (mph)	Max. Wind (mph)	North	East	South	West
7.14.2015	2.2	5.7	71%	0%	5%	24%
8.04.2015	2.6	5.7	67%	5%	19%	10%
8.14.2015	3.3	7.2	4%	0%	57%	39%
8.18.2015	4.0	7.7	88%	8%	0%	4%
8.25.2015	2.1	4.6	83%	0%	6%	11%
9.01.2015	2.5	4.6	21%	0%	54%	25%
9.15.2015	3.7	7.2	52%	3%	41%	3%
9.22.2015	5.0	9.3	97%	0%	0%	3%
9.29.2015	2.5	8.2	44%	0%	50%	6%

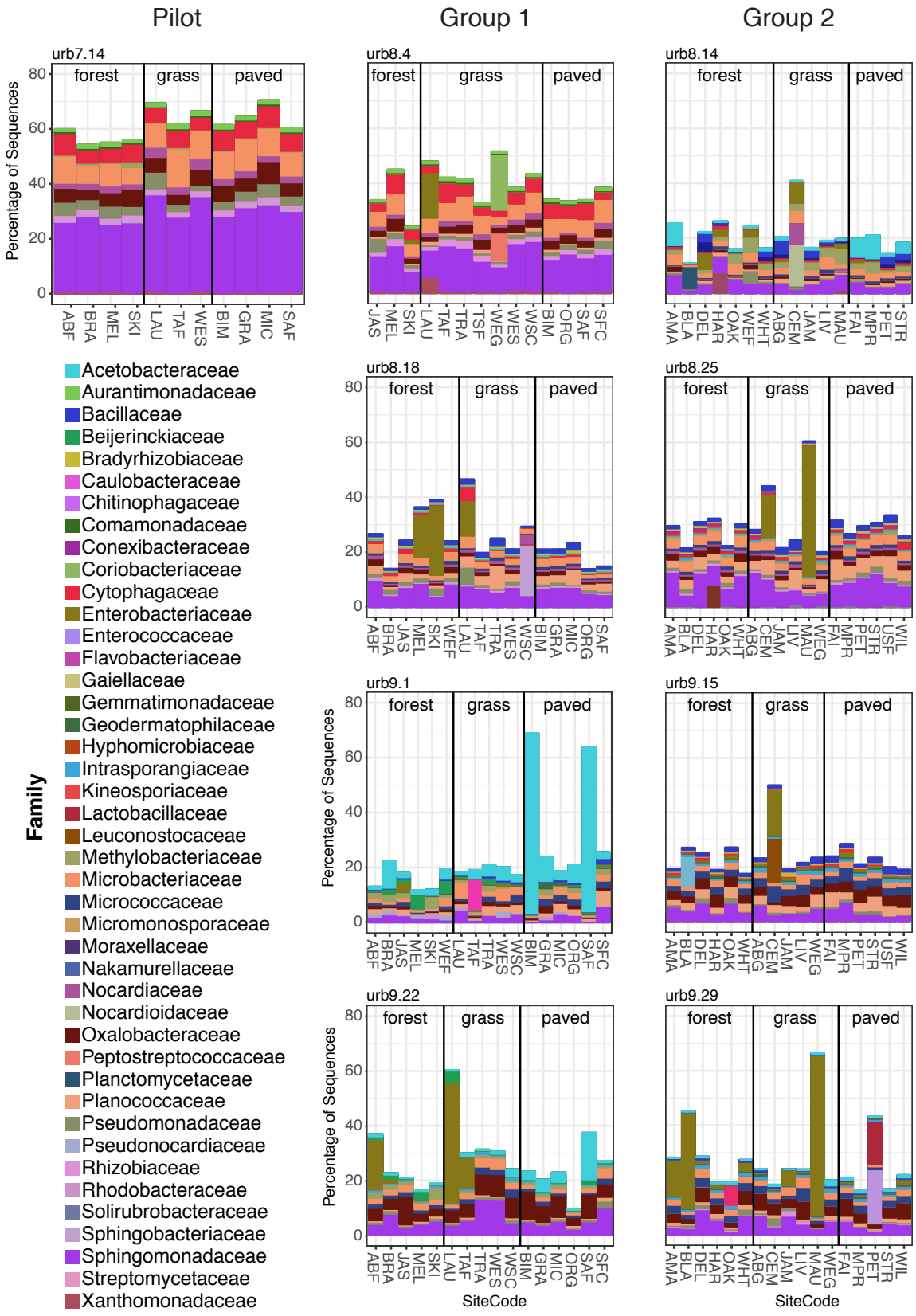


Figure 28. Top 25 most abundant taxa on each sampling date, colored by family.

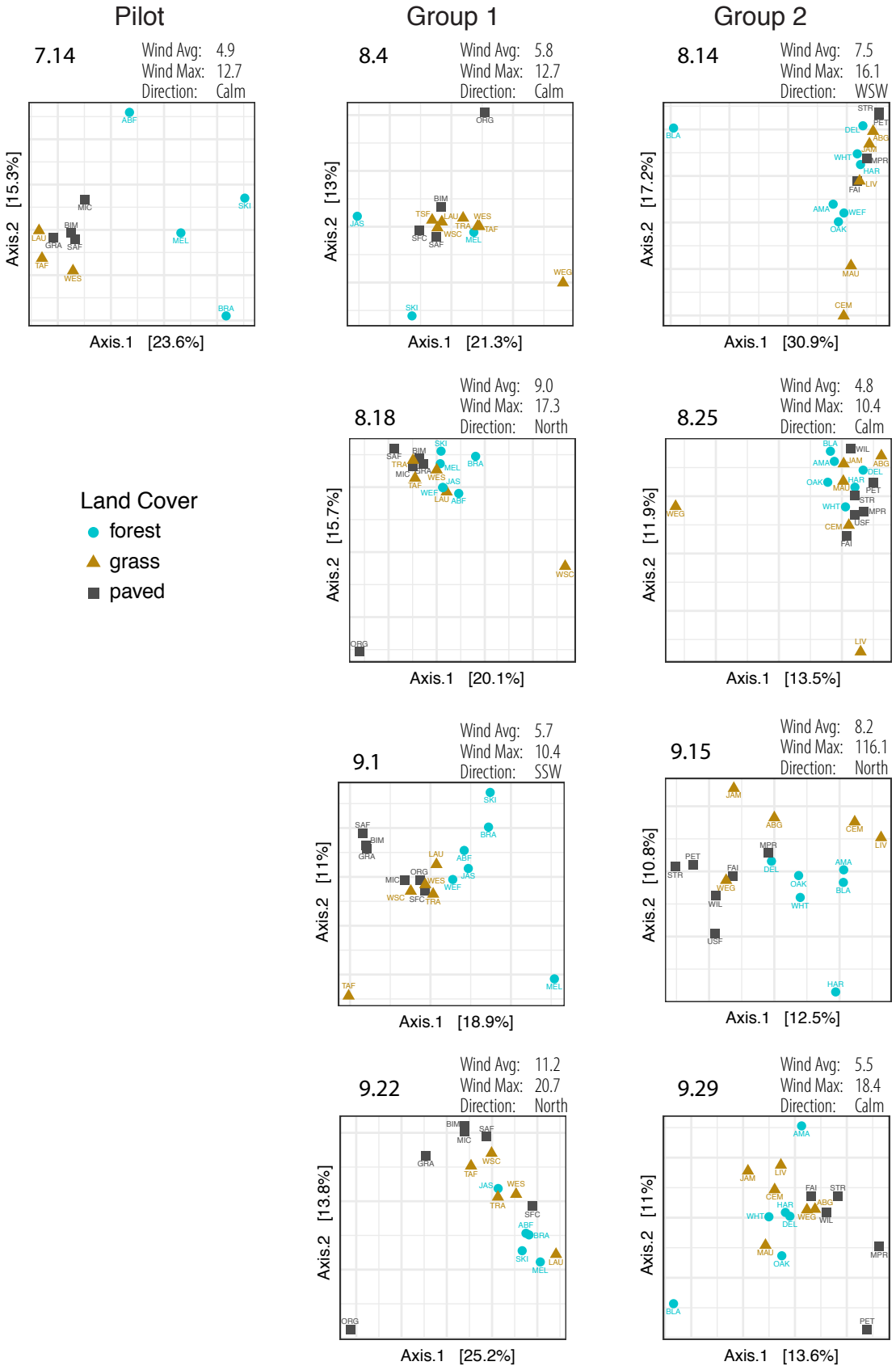


Figure 29. Unconstrained PCoA plots for each sampling date individually.

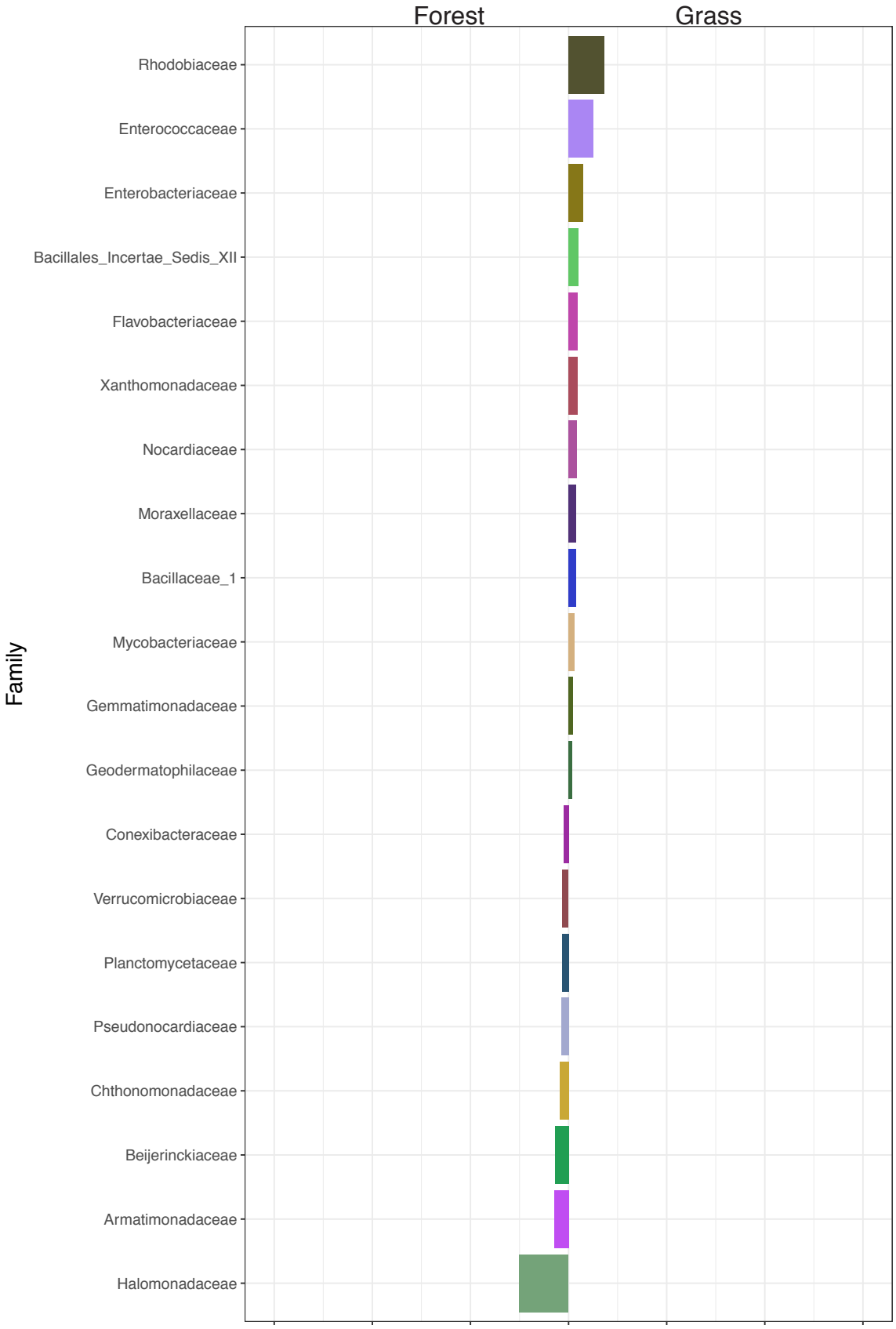


Figure 30. Differentially abundant families in forest vs. grass sites.

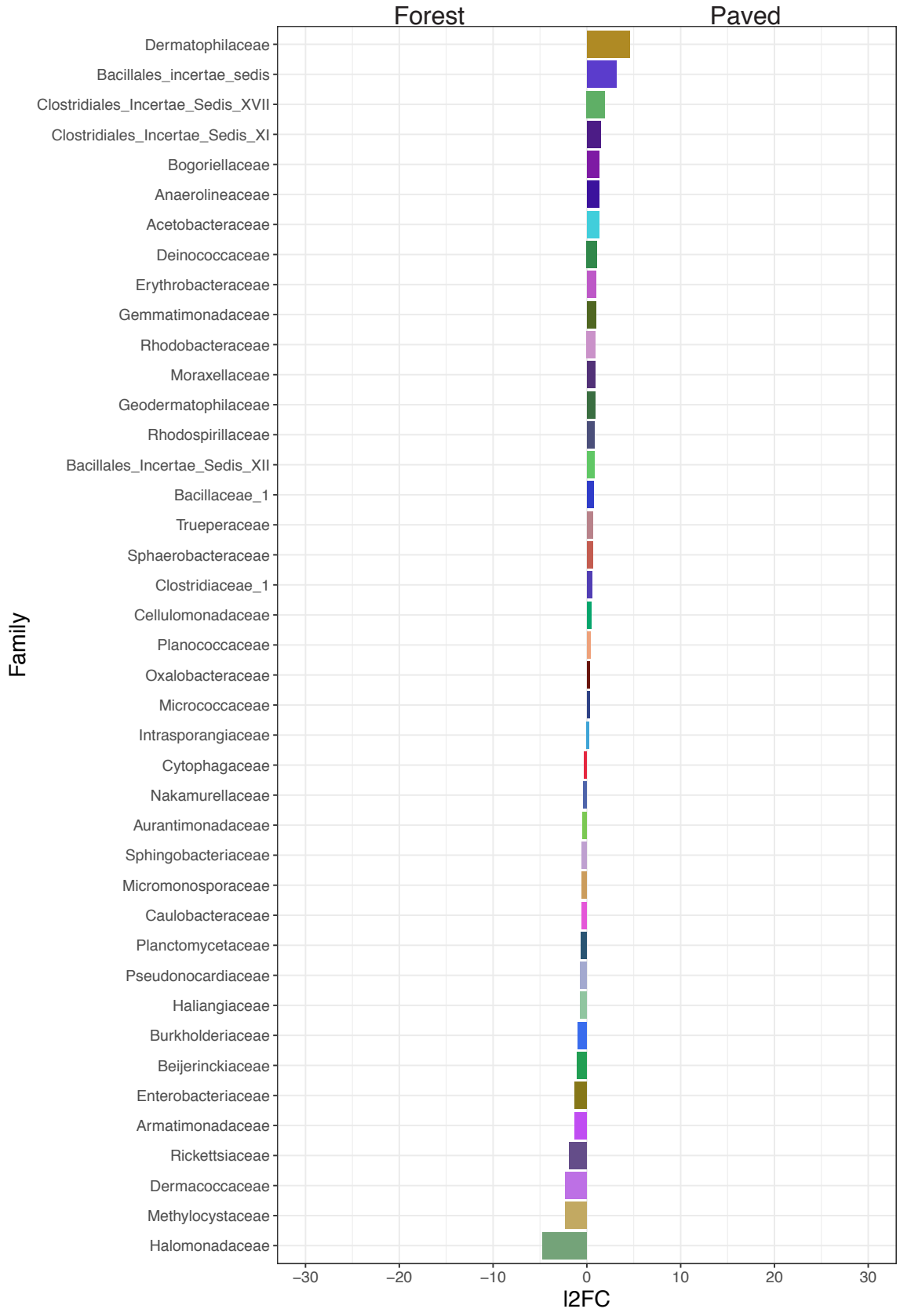


Figure 31. Differentially abundant families in forest vs. paved sites.

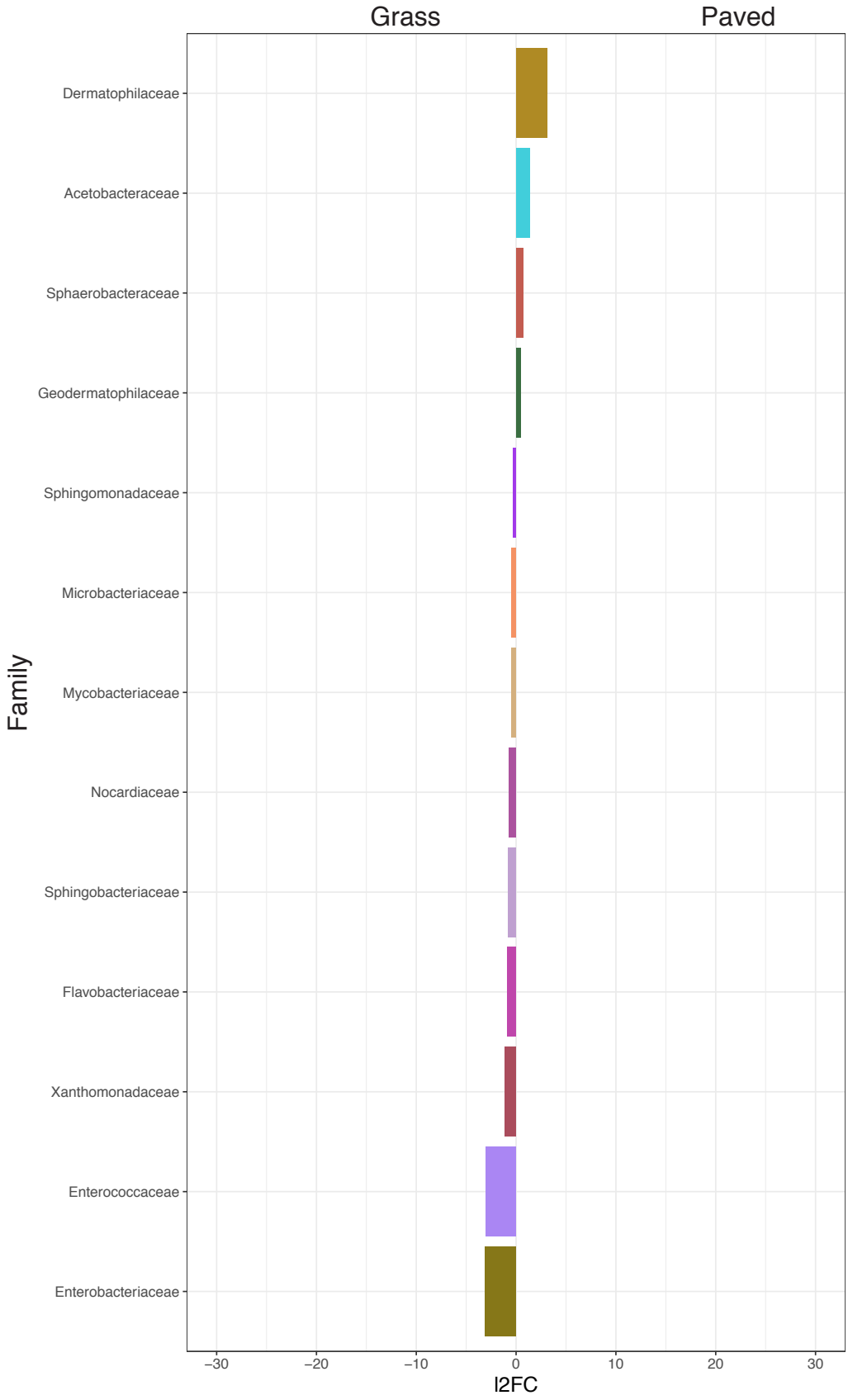


Figure 32. Differentially abundant families in grass vs. paved sites.

Table 11. Results of PERMANOVA analyses on other environmental factors for each sampling date, using Morisita-Horn dissimilarity distance on variance-stabilizing transformed counts. Values with an asterisk remain significant after Bonferroni correction.

	Land cover type		Vegetation cover (50 m)		Temperature average (°C)		Structural Diversity (25 m; forest & grass)	
Date	**R^2**	***p***	**R^2**	***p***	**R^2**	***p***	**R^2**	***p***
07.14.2015	0.31	0.008*	0.10	0.364	0.21	0.001*	0.22	0.001*
08.04.2015	0.18	0.053	0.09	0.207	0.12	0.019	0.09	0.162
08.14.2015	0.18	0.041	0.10	0.097	0.18	0.006*	0.11	0.010*
08.18.2015	0.18	0.000*	0.08	0.121	0.10	0.135	0.10	0.021
08.25.2015	0.13	0.082	0.06	0.511	0.07	0.063	0.06	0.282
09.01.2015	0.21	0.000*	0.11	0.004*	0.14	0.000*	0.15	0.000*
09.15.2015	0.18	0.000*	0.10	0.001*	0.06	0.515	0.09	0.006*
09.22.2015	0.21	0.004*	0.14	0.001*	0.09	0.106	0.14	0.001*
09.29.2015	0.16	0.001*	0.10	0.000*	0.07	0.410	0.09	0.002*

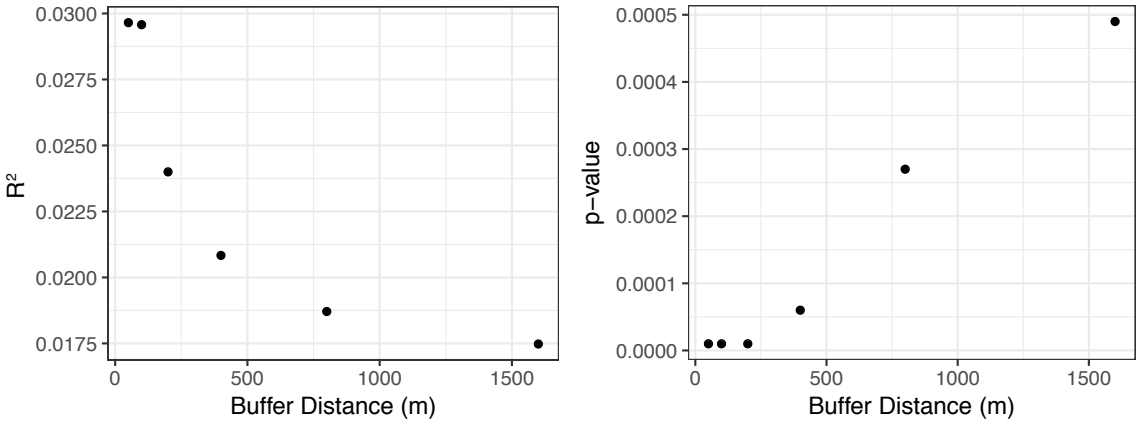


Figure 33. Change in a) explanatory power and b) statistical significance of the PERMANOVA model for vegetation cover with varying buffer zone radii.

APPENDIX D

R ANALYSIS CODE FOR CHAPTER III

````
---
title: “Influence of nearby vegetation on urban airborne microbial
communities”
author: “Mhuireach, G.A., Betancourt-Roman, C.M., Green, J.L., Johnson,
B.R”
date: ‘`r format(Sys.time(), “%B %d, %Y”)`’
header-includes:
  - \usepackage{graphicx}
output:
 pdf_document: default
 html_notebook:
  fig_caption: yes
---

```{r global_options, eval=TRUE, include=FALSE}
library(knitr)
knitr::opts_knit$set(root.dir=normalizePath(‘../’))
knitr::opts_chunk$set(dev=’pdf’, echo=FALSE, warning=FALSE,
message=FALSE, error=TRUE)
```

```{r initialSetup, include=FALSE}

set.seed(2)
options(scipen=7) # curtail scientific notation
options(digits=5) # number of digits to print on output

# install DADA2 for paired-end read assembly (instead of QIIME/Keaton’s
pipeline)
#source(“https://bioconductor.org/biocLite.R”)
#biocLite(“dada2”)
# also install ShortRead for DADA2 workflow
#biocLite(“DESeq2”)
#biocLite(“phyloseq”)

# The required package list:
reqpkg <- c(“DESeq2”, “ggplot2”, “phyloseq”, “lubridate”, “zoo”,
“vegan”, “ape”,
      “xts”, “VennDiagram”, “stargazer”, “tidyr”, “plyr”, “lidR”,
“ggmap”,
      “tiff”, “raster”, “viridis”, “data.table”)
# Load all required packages and show version
for (i in reqpkg) {
  print(i)
````
````

```
  print(packageVersion(i))
  library(i, quietly = TRUE, verbose = FALSE, warn.conflicts = FALSE,
character.only = TRUE)
}

## load extra functions
# function to extract OTU table from phyloseq and ensure samples are
rows
getTab <- function(physeq) {
 require("vegan")
 OTU = otu_table(physeq)
 if (taxa_are_rows(OTU)) {
  OTU = t(OTU)
 }
 return(as(OTU, "matrix"))
}

# set ggplot2 theme elements
theme_set(theme_bw(base_size=12))

# define palettes for coloring by family name, by Nearby_Veg, and by OTU
famPal <- read.csv("~/Documents/PhD work/dissertation/chapter3_
urbanMicrobiome/processed data/famPal.csv",
          header=FALSE, stringsAsFactors = FALSE)
famPal <- structure(famPal[,2], names=famPal[,1])
famPal <- c(famPal, Other="darkgrey")
length(famPal)

vegPal <- c(grass="darkgoldenrod", forest="turquoise3", paved="grey30")

#indicTaxa <- arrange(indicTaxa, SiteType)
#sum(indicTaxa[, indicTaxa$SiteType == "forest"])
#length(which(indicTaxa[, indicTaxa$SiteType == "grass"]))
#length(which(indicTaxa[, indicTaxa$SiteType == "paved"]))

pieColors <-

length(famPal)
length(pieColors)
pie(rep(1, length(famPal)), col = famPal) # check colors

```

```{r importDada, eval=FALSE, include=FALSE}
# import OTU tables
seqtab7.14.ec.pcr <- readRDS("~/Documents/PhD work/dissertation/
  chapter3_urbanMicrobiome/processed data/seqtab_nochimseqtab7.14.
ec.pcr.rds")
seqtab8.4.8.14 <- readRDS("~/Documents/PhD work/dissertation/
  chapter3_urbanMicrobiome/processed data/seqtab_nochimseqtab
```

```
  8.4.8.14.rds")
seqtab8.18.8.25 <- readRDS("~/Documents/PhD work/dissertation/
  chapter3_urbanMicrobiome/processed data/seqtab_nochimseqtab
  8.18.8.25.rds")
seqtab9.1 <- readRDS("~/Documents/PhD work/dissertation/
  chapter3_urbanMicrobiome/processed data/seqtab_nochimseqtab
  9.1.rds")
seqtab9.15 <- readRDS("~/Documents/PhD work/dissertation
  /chapter3_urbanMicrobiome/processed data/seqtab_nochimseqtab
  9.15.rds")
seqtab9.22 <- readRDS("~/Documents/PhD work/dissertation/
  chapter3_urbanMicrobiome/processed data/seqtab_nochimseqtab
  9.22.rds")
seqtab9.29 <- readRDS("~/Documents/PhD work/dissertation/
  chapter3_urbanMicrobiome/processed data/seqtab_nochimseqtab
  9.29.rds")

# import taxa tables
taxTab7.14 <- readRDS("~/Documents/PhD work/dissertation/
  chapter3_urbanMicrobiome/processed data/taxTab7.14.rds")
taxTab8.4 <- readRDS("~/Documents/PhD work/dissertation/
  chapter3_urbanMicrobiome/processed data/taxTab8.4.rds")
taxTab8.18 <- readRDS("~/Documents/PhD work/dissertation/
  chapter3_urbanMicrobiome/processed data/taxTab8.18.rds")
taxTab9.1 <- readRDS("~/Documents/PhD work/dissertation/
  chapter3_urbanMicrobiome/processed data/taxTab9.1.rds")
taxTab9.15 <- readRDS("~/Documents/PhD work/dissertation/
  chapter3_urbanMicrobiome/processed data/taxTab9.15.rds")
taxTab9.22 <- readRDS("~/Documents/PhD work/dissertation/
  chapter3_urbanMicrobiome/processed data/taxTab9.22.rds")
taxTab9.29 <- readRDS("~/Documents/PhD work/dissertation/
  chapter3_urbanMicrobiome/processed data/taxTab9.29.rds")

sampdata <- as.data.frame(read.csv("mapFile_2-23-2018.csv",
  header = TRUE, stringsAsFactors = FALSE))
row.names(sampdata) <- sampdata$SampleID
sampdata[1:20,1:6]
urbair7.14 <- phyloseq(otu_table(seqtab7.14.ec.pcr,
  taxa_are_rows=FALSE), tax_table(taxTab7.14))
urbair8.4 <- phyloseq(otu_table(seqtab8.4.8.14,
  taxa_are_rows=FALSE), tax_table(taxTab8.4))
urbair8.18 <- phyloseq(otu_table(seqtab8.18.8.25,
  taxa_are_rows=FALSE), tax_table(taxTab8.18))
urbair9.1 <- phyloseq(otu_table(seqtab9.1,
  taxa_are_rows=FALSE), tax_table(taxTab9.1))
urbair9.15 <- phyloseq(otu_table(seqtab9.15,
  taxa_are_rows=FALSE), tax_table(taxTab9.15))
urbair9.22 <- phyloseq(otu_table(seqtab9.22, taxa_are_rows=FALSE), tax_
table(taxTab9.22))
urbair9.29 <- phyloseq(otu_table(seqtab9.29, taxa_are_rows=FALSE), tax_
```

```
table(taxTab9.29))
urbair <- merge_phyloseq(urbair7.14, urbair8.4, urbair8.18, urbair9.1,
urbair9.15, urbair9.22, urbair9.29)
urbair <- merge_phyloseq(urbair, sample_data(sampdata))
#is.rooted(phy_tree(urbair))
#phy_tree(urbair) <- root(phy_tree(urbair), sample(taxa_names(urbair),
1), resolve.root = TRUE)

saveRDS(urbair, file="~/Documents/PhD work/dissertation/chapter3_
urbanMicrobiome/processed data/urbair_dada.rds")

# clean up
uaList <- paste("urbair", c("7.14", "8.4", "8.18", "9.1", "9.15",
"9.22", "9.29"), sep = "")
rm(list = uaList)
rm(uaList)
ttList <- paste("taxTab", c("7.14", "8.4", "8.18", "9.1", "9.15",
"9.22", "9.29"), sep = "")
rm(list = ttList)
rm(ttList)
stList <- paste("seqtab", c("7.14.ec.pcr", "8.4.8.14", "8.18.8.25",
"9.1", "9.15", "9.22", "9.29"), sep = "")
rm(list = stList)
rm(stList)

```

```{r loadDada, include=FALSE}

urbair <- readRDS(file="~/Documents/PhD work/dissertation/chapter3_
urbanMicrobiome/processed data/urbair_dada.rds")

```

Methods

```{r evalContams, include=FALSE, results="hide"}

# identify potential lab contaminants
urbair_ec1 <- prune_samples(sample_names(urbair) == "EC-6-29", urbair)
urbair_ec2 <- prune_samples(sample_names(urbair) == "EC-6-22", urbair)
urbair_ec3 <- prune_samples(sample_names(urbair) == "EC-6-13", urbair)
urbair_ec4 <- prune_samples(sample_names(urbair) == "EC-6-21", urbair)
urbair_ec5 <- prune_samples(sample_names(urbair) == "EC-7-4", urbair)
urbair_ec6 <- prune_samples(sample_names(urbair) == "EC-7-8", urbair)
urbair_ec7 <- prune_samples(sample_names(urbair) == "EC-6-20", urbair)
urbair_ec8 <- prune_samples(sample_names(urbair) == "EC-6-16", urbair)
urbair_ec9 <- prune_samples(sample_names(urbair) == "EC-6-30", urbair)
urbair_pn1 <- prune_samples(sample_names(urbair) == "PCR_NEG_01",
urbair)
```

```
urbair_pn2 <- prune_samples(sample_names(urbair) == “PCR_NEG_02”,
urbair)
urbair_pn3 <- prune_samples(sample_names(urbair) == “PCR_NEG_03”,
urbair)
urbair_pn4 <- prune_samples(sample_names(urbair) == “PCR_NEG_04”,
urbair)
urbair_pn5 <- prune_samples(sample_names(urbair) == “PCR_NEG_05”,
urbair)
urbair_pn6 <- prune_samples(sample_names(urbair) == “PCR_NEG_06”,
urbair)
urbair_neg1 <- prune_samples(sample_names(urbair) == “NEG-8-4”, urbair)
urbair_neg2 <- prune_samples(sample_names(urbair) == “NEG-8-14”,
urbair)
urbair_neg3 <- prune_samples(sample_names(urbair) == “NEG-8-25”,
urbair)
urbair_neg4 <- prune_samples(sample_names(urbair) == “NEG-9-29”,
urbair)
urbair_contams <- merge_phyloseq(urbair_ec1, urbair_ec2)
urbair_contams <- merge_phyloseq(urbair_contams, urbair_ec3)
urbair_contams <- merge_phyloseq(urbair_contams, urbair_ec4)
urbair_contams <- merge_phyloseq(urbair_contams, urbair_ec5)
urbair_contams <- merge_phyloseq(urbair_contams, urbair_ec6)
urbair_contams <- merge_phyloseq(urbair_contams, urbair_ec7)
urbair_contams <- merge_phyloseq(urbair_contams, urbair_ec8)
urbair_contams <- merge_phyloseq(urbair_contams, urbair_ec9)
urbair_contams <- merge_phyloseq(urbair_contams, urbair_pn1)
urbair_contams <- merge_phyloseq(urbair_contams, urbair_pn2)
urbair_contams <- merge_phyloseq(urbair_contams, urbair_pn3)
urbair_contams <- merge_phyloseq(urbair_contams, urbair_pn4)
urbair_contams <- merge_phyloseq(urbair_contams, urbair_pn5)
urbair_contams <- merge_phyloseq(urbair_contams, urbair_pn6)
urbair_contams <- merge_phyloseq(urbair_contams, urbair_neg1)
urbair_contams <- merge_phyloseq(urbair_contams, urbair_neg2)
urbair_contams <- merge_phyloseq(urbair_contams, urbair_neg3)
urbair_contams <- merge_phyloseq(urbair_contams, urbair_neg4)

urbair_contams <- prune_taxa(taxa_sums(urbair_contams) > 0, urbair_
contams) # remove taxa with counts of 0
contamAbund <- data.frame(tax_table(urbair_contams))
contamAbund$Count <- taxa_sums(urbair_contams)
contamAbund$RelAbund <- contamAbund$Count/sum(sample_sums(urbair))
write.csv(contamAbund, file=”contamAbund.csv”)
# what are the 3 most abundant contaminants?
top3cont <- sort(taxa_sums(urbair_contams), TRUE)[1:3]
top3cont <- prune_taxa(names(top3cont), urbair_contams)
# examine prevalence of top 3 potential contaminants across all samples
write.csv(as.data.frame(get_sample(urbair, otu_table(top3cont))),
file=”sampleContams.csv”)

################# THIS IS FROM JAMES’S CELL PHONE PAPER
```

```
# plot relative abundances of potential contaminants in controls vs
samples

cont.table <- otu_table(urbair_contams) # get otu table of potential
contaminants
#cont.table[1:10,1:10]
cont.otus <- colnames(cont.table) # which otus are they
#cont.otus[1:10]
cont.otus.names <- tax_table(urbair_contams) # what are their names
#cont.otus.names[1:10,1:8]
# cont.taxo <- makeTaxo(taxo.in=rw.taxo.tmp$taxa.names, otu.table=cont.
table) # don't seem to need this
cont3.otus <- which(colSums(cont.table)/sum(cont.table) > 0.05) # pick
out some big ones
plotY <- colSums(cont.table[, cont.otus]/sum(cont.table)) # Y
coordinates for relative abundance of potential contaminants in
controls
rw.table.tmp <- otu_table(urbair) #??????
#rw.table.tmp[1:10,1:10]
rw.taxo.tmp <- tax_table((urbair))
#rw.taxo.tmp[1:5,1:8]
plotX <- colSums(rw.table.tmp[, cont.otus]/sum(rw.table.tmp)) # X
coordinates for relative abundance of potential contaminants in samples
#pdf("fig1.pdf", height = 5, width = 7)
plot(plotY ~ plotX,
   pch=21, bg=rgb(0,0,0,.3), cex=2, las=1,
   xlab='Rel Abundance in Experiment', ylab='Rel Abundance in
Controls')
segments(0,0,1,1, lty=3, lwd=2, col='gray')
segments(0, .05, 1, .05, lty=1, lwd=2, col='tomato')
text(.12, .13, '1:1', font=3, col='gray30')
text(.1, .05, 'RA=0.05', font=3, pos=3, col='tomato')
text(plotX[names(cont3.otus)[1:3]], plotY[names(cont3.otus)[1:3]],
   rw.taxo.tmp[names(cont3.otus)[1:3], 'Genus'], pos=c(1))
#dev.off()

# clean up
rm(cont.table, cont.otus, cont.otus.names, plotY, rw.table.tmp,
rw.taxo.tmp, plotX)
#rm(urbair_contams)
# clean up
ecList <- paste("urbair_ec", 1:9, sep = "")
rm(list = ecList)
rm(ecList)
pnList <- paste("urbair_pn", 1:6, sep = "")
rm(list = pnList)
rm(pnList)
negList <- paste("urbair_neg", 1:4, sep = "")
rm(list = negList)
rm(negList)
```

```

```{r noFails, include=FALSE, results="hide"}

# remove samples with < 10000 reads
sort(sample_sums(urbair))
fails <- prune_samples(sample_sums(urbair) < 10000, urbair)
urbair <- prune_samples(sample_sums(urbair) > 10000, urbair)
sample_sums(fails)
# do test ordination
UAord <- ordinate(urbair)
#pdf("fig2.pdf", height = 5, width = 7)
plot_ordination(urbair, UAord, type = "samples", color = "Nearby_Veg",
        label = "SampleID", title = "Only failed samples removed")
#dev.off()

```

```{r noPosPCR, include=FALSE, results="hide"}
# remove PCR_POS controls
UA_noneg <- prune_samples(sample_names(urbair) != "PCR_POS_01", urbair)
UA_noneg <- prune_samples(sample_names(UA_noneg) != "PCR_POS_02", UA_
noneg)
UA_noneg <- prune_samples(sample_names(UA_noneg) != "PCR_POS_03", UA_
noneg)
UA_noneg <- prune_samples(sample_names(UA_noneg) != "PCR_POS_04", UA_
noneg)
UA_noneg <- prune_samples(sample_names(UA_noneg) != "PCR_POS_05", UA_
noneg)
UA_noneg <- prune_samples(sample_names(UA_noneg) != "PCR_POS_06", UA_
noneg)
# do test ordination
#pdf("fig3.pdf", height = 5, width = 7)
UAord <- ordinate(UA_noneg)
plot_ordination(UA_noneg, UAord, type = "samples", color = "Nearby_
Veg",
        label = "SampleID", title = "PCR_POS removed")
#dev.off()

```

```{r noNegPCR, include=FALSE, results="hide"}
# remove negative PCR controls
UA_noneg <- prune_samples(sample_names(UA_noneg) != "PCR_NEG_01", UA_
noneg)
UA_noneg <- prune_samples(sample_names(UA_noneg) != "PCR_NEG_02", UA_
noneg)
UA_noneg <- prune_samples(sample_names(UA_noneg) != "PCR_NEG_03", UA_
noneg)

```
UA_noneg <- prune_samples(sample_names(UA_noneg) != “PCR_NEG_04”, UA_
noneg)
UA_noneg <- prune_samples(sample_names(UA_noneg) != “PCR_NEG_05”, UA_
noneg)
UA_noneg <- prune_samples(sample_names(UA_noneg) != “PCR_NEG_06”, UA_
noneg)

# test ordination
UAord <- ordinate(UA_noneg)
#pdf(“fig4.pdf”, height = 5, width = 7)
plot_ordination(UA_noneg, UAord, type = “samples”, color = “Nearby_
Veg”, label = “SampleID”, title = “PCR_POS & PCR_NEG removed”)
#dev.off()
colnames(sample_data(UA_noneg))[1] <- “SampleID”
```

```{r noEC, include=FALSE, results=”hide”}
# remove extraction control samples
UA_noneg <- prune_samples(sample_names(UA_noneg) != “EC-6-29”, UA_
noneg)
UA_noneg <- prune_samples(sample_names(UA_noneg) != “EC-6-22”, UA_
noneg)
UA_noneg <- prune_samples(sample_names(UA_noneg) != “EC-6-13”, UA_
noneg)
UA_noneg <- prune_samples(sample_names(UA_noneg) != “EC-6-21”, UA_
noneg)
UA_noneg <- prune_samples(sample_names(UA_noneg) != “EC-7-4”, UA_noneg)
UA_noneg <- prune_samples(sample_names(UA_noneg) != “EC-7-8”, UA_noneg)
UA_noneg <- prune_samples(sample_names(UA_noneg) != “EC-6-20”, UA_
noneg)
UA_noneg <- prune_samples(sample_names(UA_noneg) != “EC-6-16”, UA_
noneg)
UA_noneg <- prune_samples(sample_names(UA_noneg) != “EC-6-30”, UA_
noneg)

# test ordination
UAord <- ordinate(UA_noneg)
#pdf(“fig5.pdf”, height = 5, width = 7)
plot_ordination(UA_noneg, UAord, type = “samples”, color = “Nearby_
Veg”, label = “SampleID”, title = “PCR_POS, PCR_NEG & EC removed”)
#dev.off()

```

```{r noNegPet, include=FALSE, results=”hide”}
# remove negative petri controls
UA_noneg <- prune_samples(sample_names(UA_noneg) != “NEG-8-4”, UA_
noneg)
UA_noneg <- prune_samples(sample_names(UA_noneg) != “NEG-8-14”, UA_
noneg)
```

```
UA_noneg <- prune_samples(sample_names(UA_noneg) != “NEG-8-25”, UA_
noneg)
UA_noneg <- prune_samples(sample_names(UA_noneg) != “NEG-9-29”, UA_
noneg)

# test ordination
UAord <- ordinate(UA_noneg)
#pdf(“fig6.pdf”, height = 5, width = 7)
plot_ordination(UA_noneg, UAord, type = “samples”, color = “Nearby_
Veg”, label = “SampleID”, title = “PCR_POS, EC, PCR_NEG & NEG removed”)
#dev.off()
#pdf(“fig7.pdf”, height = 5, width = 7)
plot_ordination(UA_noneg, UAord, type = “samples”, color = “Date”,
label = “SampleID”, title = “PCR_POS, EC, PCR_NEG & NEG removed”)
#dev.off()

```

```{r noExtras, include=FALSE, results=”hide”}
# remove extra samples (7/14 test samples and 9/8 Pisgah samples,
CEM_8_14, also STE and URB bc they are in UGB)
prune <- prune_samples(sample_names(UA_noneg) != “PIC_9_8”, UA_noneg)
prune <- prune_samples(sample_names(prune) != “PIO_9_8”, prune)
prune <- prune_samples(sample_names(prune) != “CEM-8-14”, prune)
prune <- prune_samples(sample_names(prune) != “URB-9-1”, prune)
prune <- prune_samples(sample_names(prune) != “STE-9-1”, prune)
prune <- prune_samples(sample_names(prune) != “URB-9-15”, prune)
prune <- prune_samples(sample_names(prune) != “STE-9-15”, prune)
prune <- prune_samples(sample_names(prune) != “URB-9-22”, prune)
prune <- prune_samples(sample_names(prune) != “STE-9-22”, prune)
prune <- prune_samples(sample_names(prune) != “URB-9-29”, prune)

# clean up
any(taxa_sums(prune) == 0)
sum(taxa_sums(prune) == 0)
prune <- prune_taxa(taxa_sums(prune) != 0, prune)

# test ordination
UAord <- ordinate(prune)
#pdf(“fig8.pdf”, height = 5, width = 7)
plot_ordination(prune, UAord, type = “samples”, color = “Nearby_Veg”,
label = “SampleID”, title = “Controls & extras removed”)
#dev.off()
#pdf(“fig9.pdf”, height = 5, width = 7)
plot_ordination(prune, UAord, type = “samples”, color = “Date”, label =
“SampleID”, title = “Controls & extras removed”)
#dev.off()

# clean up
#rm(UA_noneg)
```

```

```{r noContams, include=FALSE, results="hide"}
sort(sample_sums(prune))
sum(taxa_sums(prune))
# remove potential contaminants above 0.05 relative abundance in
controls (see contams plot), plus Halomonas and Shewanella
prune <- prune_taxa(taxa_names(prune) != names(cont3.otus)[1], prune)
prune <- prune_taxa(taxa_names(prune) != names(cont3.otus)[2], prune)
prune <- prune_taxa(taxa_names(prune) != names(cont3.otus)[3], prune)

sort(sample_sums(prune))
sum(taxa_sums(prune))

#prune <- subset_taxa(prune, Genus!="Halomonas")
#prune <- subset_taxa(prune, Genus!="Shewanella")

# remove plant and mitochondrial sequences (RDP classifier removes
mitochondria already)
sort(get_taxa_unique(prune, "Family")) # none are present
prune <- subset_taxa(prune, Family!="Streptophyta")
prune <- subset_taxa(prune, Family!="Chlorophyta")

# check and clean up
any(taxa_sums(prune) == 0)
sum(taxa_sums(prune) == 0)
prune <- prune_taxa(taxa_sums(prune) != 0, prune)

# test ordination
UAord <- ordinate(prune)
#pdf("fig10.pdf", height = 5, width = 7)
plot_ordination(prune, UAord, type = "samples", color = "Nearby_Veg",
label = "SampleID", title = "Controls, extras & contams removed")
#dev.off()
#pdf("fig11.pdf", height = 5, width = 7)
plot_ordination(prune, UAord, type = "samples", color = "Date", label =
"SampleID", title = "Controls, extras & contams removed")
#dev.off()

```

```{r noSing, include=FALSE, results="hide"}
sum(sample_sums(prune))
sum(taxa_sums(prune))
# how many single- and doubletons?
sing <- prune_taxa(taxa_sums(prune) < 3, prune)
sing
sum(sample_sums(sing))
sum(taxa_sums(sing))
```
```

```
# remove them
nosing <- prune_taxa(taxa_sums(prune) > 2, prune)

# test ordination
UAord <- ordinate(nosing)
#pdf("fig12.pdf", height = 5, width = 7)
plot_ordination(nosing, UAord, type = "samples", color = "Nearby_Veg",
label = "SampleID", title = "Controls, extras, contams & single/
doubletons removed")
#dev.off()
#pdf("fig13.pdf", height = 5, width = 7)
plot_ordination(nosing, UAord, type = "samples", color = "Date", label
= "SampleID", title = "Controls, extras, contams & single/doubletons
removed")
#dev.off()

saveRDS(nosing, file="~/Documents/PhD work/dissertation/chapter3_
urbanMicrobiome/processed data/nosing_dada.rds")

saveRDS(prune, file="~/Documents/PhD work/dissertation/chapter3_
urbanMicrobiome/processed data/urbair_sing_dada.rds")

#clean up

```


```{r splitSamples, include=FALSE, results="hide"}
# separate peri-urban sites
urb.sing <- prune_samples(sample_data(prune)$Urbanicity == "urban",
prune)
urb.sing <- prune_taxa(taxa_sums(urb.sing) != 0, urb.sing)
urb.nosing <- prune_samples(sample_data(nosing)$Urbanicity == "urban",
nosing)
urb.nosing <- prune_taxa(taxa_sums(urb.nosing) != 0, urb.nosing)

# separate by date, single- and doubletons removed
urb7.14 <- prune_samples(sample_data(urb.nosing)$Date=="07.14.2015",
urb.nosing)
urb7.14 <- prune_taxa(taxa_sums(urb7.14) != 0, urb7.14)
urb8.4 <- prune_samples(sample_data(urb.nosing)$Date=="08.04.2015",
urb.nosing)
urb8.4 <- prune_taxa(taxa_sums(urb8.4) != 0, urb8.4)
urb8.14 <- prune_samples(sample_data(urb.nosing)$Date=="08.14.2015",
urb.nosing)
urb8.14 <- prune_taxa(taxa_sums(urb8.14) != 0, urb8.14)
urb8.18 <- prune_samples(sample_data(urb.nosing)$Date=="08.18.2015",
urb.nosing)
urb8.18 <- prune_taxa(taxa_sums(urb8.18) != 0, urb8.18)
urb8.25 <- prune_samples(sample_data(urb.nosing)$Date=="08.25.2015",
urb.nosing)
```

```
urb8.25 <- prune_taxa(taxa_sums(urb8.25) != 0, urb8.25)
urb9.1 <- prune_samples(sample_data(urb.nosing)$Date==”09.01.2015”,
urb.nosing)
urb9.1 <- prune_taxa(taxa_sums(urb9.1) != 0, urb9.1)
urb9.15 <- prune_samples(sample_data(urb.nosing)$Date==”09.15.2015”,
urb.nosing)
urb9.15 <- prune_taxa(taxa_sums(urb9.15) != 0, urb9.15)
urb9.22 <- prune_samples(sample_data(urb.nosing)$Date==”09.22.2015”,
urb.nosing)
urb9.22 <- prune_taxa(taxa_sums(urb9.22) != 0, urb9.22)
urb9.29 <- prune_samples(sample_data(urb.nosing)$Date==”09.29.2015”,
urb.nosing)
urb9.29 <- prune_taxa(taxa_sums(urb9.29) != 0, urb9.29)
dateList <- list(urb7.14=urb7.14, urb8.4=urb8.4, urb8.14=urb8.14,
urb8.18=urb8.18,
        urb8.25=urb8.25, urb9.1=urb9.1, urb9.15=urb9.15,
urb9.22=urb9.22,
        urb9.29=urb9.29)
saveRDS(dateList, file=”~/Documents/PhD work/dissertation/chapter3_
urbanMicrobiome/processed data/sampDate_dada.rds”)

#clean up
garbage <- paste0(“urb”, c(“7.14”,”8.4”,”8.14”,”8.25”,”9.1”,”9.15”,”9.2
2”,”9.29”))
rm(list = garbage)

# separate by site, single- and doubletons removed
ABF <- prune_samples(sample_data(urb.nosing)$SiteCode==”ABF”, urb.
nosing)
ABF <- prune_taxa(taxa_sums(ABF) != 0, ABF)
BIM <- prune_samples(sample_data(urb.nosing)$SiteCode==”BIM”, urb.
nosing)
BIM <- prune_taxa(taxa_sums(BIM) != 0, BIM)
BRA <- prune_samples(sample_data(urb.nosing)$SiteCode==”BRA”, urb.
nosing)
BRA <- prune_taxa(taxa_sums(BRA) != 0, BRA)
GRA <- prune_samples(sample_data(urb.nosing)$SiteCode==”GRA”, urb.
nosing)
GRA <- prune_taxa(taxa_sums(GRA) != 0, GRA)
LAU <- prune_samples(sample_data(urb.nosing)$SiteCode==”LAU”, urb.
nosing)
LAU <- prune_taxa(taxa_sums(LAU) != 0, LAU)
MEL <- prune_samples(sample_data(urb.nosing)$SiteCode==”MEL”, urb.
nosing)
MEL <- prune_taxa(taxa_sums(MEL) != 0, MEL)
MIC <- prune_samples(sample_data(urb.nosing)$SiteCode==”MIC”, urb.
nosing)
MIC <- prune_taxa(taxa_sums(MIC) != 0, MIC)
SAF <- prune_samples(sample_data(urb.nosing)$SiteCode==”SAF”, urb.
nosing)
```

```
SAF <- prune_taxa(taxa_sums(SAF) != 0, SAF)
SKI <- prune_samples(sample_data(urb.nosing)$SiteCode==”SKI”, urb.
nosing)
SKI <- prune_taxa(taxa_sums(SKI) != 0, SKI)
TAF <- prune_samples(sample_data(urb.nosing)$SiteCode==”TAF”, urb.
nosing)
TAF <- prune_taxa(taxa_sums(TAF) != 0, TAF)
WES <- prune_samples(sample_data(urb.nosing)$SiteCode==”WES”, urb.
nosing)
WES <- prune_taxa(taxa_sums(WES) != 0, WES)
JAS <- prune_samples(sample_data(urb.nosing)$SiteCode==”JAS”, urb.
nosing)
JAS <- prune_taxa(taxa_sums(JAS) != 0, JAS)
ORG <- prune_samples(sample_data(urb.nosing)$SiteCode==”ORG”, urb.
nosing)
ORG <- prune_taxa(taxa_sums(ORG) != 0, ORG)
SFC <- prune_samples(sample_data(urb.nosing)$SiteCode==”SFC”, urb.
nosing)
SFC <- prune_taxa(taxa_sums(SFC) != 0, SFC)
TRA <- prune_samples(sample_data(urb.nosing)$SiteCode==”TRA”, urb.
nosing)
TRA <- prune_taxa(taxa_sums(TRA) != 0, TRA)
TSF <- prune_samples(sample_data(urb.nosing)$SiteCode==”TSF”, urb.
nosing)
TSF <- prune_taxa(taxa_sums(TSF) != 0, TSF)
WEG <- prune_samples(sample_data(urb.nosing)$SiteCode==”WEG”, urb.
nosing)
WEG <- prune_taxa(taxa_sums(WEG) != 0, WEG)
WSC <- prune_samples(sample_data(urb.nosing)$SiteCode==”WSC”, urb.
nosing)
WSC <- prune_taxa(taxa_sums(WSC) != 0, WSC)
ABG <- prune_samples(sample_data(urb.nosing)$SiteCode==”ABG”, urb.
nosing)
ABG <- prune_taxa(taxa_sums(ABG) != 0, ABG)
AMA <- prune_samples(sample_data(urb.nosing)$SiteCode==”AMA”, urb.
nosing)
AMA <- prune_taxa(taxa_sums(AMA) != 0, AMA)
BLA <- prune_samples(sample_data(urb.nosing)$SiteCode==”BLA”, urb.
nosing)
BLA <- prune_taxa(taxa_sums(BLA) != 0, BLA)
CEM <- prune_samples(sample_data(urb.nosing)$SiteCode==”CEM”, urb.
nosing)
CEM <- prune_taxa(taxa_sums(CEM) != 0, CEM)
DEL <- prune_samples(sample_data(urb.nosing)$SiteCode==”DEL”, urb.
nosing)
DEL <- prune_taxa(taxa_sums(DEL) != 0, DEL)
FAI <- prune_samples(sample_data(urb.nosing)$SiteCode==”FAI”, urb.
nosing)
FAI <- prune_taxa(taxa_sums(FAI) != 0, FAI)
HAR <- prune_samples(sample_data(urb.nosing)$SiteCode==”HAR”, urb.
```

```
nosing)
HAR <- prune_taxa(taxa_sums(HAR) != 0, HAR)
JAM <- prune_samples(sample_data(urb.nosing)$SiteCode==”JAM”, urb.
nosing)
JAM <- prune_taxa(taxa_sums(JAM) != 0, JAM)
LIV <- prune_samples(sample_data(urb.nosing)$SiteCode==”LIV”, urb.
nosing)
LIV <- prune_taxa(taxa_sums(LIV) != 0, LIV)
MAU <- prune_samples(sample_data(urb.nosing)$SiteCode==”MAU”, urb.
nosing)
MAU <- prune_taxa(taxa_sums(MAU) != 0, MAU)
MPR <- prune_samples(sample_data(urb.nosing)$SiteCode==”MPR”, urb.
nosing)
MPR <- prune_taxa(taxa_sums(MPR) != 0, MPR)
OAK <- prune_samples(sample_data(urb.nosing)$SiteCode==”OAK”, urb.
nosing)
OAK <- prune_taxa(taxa_sums(OAK) != 0, OAK)
PET <- prune_samples(sample_data(urb.nosing)$SiteCode==”PET”, urb.
nosing)
PET <- prune_taxa(taxa_sums(PET) != 0, PET)
STR <- prune_samples(sample_data(urb.nosing)$SiteCode==”STR”, urb.
nosing)
STR <- prune_taxa(taxa_sums(STR) != 0, STR)
WEF <- prune_samples(sample_data(urb.nosing)$SiteCode==”WEF”, urb.
nosing)
WEF <- prune_taxa(taxa_sums(WEF) != 0, WEF)
WHT <- prune_samples(sample_data(urb.nosing)$SiteCode==”WHT”, urb.
nosing)
WHT <- prune_taxa(taxa_sums(WHT) != 0, WHT)
USF <- prune_samples(sample_data(urb.nosing)$SiteCode==”USF”, urb.
nosing)
USF <- prune_taxa(taxa_sums(USF) != 0, USF)
WIL <- prune_samples(sample_data(urb.nosing)$SiteCode==”WIL”, urb.
nosing)
WIL <- prune_taxa(taxa_sums(WIL) != 0, WIL)

siteList <- list(ABF=ABF, BIM=BIM, BRA=BRA, GRA=GRA, LAU=LAU, MEL=MEL,
MIC=MIC, SAF=SAF,
         SKI=SKI, TAF=TAF, WES=WES, JAS=JAS, ORG=ORG, SFC=SFC, TRA=TRA,
TSF=TSF, WEG=WEG,
         WSC=WSC, ABG=ABG, AMA=AMA, BLA=BLA, CEM=CEM, DEL=DEL, FAI=FAI,
HAR=HAR, JAM=JAM,
         LIV=LIV, MAU=MAU, MPR=MPR, OAK=OAK, PET=PET, STR=STR, WEF=WEF,
WHT=WHT, USF=USF,
         WIL=WIL)
saveRDS(siteList, file=”~/Documents/PhD work/dissertation/chapter3_
urbanMicrobiome/processed data/sampSite_dada.rds”)

#clean up
garbage <- c(“ABF”, “BIM”, “BRA”, “GRA”, “LAU”, “MEL”, “MIC”, “SAF”,
```

```
"SKI", "TAF", "WES", "JAS",
       "ORG", "SFC", "TRA", "TSF", "WEG", "WSC", "ABG", "AMA", "BLA",
"CEM", "DEL", "FAI",
       "HAR", "JAM", "LIV", "MAU", "MPR", "OAK", "PET", "STR", "WEF",
"WHT", "USF", "WIL")
rm(list = garbage)

## rural site list, single- and doubletons removed
BUF <- prune_samples(sample_data(nosing)$SiteCode=="BUF", nosing)
BUF <- prune_taxa(taxa_sums(BUF) != 0, BUF)
COG <- prune_samples(sample_data(nosing)$SiteCode=="COG", nosing)
COG <- prune_taxa(taxa_sums(COG) != 0, COG)
COW <- prune_samples(sample_data(nosing)$SiteCode=="COW", nosing)
COW <- prune_taxa(taxa_sums(COW) != 0, COW)
DAL <- prune_samples(sample_data(nosing)$SiteCode=="DAL", nosing)
DAL <- prune_taxa(taxa_sums(DAL) != 0, DAL)
DEB <- prune_samples(sample_data(nosing)$SiteCode=="DEB", nosing)
DEB <- prune_taxa(taxa_sums(DEB) != 0, DEB)
KAR <- prune_samples(sample_data(nosing)$SiteCode=="KAR", nosing)
KAR <- prune_taxa(taxa_sums(KAR) != 0, KAR)
PHY <- prune_samples(sample_data(nosing)$SiteCode=="PHY", nosing)
PHY <- prune_taxa(taxa_sums(PHY) != 0, PHY)
PIC <- prune_samples(sample_data(nosing)$SiteCode=="PIC", nosing)
PIC <- prune_taxa(taxa_sums(PIC) != 0, PIC)
PIO <- prune_samples(sample_data(nosing)$SiteCode=="PIO", nosing)
PIO <- prune_taxa(taxa_sums(PIO) != 0, PIO)
#STE <- prune_samples(sample_data(nosing)$SiteCode=="STE", nosing)
#STE <- prune_taxa(taxa_sums(STE) != 0, STE)
TEP <- prune_samples(sample_data(nosing)$SiteCode=="TEP", nosing)
TEP <- prune_taxa(taxa_sums(TEP) != 0, TEP)
TNC <- prune_samples(sample_data(nosing)$SiteCode=="TNC", nosing)
TNC <- prune_taxa(taxa_sums(TNC) != 0, TNC)
#URB <- prune_samples(sample_data(nosing)$SiteCode=="URB", nosing)
#URB <- prune_taxa(taxa_sums(URB) != 0, URB)
WLF <- prune_samples(sample_data(nosing)$SiteCode=="WLF", nosing)
WLF <- prune_taxa(taxa_sums(WLF) != 0, WLF)

RURsiteList <- list(BUF=BUF, COG=COG, COW=COW, DAL=DAL, DEB=DEB,
KAR=KAR, PHY=PHY, PIC=PIC, PIO=PIO,
          TEP=TEP, TNC=TNC, WLF=WLF)

#clean up
garbage <- c("BUF", "COG", "COW", "DAL", "DEB", "KAR", "PHY", "PIC",
"PIO", "TEP", "TNC", "WLF")
rm(list = garbage)

```

```{r noJuly, eval=FALSE, include=FALSE, results="hide"}
# remove July samples
```

```
nojuly <- prune_samples(sample_data(urb.nosing)$Date != “07.14.2015”,
urb.nosing)
nojuly <- prune_taxa(taxa_sums(nojuly) != 0, nojuly)

#nojuly.des <- phyloseq_to_deseq2(nojuly, ~ Date + Nearby_Veg)
# calculate geometric means prior to estimate size factors if error
with too many zeros
#gm_mean = function(x, na.rm=TRUE){
 # exp(sum(log(x[x > 0]), na.rm=na.rm) / length(x))
#}
#geoMeans = apply(counts(nojuly.des), 1, gm_mean)
nojuly.des <- estimateSizeFactors(nojuly.des)#, geoMeans = geoMeans)
nojuly.des <- estimateDispersions(nojuly.des, fitType=”local”,
maxit=260)
# perform DESeq2 variance stabilization instead of rarefying
nojuly.vst <- getVarianceStabilizedData(nojuly.des)
# Save the untransformed data as a separate variable so you can go back
to it
nojuly0 <- nojuly
otu_table(nojuly) <- otu_table(nojuly.vst, taxa_are_rows=TRUE)
nojuly.vst <- nojuly
nojuly <- nojuly0
rm(nojuly0)
# Set values below zero, to zero
otu_table(nojuly.vst)[otu_table(nojuly.vst) < 0.0] <- 0

#no july samples
UAord <- ordinate(nojuly, method = “PCoA”, distance = “horn”)
plot_ordination(urb.nosing.vst, UAord, type = “samples”, color =
“Date”, shape = “Nearby_Veg”) +
 geom_point(size=3)
        #label = “SampleID”, title = “Quality filtered, all urban sites,
no July (PCoA, horn)”) +
ggsave(“ordALL_nearbyveg_nojuly_horn.pdf”, device=”pdf”, width=5,
height=3.25, units=”in”, useDingbats=FALSE)
ggsave(“ordALL_nearbyveg_nojuly_horn.png”, device=”png”, width=6.5,
height=5, units=”in”)


```

```{r prevFilt, eval=FALSE, include=FALSE, results=”hide”}
# Prevalence filtering
# Define prevalence of each taxa
# (in how many samples did each taxa appear at least once)
prev0 <- apply(X = otu_table(urb.nosing),
        MARGIN = ifelse(taxa_are_rows(urb.nosing), yes = 1, no = 2),
        FUN = function(x){sum(x > 0)})
prevdf <- data.frame(Prevalence = prev0,
           TotalAbundance = taxa_sums(urb.nosing),
```

```
           tax_table(urb.nosing))
write.csv(prevdf, "taxaPrevalence.csv")
keepPhyla <- table(prevdf$Phylum)[(table(prevdf$Phylum) > 5)]
prevdf1 <- subset(prevdf, Phylum %in% names(keepPhyla))
# Define prevalence threshold as 5% of total samples
prevalenceThreshold <- 0.05 * nsamples(urb.nosing)
prevalenceThreshold

## [1] 18
# Execute prevalence filter, using `prune_taxa()` function
ps1 = prune_taxa((prev0 > prevalenceThreshold), urb.nosing)
ps1

## phyloseq-class experiment-level object
## otu_table()  OTU Table:     [ 353 taxa and 360 samples ]
## sample_data() Sample Data:    [ 360 samples by 14 sample variables ]
## tax_table()  Taxonomy Table:  [ 353 taxa by 6 taxonomic ranks ]
## phy_tree()  Phylogenetic Tree: [ 353 tips and 351 internal nodes ]

ggplot(prevdf1, aes(TotalAbundance, Prevalence, color = Phylum)) +
 geom_hline(yintercept = prevalenceThreshold, alpha = 0.5, linetype =
2) +
 geom_point(size = 2, alpha = 0.7) +
 scale_y_log10() + scale_x_log10() +
 xlab("Total Abundance") +
 facet_wrap(~Phylum) +
 theme(legend.position="none", text=element_text(size=11))
ggsave("prevalenceFiltered.pdf", device="pdf", width=6.5, height=4,
units="in", useDingbats=FALSE)

# below are taxa identified by the prevalence filter
# this is Legionella sp. (putatively L. fallonii, isolated from ship
air conditioning) present only at BLA-9-15
get_sample(urbair,
"TACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGAGTGCGTAGGTGGTTAATTAA
GTTATCTGTAAAATCCCTGGGCTCAACCTGGGCAGGTCAGATAATACTGGTTAACTCGAGTATG
GGAGAGGGTAGTGGAATTTCCGGTGTAGCGGTGAAATGCGTAGAGATCGGAAGGAACACCAGTG
GCGAAGGCGGCTACCTGGACTAATACTGACACTGAGGCACGAAAGCGTGGGGAGCAAACAGG")
# this is Collinsella   tanakaei (isolated from human feces) present
only at WEG-8-4
get_sample(urbair,
"TACGTAGGGGGCGAGCGTTATCCGGATTCATTGGGCGTAAAGCGCGCGTAGGCGGCCGCGTAG
GCGGGGGGTCAAATCCCGGGGCTCAACCCCGGTCCGCCCCCCGAACCCCGCGGCTCGGGTCCGG
TAGGGGAGGGTGGAATTCCCGGTGTAGCGGTGGAATGCGCAGATATCGGGAGGAACACCGGTGG
CGAAGGCGGCCCTCTGGGCCGAGACCGACGCTGAGGCGCGAAAGCTGGGGGAGCGAACAGG")
# Buchnera sp. (obligate endosymbionts of aphids) SKI-8-18
get_sample(urbair, "TACGGAGGGTGCTAGCGTTAATCAGAATTACTGGGCGTAAAGAG
CGCGTAGGTGGTTTTTTAAGTCAGATGTGAAATCCCTGGGCTTAACCTAGGAACTGCATTTGAA
ACTGAAATACTAGAGTATCGTAGAGGGAGGTAGAATTCTAGGTGTAGCGGTGAAATGCGTAGAT
ATCTGGAGGAATACCCGTGGCGAAAGCGGCCTCCTAAACGAATACTGACACTGAGGTGCGAAAG
```

```
CGTGGGGAGCAAACAGG")
# Aeromicrobium ponti (isolated from seawater) CEM-8-14
get_sample(urbair, "TACGTAGGGTCCGAGCGTTGTCCGGAATTATTGGGCGTAAAGGG
CTCGTAGGCGGTCTGTCGCGTCGGGAGTGAAAACTCAGGGCTCAACCCTGAGCGTGCTTCCGAT
ACGGGCAGACTAGAGGTATGCAGGGGAGAACGGAATTCCTGGTGTAGCGGTGGAATGCGCAGAT
ATCAGGAGGAACACCGGTGGCGAAGGCGGTTCTCTGGGCATTACCTGACGCTGAGGAGCGAAAG
CATGGGGAGCGAACAGG")
# Fructobacillus  fructosus (found in fructose-rich environments
such as flowers, (fermented) fruits, or bee guts, and are
characterized as fructophilic lactic acid bacteria) mostly from
CEM-9-15, a few at WLF, BUF and PIO and SAF and AMA
get_sample(urbair, "TACGTATGTCCCGAGCGTTATCCGGATTTATTGGGCGTAAAGCG
AGCGCAGACGGTTGCTTAAGTCTGAAGTGAAAGCCCACAGCTCAACTGTGGAATGGCTTTGGAA
ACTGGGCAACTTGAGTGCAGTAGAGGTAAGTGGAACTCCATGTGTAGCGGTGGAATGCGTAGAT
ATATGGAAGAACACCAGTGGCGAAGGCGGCTTACTGGACTGCAACTGACGTTGAGGCTCGAAAG
TGTGGGTAGCAAACAGG")

#ps1.des <- phyloseq_to_deseq2(ps1, ~ Date + Nearby_Veg)
# calculate geometric means prior to estimate size factors if error
with too many zeros
#gm_mean = function(x, na.rm=TRUE){
 # exp(sum(log(x[x > 0]), na.rm=na.rm) / length(x))
#}
#geoMeans = apply(counts(urb.nosing.des), 1, gm_mean)
#ps1.des <- estimateSizeFactors(ps1.des)#, geoMeans = geoMeans)
#ps1.des <- estimateDispersions(ps1.des, fitType="local", maxit=260)
# perform DESeq2 variance stabilization instead of rarefying
#ps1.vst <- getVarianceStabilizedData(ps1.des)
# Save the untransformed data as a separate variable so you can go back
to it
#ps10 <- ps1
#otu_table(ps1) <- otu_table(ps1.vst, taxa_are_rows=TRUE)
#ps1.vst <- ps1
#ps1 <- ps10
#rm(ps10)
# Set values below zero, to zero
#otu_table(ps1.vst)[otu_table(ps1.vst) < 0.0] <- 0
#ps1.tab <- getTab(ps1.vst)

UAord <- ordinate(ps1.vst, method = "PCoA", distance = "horn")
plot_ordination(ps1.vst, UAord, type = "samples", color = "Nearby_Veg",
        label = "SampleID", title = "Quality filtered, urban sites, set
1 (PCoA, horn)") +
 scale_color_manual(values=c("turquoise3", "darkgoldenrod", "grey30"))
ggsave("ordALL_nearbyveg_prevFilt_horn.pdf", device="pdf", width=6.5,
height=5, units="in", useDingbats=FALSE)
ggsave("ordALL_nearbyveg_prevFilt_horn.png", device="png", width=6.5,
height=5, units="in")

#try plotting with land cover types separate
```

```
forest <- prune_samples(sample_data(urb.nosing)$Nearby_Veg==”forest”,
urb.nosing)
prev0 <- apply(X = otu_table(forest),
        MARGIN = ifelse(taxa_are_rows(forest), yes = 1, no = 2),
        FUN = function(x){sum(x > 0)})
prevdf <- data.frame(Prevalence = prev0,
           TotalAbundance = taxa_sums(forest),
           tax_table(forest))
keepPhyla <- table(prevdf$Phylum)[(table(prevdf$Phylum) > 5)]
prevdf1 <- subset(prevdf, Phylum %in% names(keepPhyla))
prevalenceThreshold <- 0.05 * nsamples(forest)
ggplot(prevdf1, aes(TotalAbundance, Prevalence, color = Phylum)) +
 geom_hline(yintercept = prevalenceThreshold, alpha = 0.5, linetype =
2) +
 geom_point(size = 2, alpha = 0.7) +
 scale_y_log10() + scale_x_log10() +
 xlab(“Total Abundance”) +
 facet_wrap(~Phylum) +
 theme(legend.position=”none”, text=element_text(size=11))
ggsave(“prevalenceFiltered_forest.pdf”, device=”pdf”, width=6.5,
height=4, units=”in”, useDingbats=FALSE)

grass <- prune_samples(sample_data(urb.nosing)$Nearby_Veg==”grass”,
urb.nosing)
prev0 <- apply(X = otu_table(grass),
        MARGIN = ifelse(taxa_are_rows(grass), yes = 1, no = 2),
        FUN = function(x){sum(x > 0)})
prevdf <- data.frame(Prevalence = prev0,
           TotalAbundance = taxa_sums(grass),
           tax_table(grass))
keepPhyla <- table(prevdf$Phylum)[(table(prevdf$Phylum) > 5)]
prevdf1 <- subset(prevdf, Phylum %in% names(keepPhyla))
prevalenceThreshold <- 0.05 * nsamples(grass)
ggplot(prevdf1, aes(TotalAbundance, Prevalence, color = Phylum)) +
 geom_hline(yintercept = prevalenceThreshold, alpha = 0.5, linetype =
2) +
 geom_point(size = 2, alpha = 0.7) +
 scale_y_log10() + scale_x_log10() +
 xlab(“Total Abundance”) +
 facet_wrap(~Phylum) +
 theme(legend.position=”none”, text=element_text(size=11))
ggsave(“prevalenceFiltered_grass.pdf”, device=”pdf”, width=6.5,
height=4, units=”in”, useDingbats=FALSE)

paved <- prune_samples(sample_data(urb.nosing)$Nearby_Veg==”paved”,
urb.nosing)
prev0 <- apply(X = otu_table(paved),
        MARGIN = ifelse(taxa_are_rows(paved), yes = 1, no = 2),
        FUN = function(x){sum(x > 0)})
prevdf <- data.frame(Prevalence = prev0,
```

```
          TotalAbundance = taxa_sums(paved),
          tax_table(paved))
keepPhyla <- table(prevdf$Phylum)[(table(prevdf$Phylum) > 5)]
prevdf1 <- subset(prevdf, Phylum %in% names(keepPhyla))
prevalenceThreshold <- 0.05 * nsamples(paved)
ggplot(prevdf1, aes(TotalAbundance, Prevalence, color = Phylum)) +
 geom_hline(yintercept = prevalenceThreshold, alpha = 0.5, linetype =
2) +
 geom_point(size = 2, alpha = 0.7) +
 scale_y_log10() + scale_x_log10() +
 xlab("Total Abundance") +
 facet_wrap(~Phylum) +
 theme(legend.position="none", text=element_text(size=11))
ggsave("prevalenceFiltered_paved.pdf", device="pdf", width=6.5,
height=4, units="in", useDingbats=FALSE)


```

```{r plotMap, include=TRUE, results="hide"}
SFC <- prune_samples(sample_data(urbair)$SiteCode=="SFC", urbair)
# plot the basemap using sample location long and lat
sampLong <- as.numeric(as.character(unique(sample_data(SFC)$Long)))
sampLat <- as.numeric(as.character(unique(sample_data(SFC)$Lat)))
map <- get_map(location=c(lon= sampLong, lat= sampLat), zoom = 10,
maptype = "satellite", source = "google") #get map

#plot Sample Locations
sitePts <- as.data.frame(cbind(as.character(unique(sample_
data(prune)$SiteCode)),
         as.numeric(as.character(unique(sample_data(prune)$Lat))),
         as.numeric(as.character(unique(sample_data(prune)$Long))),
         as.character(unique(sample_data(prune)$Urbanicity))
         ), stringsAsFactors=FALSE)
colnames(sitePts) <- c("SiteCode", "Lat", "Long", "Location")
sitePts$Veg <- sample_data(prune)$Nearby_Veg[match(sitePts$SiteCode,
sample_data(prune)$SiteCode)]
sitePts$Lat <- as.numeric(sitePts$Lat)
sitePts$Long <- as.numeric(sitePts$Long)
sitePts$Location <- sample_
data(prune)$Urbanicity[match(sitePts$SiteCode, sample_
data(prune)$SiteCode)]
sitePts$Location <- factor(sitePts$Location, levels=c("urban",
"rural"))

ggmap(map, darken=c(0.25, "black"), extent="panel") +
 geom_point(data=sitePts, aes(x=Long, y=Lat, group=Veg, color=Veg,
shape=Location), size=2, alpha=1) +
 scale_color_manual(values=c("turquoise3", "darkgoldenrod1", "grey20"))
+
```

```
 geom_text(data=sitePts, aes(x=Long, y=Lat, label=SiteCode, vjust=2,
group=Veg, color=Veg), size=2, check_overlap = TRUE) +
 xlim(c(-123.25, -122.78)) +
 ylim(c(43.95, 44.27)) #+ theme(legend.position = "none")

ggsave("sitesMap_rurUrb_labels.pdf", device="pdf", width=6.5, height=4,
units="in", useDingbats=FALSE)

ggsave("sitesMap_rurUrb_labels.png", device="png", width=6.5, height=4,
units="in")
```

Vegetation Analysis

```{r vegEval, include=FALSE, results="hide"}
vegStats <- data.frame(sample_data(urb.sing)$SiteCode, sample_data(urb.
sing)$Nearby_Veg,
            sample_data(urb.sing)$Context, sample_data(urb.
sing)$green50, sample_data(urb.sing)$green100,
            sample_data(urb.sing)$green200, sample_data(urb.
sing)$green400, sample_data(urb.sing)$green800,
            sample_data(urb.sing)$structDiv, sample_data(urb.
sing)$vegDiv)
vegStats <- unique(vegStats)
colnames(vegStats) <- c("SiteCode", "LandCover", "Context", "green50",
"green100", "green200", "green400", "green800", "structDiv", "vegDiv")
vegStats <- vegStats[with(vegStats, order(LandCover, green800)), ]
write.csv(vegStats, "vegStats.csv")
min(vegStats$green50)
max(vegStats$green50)
max(vegStats$green50[vegStats$LandCover == "paved"])
min(vegStats$green800)
max(vegStats$green800)

mean(vegStats$green50[vegStats$LandCover == "paved"])
mean(vegStats$green50[vegStats$LandCover == "grass"])
mean(vegStats$green50[vegStats$LandCover == "forest"])
mean(vegStats$green800[vegStats$LandCover == "paved"])
mean(vegStats$green800[vegStats$LandCover == "grass"])
mean(vegStats$green800[vegStats$LandCover == "forest"])

se <- function(x) sqrt(var(x)/length(x))
vegMelt <- melt(vegStats[,2:8], id.vars=c("LandCover", "Context"))
avgs <- ddply(vegMelt, .(LandCover, Context, variable), summarize,
       mean = mean(value),
       se = se(value)
       )

# VEG COVER STATS - plot bar graph with standard deviation as error
bars
```

```
ggplot(avgs) +
 geom_bar(aes(x=variable, y=mean, fill=variable), stat="identity",
position=position_dodge(width=0.8), width=0.55) +
 geom_errorbar(aes(x=variable, ymin=mean-se, ymax=mean+se),
position=position_dodge(width=0.8), width=0.5, size=0.15) +
 scale_fill_manual(labels = c("green50" = "50m", "green100" = "100m",
"green200" = "200m", "green400" = "400m", "green800" = "800m"),
          values=rainbow(5, s = 0.8, v = 0.6, start = 0.2, end = 0.5,
alpha = 1)) +
 facet_grid(LandCover ~ Context) +
 labs(x="Buffer Size", y="Average Vegetation Proportion", fill="Buffer
Size") +
 scale_x_discrete(labels=c("green50" = "50m", "green100" = "100m",
"green200" = "200m", "green400" = "400m", "green800" = "800m")) +
 theme(axis.text=element_text(size=11), axis.title=element_
text(size=12),
    legend.text=element_text(size=11), legend.title=element_
text(size=12), legend.position = "none")

ggsave("bar_vegStats.pdf", device="pdf", width=6.5, height=4,
units="in", useDingbats=FALSE)
ggsave("bar_vegStats.png", device="png", width=6.5, height=4,
units="in")

# VEG STRUCTURAL & SPECIES DIVERSITY - plot bar graph with standard
deviation as error bars
vegMelt <- melt(vegStats[,c(2:3,9:10)], id.vars=c("LandCover",
"Context"))
avgs <- ddply(vegMelt, .(LandCover, Context, variable), summarize,
       mean = mean(value),
       se = se(value)
       )
avgs <- subset(avgs, LandCover %in% c("forest", "grass"))

ggplot(avgs) +
 geom_bar(aes(x=variable, y=mean, fill=variable), stat="identity",
position=position_dodge(width=0.8), width=0.55) +
 geom_errorbar(aes(x=variable, ymin=mean-se, ymax=mean+se),
position=position_dodge(width=0.8), width=0.5, size=0.15) +
 scale_fill_manual(labels = c("structDiv" = "Structural", "vegDiv" =
"Species"),
          values=rainbow(2, s = 0.8, v = 0.6, start = 0.1, end = 0.6,
alpha = 1)) +
 facet_grid(LandCover ~ Context) +
 labs(x="Diversity (Shannon)", y="Average Vegetation Proportion",
fill="Diversity") +
 scale_x_discrete(labels=c("structDiv" = "Structural", "vegDiv" =
"Species")) +
 theme(axis.text=element_text(size=11), axis.title=element_
text(size=12),
```

```
    legend.text=element_text(size=11), legend.title=element_
text(size=12), legend.position = "none")

ggsave("bar_vegStats2.pdf", device="pdf", width=6.5, height=4,
units="in", useDingbats=FALSE)
ggsave("bar_vegStats2.png", device="png", width=6.5, height=4,
units="in")

```

Meteorological Analysis

``` {r trimTime, eval=FALSE, echo=FALSE}
####### trim time data --------
iButList <- list.files("/Users/Gwynne/UO/PhD work/dissertation/raw data/
iButton/iButtonData_calibrate")
for (i in iButList) {
 filepath <- file.path("/Users/Gwynne/UO/PhD work/dissertation/raw data/
iButton/iButtonData_calibrate",
             paste(i,sep="/"))
 temp <- read.csv(filepath, header=TRUE, skip=13)
 temp[,1] <- mdy_hms(temp[,1])
 temp.x <- xts(temp[,-1], order.by=temp[,1])
 temp.x <- temp.x['20150709/20150711 12:00']
 temp <- fortify(temp.x)
 trimBut <- paste("/Users/Gwynne/UO/PhD work/dissertation/processed
data/iButtonData_calibrate", i, sep="/")
 write.csv(temp, trimBut)
}

```

``` {r meteorEval, eval=FALSE, echo=FALSE}

foldList <- list.files("/Users/gwynhwyfer/Documents/PhD work/
dissertation/chapter3_urbanMicrobiome/processed data/iButton_processed
with site codes/")

totBut <- data.frame(Index=character(),
           Unit=factor(),
           Value=numeric(),
           Location=factor(),
           iButton=factor(),
           Period=factor(),
           stringsAsFactors=TRUE)
for (h in foldList) {
 foldpath <- file.path("/Users/gwynhwyfer/Documents/PhD work/
dissertation/chapter3_urbanMicrobiome/processed data/iButton_processed
with site codes",
             paste(h,sep=""))
```

```
 iButList <- list.files(foldpath)
 allBut <- data.frame(Index=character(),
            Unit=factor(),
            Value=numeric(),
            Location=factor(),
            iButton=factor(),
            stringsAsFactors=TRUE)
 for (i in iButList) {
  filepath <- file.path("/Users/gwynhwyfer/Documents/PhD work/
dissertation/chapter3_urbanMicrobiome/processed data/iButton_processed
with site codes",
             paste(h,i,sep="/"))
  iBut <- read.csv(filepath, header=TRUE,
            stringsAsFactors=TRUE)
  iBut <- iBut[,-1]
  iBut$iButton <- "iButton1"
  iBut$iButton <- gsub("iButton1", paste(i,sep=""), iBut[,5])
  iBut$iButton <- gsub(".csv","", iBut[,5])
  iBut[,1] <- mdy_hm(iBut[,1])
  allBut <- rbind(allBut, iBut)
 }
 allBut$Period <- "9-29"
 allBut$Period <- gsub("9-29", paste(h,sep=""), allBut[,6])
 allBut$Period <- gsub("iButtonData_","", allBut[,6])
 allBut$Period <- gsub("-2015","", allBut[,6])
 allBut <- subset(allBut, Location != "CONTROL") # remove controls
 allBut <- subset(allBut, Location != "UNKNOWN")
 pdf(paste(h, ".pdf", sep=""), height=7, width=8.5,
  useDingbats=FALSE)
 par(mfrow=c(1,1), pty="m", oma=c(0,0,0,0), mar=c(4,4,2,2))
 print(ggplot(allBut, aes(x=Index, y=Value, colour=Location)) + geom_
line())
 dev.off()
 totBut <- rbind(totBut, allBut)
}

#totBut$Location <- as.factor(totBut$Location) #don't know if I want
this as a factor
totBut.sub <- subset(totBut, Location != "CONTROL") # remove controls
totBut.sub <- subset(totBut.sub, Location != "UNKNOWN") # remove "lost"
iButtons
meta <- read.csv("/Users/gwynhwyfer/Documents/PhD work/dissertation/
chapter3_urbanMicrobiome/processed data/mapFile_2-15-2017.csv")
totBut.sub$Type <- meta[match(totBut.sub$Location, meta$SiteCode), 6]
totBut.sub$iButton <- as.factor(totBut.sub$iButton)
totBut.sub$Period <- as.factor(totBut.sub$Period)

# convert to zoo/xts object to plot time series
totBut.sub$Time <- hour(totBut.sub$Index) + minute(totBut.sub$Index)/60
totBut.sub$Time <- format(as.POSIXct(totBut.sub$Time*3600,
```

```
origin="2001-01-01", "GMT"), "%H:%M")
#totBut.sub$Time <- hm(totBut.sub$Time)
#totBut.sub$Time <- parse_date_time(totBut.sub$Time, "HM")
write.csv(totBut.sub, "/Users/gwynhwyfer/Documents/PhD work/
dissertation/chapter3_urbanMicrobiome/processed data/totBut.csv")
totBut.noOuts <- totBut.sub[!(totBut.sub$Location=="WEG" & totBut.
sub$Period=="8-3"), ] #remove crazy iButton
totBut.noOuts <- totBut.noOuts[!(totBut.noOuts$Value>=42), ] #and
remove random outlier
sort(totBut.noOuts$Value)[1:10]

ggplot(totBut.sub, aes(x=Time, y=Value, colour=Period)) + geom_line() +
     facet_wrap("Location")
ggplot(totBut.sub, aes(x=Time, y=Value, colour=Location)) + geom_line()
+
     facet_wrap("Period")
ggplot(totBut.noOuts, aes(x=Time, y=Value, colour=Location)) +
 geom_point() +
 scale_y_continuous(limits=c(35, 45))
ggplot(totBut.noOuts, aes(x=Type, y=Value, colour=Type)) +
 geom_boxplot() +
 facet_wrap(~Period)

iButSumm <- data.frame(ddply(totBut.noOuts, .(Location, Period),
summarize, meanTemp = mean(Value, na.rm = TRUE),
               minTemp = min(Value, na.rm = TRUE), maxTemp = max(Value,
na.rm = TRUE)))
write.csv(iButSumm, "/Users/gwynhwyfer/Documents/PhD work/dissertation/
chapter3_urbanMicrobiome/processed data/iButSumm.csv")
mean(iButSumm$meanTemp) #average mean temp
min(iButSumm$minTemp)
max(iButSumm$maxTemp)

iButSumm.date <- data.frame(ddply(totBut.noOuts, .(Period), summarize,
meanTemp = mean(Value, na.rm = TRUE),
               minTemp = min(Value, na.rm = TRUE), maxTemp = max(Value,
na.rm = TRUE)))
iButSumm.site <- data.frame(ddply(totBut.noOuts, .(Location),
summarize, meanTemp = mean(Value, na.rm = TRUE),
               minTemp = min(Value, na.rm = TRUE), maxTemp = max(Value,
na.rm = TRUE)))
iButSumm.site[order(iButSumm.site$meanTemp),]
iButSumm.lc <- data.frame(ddply(totBut.noOuts, .(Type, Period),
summarize, meanTemp = mean(Value, na.rm = TRUE),
               minTemp = min(Value, na.rm = TRUE), maxTemp = max(Value,
na.rm = TRUE)))

write.csv(iButSumm.lc, "tempSummary.csv")
iButSumm.melt <- melt(iButSumm, id.vars=c("Period", "Location"),
measure.vars="meanTemp") # convert to long format
```

```
ggplot(data=iButSumm.melt) +
 geom_line(aes(x=Period, y=value, color=factor(Location),
group=interaction(Location))) +
 facet_wrap(~Location)

ggplot(data=iButSumm, aes(x=Period, y=meanTemp), fill=Location) +
 geom_point() +
 scale_fill_manual(values=rainbow(36))
length(unique(iButSumm$Location))
```

Bacterial Analysis
Overview

```` {r basicInfo, include=FALSE, results="hide"}

# access basic info (with single/doubletons and extra samples)
sum(sample_sums(urb.sing)) # total number of reads, including single/
doubletons
sum(sample_sums(urb.nosing)) # total number of reads, excluding single/
doubletons
sum(sample_sums(urb.sing))-sum(sample_sums(urb.nosing))
sort(sample_sums(urb.nosing)) # numbers of reads in each sample
ntaxa(urb.nosing) # total number of taxa, excluding single/doubletons
length(get_taxa_unique(urb.nosing, "Phylum")) # number of different
Phyla represented
rev(sort(taxa_sums(urb.sing)))[1:10] # top ten OTU abundances
tax_table(urb.sing)[row.names(tax_table(urb.sing))=="TACGGAGGGAG
CTAGCGTTATTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGCTTTGTAAGTAAGAGGTGAA
AGCCCAGAGCTCAACTCTGGAATTGCCTTTTAGACTGCATCGCTTGAATCATGGAGAGGTCAGT
GGAATTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAAGAACACCAGTGGCGAAGGCGGCT
GACTGGACATGTATTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGG",]
rev(sort(taxa_sums(urb.nosing)))[1]/sum(sample_sums(urb.nosing)) #
proportional abundance of top taxon

phyGlom <- tax_glom(urb.nosing, taxrank="Phylum") #agglomerate at
phylum level
taxa_sums(phyGlom)/sum(sample_sums(phyGlom)) # proportional abundance
of phylum
top10phy <- names(sort(taxa_sums(phyGlom), TRUE))
top10 <- prune_taxa(top10phy, phyGlom)
top10_sums <- data.frame(taxa_sums(top10)/sum(sample_sums(top10)))
# get taxonomic Families of top 10
top10tax <- tax_table(top10)[, c("Phylum", "Class")]
top10matrix <- data.frame(as(top10tax, "matrix"))
top10matrix <- cbind(top10matrix, top10_sums)
names(top10matrix)[names(top10matrix) == "taxa_sums.top10."] <-
"Abundance"
top10matrix <- top10matrix[order(as.numeric(as.character(-
top10matrix[,3]))), ]
```

```
write.csv(top10matrix, “phyAbund.csv”)

classGlom <- tax_glom(urb.nosing, taxrank=”Class”) #agglomerate at
class level
taxa_sums(classGlom)/sum(sample_sums(classGlom)) # proportional
abundance of class
top10phy <- names(sort(taxa_sums(classGlom), TRUE))
top10 <- prune_taxa(top10phy, classGlom)
top10_sums <- data.frame(taxa_sums(top10)/sum(sample_sums(top10)))
# get taxonomic Families of top 10
top10tax <- tax_table(top10)[, c(“Phylum”, “Class”)]
top10matrix <- data.frame(as(top10tax, “matrix”))
top10matrix <- cbind(top10matrix, top10_sums)
names(top10matrix)[names(top10matrix) == “taxa_sums.top10.”] <-
“Abundance”
top10matrix <- top10matrix[order(as.numeric(as.character(-
top10matrix[,3]))), ]
write.csv(top10matrix, “classAbund.csv”)

# What about the total reads per sample, and what does the distribution
look like?
readsumsdf <- data.frame(nreads = sort(taxa_sums(urb.nosing), TRUE),
sorted = 1:ntaxa(urb.nosing), type = “OTUs”)
readsumsdf <- rbind(readsumsdf, data.frame(nreads = sort(sample_
sums(urb.nosing), TRUE), sorted = 1:nsamples(urb.nosing), type =
“Samples”))
ggplot(readsumsdf, aes(x = sorted, y = nreads)) + geom_bar(stat =
“identity”) +
 ggtitle(“Total number of reads”) + scale_y_log10() + facet_wrap(~type,
1, scales = “free”)
#dev.off()

#clean up
rm(readsumsdf)

```


### Alpha diversity
Richness across all urban sites is shown in Figure \ref{fig:alpha},
measured using the Shannon index. We found no significant pattern to
variation in diversity by nearby vegetation type.

```{r alphaDiv, include=TRUE, results=”hide”, fig.width=5, fig.height=4,
fig.show=”hold”, fig.align=”center”, fig.cap=”\\label{fig:alpha}Alpha
diversity across dates by nearby vegetation type.”}
## CAN’T COMPARE DIVERSITY INDICES BC OF DIFFERENCE IN SAMPLE SIZES
## USE RAREFACTION CURVES INSTEAD per Bowers
# get raw OTU and sample data tables out of phyloseq, samples as rows
pruneTab <- getTab(urb.sing)
pruneEnv <- sample_data(urb.sing)
```

```
pruneTab[1:5, 1:5] # check
raremin <- min(rowSums(pruneTab))
expTaxa <- data.frame("Expected_Taxa" = rarefy(pruneTab, raremin))
expTaxa$Date <- as.factor(sample_data(urb.sing)$Date[match(row.
names(expTaxa), sample_data(urb.sing)$SampleID)])
expTaxa$SiteCode <- as.factor(sample_data(urb.sing)$SiteCode[match(row.
names(expTaxa), sample_data(urb.sing)$SampleID)])
expTaxa$Nearby_Veg <- as.factor(sample_data(urb.sing)$Nearby_
Veg[match(row.names(expTaxa), sample_data(urb.sing)$SampleID)])
expTaxa$Context <- as.factor(sample_data(urb.sing)$Context[match(row.
names(expTaxa), sample_data(urb.sing)$SampleID)])
expTaxa$Reads <- sample_sums(urb.sing)[match(row.names(expTaxa),
names(sample_sums(urb.sing)))]

#pdf("rarecurve_allSites.pdf", height = 9, width = 6.5)
#rc <- rarecurve(pruneTab, step = 20, sample = raremin, col = "blue",
cex = 0.6)
#dev.off()
write.csv(expTaxa, "rarecurve_expTaxa.csv")

##TEST FOR DIFFERENCES IN DIVERSITY?
min(expTaxa$Expected_Taxa)
max(expTaxa$Expected_Taxa)
hist(expTaxa) #Distribution of the alpha diversity index is close to
normal
summary(aov(Expected_Taxa ~ Date, data = expTaxa))
summary(aov(Expected_Taxa ~ SiteCode, data = expTaxa))
summary(aov(Expected_Taxa ~ Nearby_Veg, data = expTaxa))
summary(aov(Expected_Taxa ~ Context, data = expTaxa))
#summary(aov(Expected_Taxa ~ Urbanicity, data = expTaxa))

```

```{r dateBar, include=TRUE, results="hide", fig.width=5, fig.height=4,
fig.show="hold", fig.align="center", fig.cap="\\label{fig:barAll}Barplot
showing composition by date for all sites aggregated."}

# transform only urban sample counts to relative abundance by sample
urb.nosingRel <- transform_sample_counts(urb.nosing, function(x) 100 *
x/sum(x))

# top 25 taxa overall (urban sites) - output csv
top25otus <- names(sort(taxa_sums(urb.nosingRel), TRUE)[1:25])
top50otus <- names(sort(taxa_sums(urb.nosingRel), TRUE)[1:50])
top100otus <- names(sort(taxa_sums(urb.nosingRel), TRUE)[1:100])
top25 <- prune_taxa(top25otus, urb.nosingRel)
top25_sums <- data.frame(taxa_sums(top25))
# get taxonomic Families of top 25
top25tax <- tax_table(top25)[, c("Phylum", "Class", "Order", "Family",
"Genus", "Species")]
```

```
top25matrix <- data.frame(as(top25tax, “matrix”))
top25matrix <- cbind(top25matrix, top25_sums)
names(top25matrix)[names(top25matrix) == “taxa_sums.top25.”] <-
“Abundance”
top25matrix <- top25matrix[order(as.numeric(as.character(-
top25matrix[,7]))), ]
write.csv(top25matrix, “top25_all.csv”)

genus.merge <- tax_glom(urb.nosing, taxrank=”Genus”)
genus.mergeRel <- transform_sample_counts(genus.merge, function(x) 100
* x/sum(x))
fam.merge <- tax_glom(urb.nosing, taxrank=”Family”)
fam.mergeRel <- transform_sample_counts(fam.merge, function(x) 100 * x/
sum(x))
order.merge <- tax_glom(urb.nosing, taxrank=”Order”)
class.merge <- tax_glom(urb.nosing, taxrank=”Class”)
class.mergeRel <- transform_sample_counts(class.merge, function(x) 100
* x/sum(x))

## stacked barplots to compare proportional composition for all urban
sites (family level) by Site
sort.class <- sort(tapply(taxa_sums(urb.nosing), tax_table(urb.nosing)
[, “Family”], sum), TRUE)
length(sort.class)
top.class <- sort.class[1:25] #what are the top 25 most abundant
Families?
bottom.class <- sort.class[26:length(sort.class)]
urb.nosing1 <- subset_taxa(urb.nosing, Family %in% names(top.class))
#get top 25 most abundant Family
urb.nosing2 <- subset_taxa(urb.nosing, Family %in% names(bottom.class))
#get all other taxa
urb.nosing2 <- merge_taxa(urb.nosing, taxa_names(urb.nosing2),
archetype=1) #merge all other taxa into Family “Other”
tax_table(urb.nosing2)[,5][is.na(tax_table(urb.nosing2)[,5])] <-
“Other”
urb.nosing2 <- tax_glom(urb.nosing2, taxrank=”Family”)
get_taxa_unique(urb.nosing2, “Family”)

##tried it with SVs instead of merged families
#sort.sv <- sort(taxa_sums(urb.nosing), TRUE)[101:length(taxa_sums(urb.
nosing))]
#otherTaxa <- prune_taxa(names(sort.sv), urb.nosing)
#urb.nosing2 <- merge_taxa(urb.nosing, taxa_names(otherTaxa),
archetype=1) #merge all other taxa into Family “Other”
#taxa_names(otherTaxa)[1]
#tax_table(urb.nosing2)[grep(taxa_names(otherTaxa)[1], row.names(tax_
table(urb.nosing2))),5] <- “Other”
#tax_table(urb.nosing2)[,5]

## stacked barplots to compare proportional composition for all urban
```

```
sites (family level) by Date
urb.nosing.m <- merge_samples(urb.nosing2, “Date”)
sample_data(urb.nosing.m)$Date <- levels(factor(sample_data(urb.
nosing2)$Date))
urb.nosing.m <- transform_sample_counts(urb.nosing.m, function(x) 100 *
x/sum(x))
df_long <- psmelt(urb.nosing.m) # first change to long format
old.lvl <- levels(df_long$Family)
df_long$Family <- factor(df_long$Family, levels=c(“Other”, sort(old.
lvl[old.lvl!=”Other”], decreasing=F)))
ggplot(df_long, aes(x=Date, y=Abundance, fill = Family)) +
 geom_bar(stat=”identity”, position=”stack”) +
 ylab(“Percentage of Sequences”) +
 scale_fill_manual(values=famPal, na.value=”darkgrey”) +
 scale_color_manual(values=famPal, na.value=”darkgrey”) +
 theme(axis.title=element_text(size=12), legend.text=element_
text(size=8), legend.key.size = unit(0.16,”in”))

ggsave(“date_stackedBar_allbyFamily.pdf”, device=”pdf”, width=7,
height=4, units=”in”, useDingbats=FALSE)
ggsave(“date_stackedBar_allbyFamily.png”, device=”png”, width=6.5,
height=4, units=”in”)
```


### Community composition by site location
```{r siteBar, include=TRUE, results=”hide”, fig.width=6.5, fig.
height=6.5, fig.show=”hold”, fig.align=”center”, fig.cap=”\\
label{fig:barSite}Barplot showing bacterial family composition
aggregated by site location.”}
siteList.rel <- llply(siteList, function(x){transform_sample_counts(x,
function(x) 100 * x/sum(x))})
dateList.rel <- llply(dateList, function(x){transform_sample_counts(x,
function(x) 100 * x/sum(x))})

RURsiteList.rel <- llply(RURsiteList, function(x){transform_sample_
counts(x, function(x) 100 * x/sum(x))})

## stacked barplots to compare proportional composition (colored at
family level) of top 25 taxa for every site BY DATE
##FIGURE OUT HOW TO GROUP BY SITE TYPE
sort_plot <- function(x){
 top25 <- sort(taxa_sums(x), TRUE)[1:25]
 top25 <- prune_taxa(names(top25), x)
 top25_sums <- data.frame(taxa_sums(top25))
 # get taxonomic Families of top 25
 top25tax <- tax_table(top25)[, c(“Phylum”, “Class”, “Order”, “Family”,
“Genus”)]
 top25matrix <- data.frame(as(top25tax, “matrix”))
 top25matrix <- cbind(top25matrix, top25_sums)
 # reorder levels by Nearby_Veg
```

```
 sample_data(top25)$SampleID <- factor(
  sample_data(top25)$SampleID, levels=sample_
data(top25)$SampleID[order(sample_data(top25)$Nearby_Veg)])
 plot_bar(top25, x="SampleID", fill = "Family") +
  ylab("Percentage of Sequences") +
  geom_bar(aes(fill=Family, color=Family), stat="identity",
position="stack") +
  scale_fill_manual(values=famPal, na.value="darkgrey") +
  scale_color_manual(values=famPal, na.value="darkgrey") +
  #scale_x_discrete(labels=c("7.14","8.4","8.14","8.18",
"8.25","9.1","9.15","9.22", "9.29")) +
  ylim(0,80) +
  theme(legend.position = "none", axis.text=element_text(size=11),
axis.title=element_text(size=12))
}

siteList.25 <- llply(siteList.rel, sort_plot)
siteList.25[1:9]
for (i in seq_along(siteList.25)){
 ggsave(paste0(names(siteList.25)[i], "_stackedBar_bySite_top50.pdf"),
plot=siteList.25[[i]], device="pdf", width=3, height=3, units="in")
}


dateList.25 <- llply(dateList.rel, sort_plot)
dateList.25[1:9]
for (i in seq_along(dateList.25)){
 ggsave(paste0(names(dateList.25)[i], "_stackedBar_byDate_top25.pdf"),
plot=dateList.25[[i]], device="pdf", width=3, height=3, units="in")
}

RURsiteList.25 <- llply(RURsiteList.rel, sort_plot)
for (i in seq_along(RURsiteList.25)){
 ggsave(paste0(names(RURsiteList.25)[i], "_stackedBar_byDate_top50_RUR.
pdf"), plot=RURsiteList.25[[i]], device="pdf", width=3, height=3,
units="in")
}

#clean up
rm(top25otus, top25_sums, top25tax, top25matrix)

```

``` {r transformData, results="hide"}
# transform urban only sample counts for downstream analysis using vst
# this design measures the effect of Nearby_Veg, controlling for
sampling Date differences. In order to benefit from the default settings
of the package, you should put the variable of interest at the end of
the formula and make sure the control level is the first level.
urb.nosing.des <- phyloseq_to_deseq2(urb.nosing, ~ Date + Nearby_Veg)
```

```
# calculate geometric means prior to estimate size factors if error
with too many zeros
#gm_mean = function(x, na.rm=TRUE){
 # exp(sum(log(x[x > 0]), na.rm=na.rm) / length(x))
#}
#geoMeans = apply(counts(urb.nosing.des), 1, gm_mean)
urb.nosing.des <- estimateSizeFactors(urb.nosing.des)#, geoMeans =
geoMeans)
urb.nosing.des <- estimateDispersions(urb.nosing.des, fitType="local",
maxit=260)
# perform DESeq2 variance stabilization instead of rarefying
urb.nosing.vst <- getVarianceStabilizedData(urb.nosing.des)
# Save the untransformed data as a separate variable so you can go back
to it
urb.nosing0 <- urb.nosing
otu_table(urb.nosing) <- otu_table(urb.nosing.vst, taxa_are_rows=TRUE)
urb.nosing.vst <- urb.nosing
urb.nosing <- urb.nosing0
rm(urb.nosing0)
# Set values below zero, to zero
otu_table(urb.nosing.vst)[otu_table(urb.nosing.vst) < 0.0] <- 0

# transform all (urban + rural) sample counts for downstream analysis
using vst
nosing.des <- phyloseq_to_deseq2(nosing, ~ Date + Context)
nosing.des <- estimateSizeFactors(nosing.des)#, geoMeans = geoMeans)
nosing.des <- estimateDispersions(nosing.des, fitType="local",
maxit=260)
# perform DESeq2 variance stabilization instead of rarefying
nosing.vst <- getVarianceStabilizedData(nosing.des)
# Save the untransformed data as a separate variable so you can go back
to it
nosing0 <- nosing
otu_table(nosing) <- otu_table(nosing.vst, taxa_are_rows=TRUE)
nosing.vst <- nosing
nosing <- nosing0
rm(nosing0)
# Set values below zero, to zero
otu_table(nosing.vst)[otu_table(nosing.vst) < 0.0] <- 0

```

```{r PCoA, eval=TRUE, include=TRUE, results="hide", fig.width=5, fig.
height=4, fig.show="hold", fig.align="center", fig.cap="\\label{fig:pcoa}
PCoA ordination plot of all samples, colored by sampling date."}
#use these
UAord <- ordinate(urb.nosing.vst, method = "PCoA", distance = "horn")
plot_ordination(urb.nosing.vst, UAord, type = "samples", color =
"Nearby_Veg",
        label = "SampleID", title = "Quality filtered, all urban sites
```

```
(PCoA, horn)") +
 scale_color_manual(values=c("turquoise3", "darkgoldenrod", "grey30"))
ggsave("ordALL_nearbyveg_horn.pdf", device="pdf", width=5, height=3.25,
units="in", useDingbats=FALSE)
ggsave("ordALL_nearbyveg_horn.png", device="png", width=6.5, height=5,
units="in")

plot_ordination(urb.nosing.vst, UAord, type = "samples", color =
"Date", shape = "Nearby_Veg") +
 geom_point(size=3)
        #label = "SampleID", title = "Quality filtered, all urban sites,
no July (PCoA, horn)") +
ggsave("ordALL_date_horn.pdf", device="pdf", width=5, height=3.25,
units="in", useDingbats=FALSE)
ggsave("ordALL_date_horn.png", device="png", width=6.5, height=5,
units="in")

#ordinate/plot sets separately
sample_data(urb.nosing.vst)$set <- "1"
sample_data(urb.nosing.vst)$set[sample_data(urb.nosing.vst)$Date ==
'08.14.2015'] <- "2"
sample_data(urb.nosing.vst)$set[sample_data(urb.nosing.vst)$Date ==
'08.25.2015'] <- "2"
sample_data(urb.nosing.vst)$set[sample_data(urb.nosing.vst)$Date ==
'09.15.2015'] <- "2"
sample_data(urb.nosing.vst)$set[sample_data(urb.nosing.vst)$Date ==
'09.29.2015'] <- "2"

set1 <- prune_samples(sample_data(urb.nosing.vst)$set == '1', urb.
nosing.vst)
UAord <- ordinate(set1, method = "PCoA", distance = "horn")
plot_ordination(set1, UAord, type = "samples", color = "Nearby_Veg",
        label = "SampleID", title = "Quality filtered, urban sites, set
1 (PCoA, horn)") +
 scale_color_manual(values=c("turquoise3", "darkgoldenrod", "grey30"))
ggsave("ordALL_nearbyveg_set1_horn.pdf", device="pdf", width=6.5,
height=5, units="in", useDingbats=FALSE)
ggsave("ordALL_nearbyveg_set1_horn.png", device="png", width=6.5,
height=5, units="in")

set2 <- prune_samples(sample_data(urb.nosing.vst)$set == '2', urb.
nosing.vst)
UAord <- ordinate(set2, method = "PCoA", distance = "horn")
plot_ordination(set2, UAord, type = "samples", color = "Nearby_Veg",
        label = "SampleID", title = "Quality filtered, all urban sites,
no July (PCoA, horn)") +
 scale_color_manual(values=c("turquoise3", "darkgoldenrod", "grey30"))
ggsave("ordALL_nearbyveg_set2_horn.pdf", device="pdf", width=6.5,
height=5, units="in", useDingbats=FALSE)
ggsave("ordALL_nearbyveg_set2_horn.png", device="png", width=6.5,
```

```
height=5, units="in")


## create an ordination plot with better control over elements
# pull out axis scores from ordination
ord.scores <- UAord$vectors[, c(1, 2)]
# get map out of phyloseq object
map.df <- data.frame(sample_data(urb.nosing.vst))
# control symbol shape & color
map.df$pch <- 17
map.df$pch[map.df$Nearby_Veg == 'paved'] <- 16
map.df$pch[map.df$Nearby_Veg == 'grass'] <- 15
map.df$col <- "turquoise3"
map.df$col[map.df$Nearby_Veg == 'paved'] <- "grey30"
map.df$col[map.df$Nearby_Veg == 'grass'] <- "darkgoldenrod"
map.df$dateColor <- "firebrick2"
map.df$dateColor[map.df$Date == '08.04.2015'] <- "sienna1"
map.df$dateColor[map.df$Date == '08.14.2015'] <- "darkgoldenrod1"
map.df$dateColor[map.df$Date == '08.18.2015'] <- "darkolivegreen3"
map.df$dateColor[map.df$Date == '08.25.2015'] <- "green2"
map.df$dateColor[map.df$Date == '09.01.2015'] <- "turquoise1"
map.df$dateColor[map.df$Date == '09.15.2015'] <- "blue"
map.df$dateColor[map.df$Date == '09.22.2015'] <- "darkviolet"
map.df$dateColor[map.df$Date == '09.29.2015'] <- "hotpink"
unique(map.df$Date)
# make a blank plot
pdf("ordALL_vegEllipse_horn.pdf")
plot(ord.scores[, 1], ord.scores[, 2],type="n",
ylab='PCoA Axis 2', xlab='PCoA Axis 1')
legend("right", unique(map.df$Date), fill = unique(map.df$dateColor))
legend("topright", unique(map.df$Nearby_Veg), pch = unique(map.df$pch))
points(ord.scores[, 1], ord.scores[, 2], pch=map.df$pch, cex=2,
col=map.df$dateColor) #bg=map.df$parkVlot.col,
ellOne <- ordiellipse(ord.scores, map.df$Date, label=FALSE, kind="se",
conf=0.95, col="firebrick2", show='07.14.2015', draw='polygon',
lwd=0.0001, border=NA)
ellTwo <- ordiellipse(ord.scores, map.df$Date, label=FALSE, kind="se",
conf=0.95, col="sienna1", show='08.04.2015', draw='polygon',
lwd=0.0001, border=NA)
ellThree <- ordiellipse(ord.scores, map.df$Date, label=FALSE,
kind="se", conf=0.95, col="darkgoldenrod1", show='08.14.2015',
draw='polygon', lwd=0.0001, border=NA)
ellFour <- ordiellipse(ord.scores, map.df$Date, label=FALSE, kind="se",
conf=0.95, col="darkolivegreen3", show='08.18.2015', draw='polygon',
lwd=0.0001, border=NA)
ellFive <- ordiellipse(ord.scores, map.df$Date, label=FALSE, kind="se",
conf=0.95, col="green2", show='08.25.2015', draw='polygon', lwd=0.0001,
border=NA)
ellSix <- ordiellipse(ord.scores, map.df$Date, label=FALSE, kind="se",
conf=0.95, col="turquoise1", show='09.01.2015', draw='polygon',
```

```
lwd=0.0001, border=NA)
ellSeven <- ordiellipse(ord.scores, map.df$Date, label=FALSE,
kind="se", conf=0.95, col="blue", show='09.15.2015', draw='polygon',
lwd=0.0001, border=NA)
ellEight <- ordiellipse(ord.scores, map.df$Date, label=FALSE,
kind="se", conf=0.95, col="darkviolet", show='09.22.2015',
draw='polygon', lwd=0.0001, border=NA)
ellNine <- ordiellipse(ord.scores, map.df$Date, label=FALSE, kind="se",
conf=0.95, col="hotpink", show='09.29.2015', draw='polygon',
lwd=0.0001, border=NA)
#forest.ellipse <- ordiellipse(ord.scores, map.df$Nearby_
Veg, label=FALSE, kind='sd', col="turquoise3", show='forest',
draw='polygon', lwd=0.0001, border=NA)
#paved.ellipse <- ordiellipse(ord.scores, map.df$Nearby_Veg,
label=FALSE, kind='sd', col="gray", show='paved', draw='polygon',
lwd=0.0001, border=NA)
#grass.ellipse <- ordiellipse(ord.scores, map.df$Nearby_Veg,
label=FALSE, kind='sd', col="darkgoldenrod", show='grass',
draw='polygon', lwd=0.0001, border=NA)
dev.off()

#text(ord.scores[, 1], ord.scores[, 2], map.df$SiteCode, cex=.7, pos=1,
col='gray30')

## ordinate by urban vs. rural
UAord <- ordinate(nosing.vst, method = "PCoA", distance = "horn")
sample_data(nosing.vst)$urbRur <- "Urban"
sample_data(nosing.vst)$urbRur[sample_data(nosing.vst)$Urbanicity ==
"rural"] <- "Rural"
plot_ordination(nosing.vst, UAord, type = "samples", color = "urbRur",
        label = "SampleID", title = "Quality filtered, urban vs. rural
(PCoA, horn)") +
 scale_color_manual(values=c("red", "purple"))
ggsave("ordALL_urb-rur_horn.pdf", device="pdf", width=5, height=3.25,
units="in", useDingbats=FALSE)


```

```{r capscale, eval=TRUE, include=TRUE, results="hide", fig.width=5,
fig.height=4, fig.show="hold", fig.align="center", fig.cap="\\label{fig:cap}
Constrained PCoA ordination of all samples, colored by nearby
vegetation type."}
## APPARENTLY NOT MUCH VARIATION IS EXPLAINED BY VEG
# test hypothesis with direct gradient analysis with constrained PCoA
ordcap <- ordinate(urb.nosing.vst, "CAP", "horn", ~ Nearby_Veg)
require(ggrepel)
require(viridis)
sample_data(urb.nosing.vst)$veg50 <- as.numeric(as.character(sample_
data(urb.nosing.vst)$veg50))
```

```
plot_ordination(urb.nosing.vst, ordcap, type="samples", color="Nearby_
Veg") + #, shape="Nearby_Veg"
 scale_color_manual(values=c("turquoise3", "darkgoldenrod", "grey30"))
+ #scale_color_viridis(begin=0, end=0.7, option="plasma") +
 geom_point(size=3) +
 labs(color="Land Cover") +
 #geom_text_repel(aes(label=sample_data(urb.nosing.vst)$Site), box.
padding=unit(0.6, "lines")) +
 theme(legend.text=element_text(size=11), legend.title=element_
text(size=12, vjust=1),
    axis.text=element_text(size=11), axis.title=element_text(size=12)
    )
ggsave("Nearby_Veg_capscale_vst_horn.pdf", device="pdf", width=5.5,
height=4, units="in", useDingbats=FALSE)
ggsave("Nearby_Veg_capscale_vst_horn.png", device="png", width=5.5,
height=4, units="in")

```

```{r dateOrd, eval=TRUE, include=TRUE, fig.width=5, fig.height=4, fig.
show="hold", fig.align="center", fig.cap="\\label{fig:capDate}Constrained
PCoA ordinations separated by sampling date, colored by nearby
vegetation type."}
# separate by date for VST
urb7.14 <- prune_samples(sample_data(urb.nosing.
vst)$Date=="07.14.2015", urb.nosing.vst)
urb7.14 <- prune_taxa(taxa_sums(urb7.14) > 9.4187, urb7.14)
urb8.4 <- prune_samples(sample_data(urb.nosing.vst)$Date=="08.04.2015",
urb.nosing.vst)
urb8.4 <- prune_taxa(taxa_sums(urb8.4) > 9.4187, urb8.4)
urb8.14 <- prune_samples(sample_data(urb.nosing.
vst)$Date=="08.14.2015", urb.nosing.vst)
urb8.14 <- prune_taxa(taxa_sums(urb8.14) > 9.4187, urb8.14)
urb8.18 <- prune_samples(sample_data(urb.nosing.
vst)$Date=="08.18.2015", urb.nosing.vst)
urb8.18 <- prune_taxa(taxa_sums(urb8.18) > 9.4187, urb8.18)
urb8.25 <- prune_samples(sample_data(urb.nosing.
vst)$Date=="08.25.2015", urb.nosing.vst)
urb8.25 <- prune_taxa(taxa_sums(urb8.25) > 9.4187, urb8.25)
urb9.1 <- prune_samples(sample_data(urb.nosing.vst)$Date=="09.01.2015",
urb.nosing.vst)
urb9.1 <- prune_taxa(taxa_sums(urb9.1) > 9.4187, urb9.1)
urb9.15 <- prune_samples(sample_data(urb.nosing.
vst)$Date=="09.15.2015", urb.nosing.vst)
urb9.15 <- prune_taxa(taxa_sums(urb9.15) > 9.4187, urb9.15)
urb9.22 <- prune_samples(sample_data(urb.nosing.
vst)$Date=="09.22.2015", urb.nosing.vst)
urb9.22 <- prune_taxa(taxa_sums(urb9.22) > 9.4187, urb9.22)
urb9.29 <- prune_samples(sample_data(urb.nosing.
vst)$Date=="09.29.2015", urb.nosing.vst)
```

```
urb9.29 <- prune_taxa(taxa_sums(urb9.29) > 9.4187, urb9.29)
urb.dateList.vst <- list(urb7.14=urb7.14, urb8.4=urb8.4,
urb8.14=urb8.14, urb8.18=urb8.18,
        urb8.25=urb8.25, urb9.1=urb9.1, urb9.15=urb9.15,
urb9.22=urb9.22,
        urb9.29=urb9.29)

# ordinations by date for urban samples, VST counts (untransformed
counts don't plot great)
dateListVST.ord <- llply(urb.dateList.vst, function(x){
 ord <- ordinate(x, method="PCoA", distance="horn")
 plot_ordination(x, ord, type="samples", shape="Nearby_Veg",
color="Nearby_Veg", label="SiteCode",
         title=paste0(names(x))) +
  scale_color_manual(values=c("turquoise3", "darkgoldenrod", "grey30"))
+ #scale_color_viridis(begin=0, end=.7)
  geom_point(size=3) +
  theme(text=element_text(size=8))}
 )
#dateListVST.ord

## NOTES: DCA/bray looks pretty good (horn looks same), NMDS not great,
PCoA/horn also good (bray and canberra not as good)
for (i in seq_along(dateListVST.ord)){
 ggsave(paste0(names(dateListVST.ord)[i], "Nearby_Veg_PCoAhorn_
vst.pdf"), dateListVST.ord[[i]], width=4, height=3, units="in",
useDingbats=FALSE)
}
for (i in seq_along(dateListVST.ord)){
 ggsave(paste0(names(dateListVST.ord)[i], "Nearby_Veg_PCoAhorn_vst.
png"), dateListVST.ord[[i]], device="png", width=4, height=3,
units="in")
}

require(ggrepel)
require(viridis)
ord.plot <- function(x){
 ord <- ordinate(x, method="CAP", distance="horn", ~"Nearby_Veg")
 plot_ordination(x, ord, type="samples", shape="Nearby_Veg",
         color="Nearby_Veg", #label="SiteCode",
         title=paste0(names(x))) +
  geom_point(size=3) +
  scale_color_manual(values=c("turquoise3", "darkgoldenrod", "grey30"))
+ #scale_color_viridis(begin=0, end=.7) +
  theme(text=element_text(size=8))
 }

dateList.cap <- llply(urb.dateList.vst, ord.plot)
#dateList.cap
for (i in seq_along(dateList.cap)){
```

```
 ggsave(paste0(names(dateList.cap)[i], “capscale_Nearby_Veg_nolabels.
pdf”), dateList.cap[[i]], width=4, height=3, units=”in”)
}
for (i in seq_along(dateList.cap)){
 ggsave(paste0(names(dateList.cap)[i], “capscale_Nearby_Veg_nolabels.
png”), dateList.cap[[i]], device=”png”, width=4, height=3, units=”in”)
}

# separate by date for VST
all7.14 <- prune_samples(sample_data(nosing.vst)$Date==”07.14.2015”,
nosing.vst)
all7.14 <- prune_taxa(taxa_sums(all7.14) > 9.4187, all7.14)
all8.4 <- prune_samples(sample_data(nosing.vst)$Date==”08.04.2015”,
nosing.vst)
all8.4 <- prune_taxa(taxa_sums(all8.4) > 9.4187, all8.4)
all8.14 <- prune_samples(sample_data(nosing.vst)$Date==”08.14.2015”,
nosing.vst)
all8.14 <- prune_taxa(taxa_sums(all8.14) > 9.4187, all8.14)
all8.18 <- prune_samples(sample_data(nosing.vst)$Date==”08.18.2015”,
nosing.vst)
all8.18 <- prune_taxa(taxa_sums(all8.18) > 9.4187, all8.18)
all8.25 <- prune_samples(sample_data(nosing.vst)$Date==”08.25.2015”,
nosing.vst)
all8.25 <- prune_taxa(taxa_sums(all8.25) > 9.4187, all8.25)
all9.1 <- prune_samples(sample_data(nosing.vst)$Date==”09.01.2015”,
nosing.vst)
all9.1 <- prune_taxa(taxa_sums(all9.1) > 9.4187, all9.1)
all9.15 <- prune_samples(sample_data(nosing.vst)$Date==”09.15.2015”,
nosing.vst)
all9.15 <- prune_taxa(taxa_sums(all9.15) > 9.4187, all9.15)
all9.22 <- prune_samples(sample_data(nosing.vst)$Date==”09.22.2015”,
nosing.vst)
all9.22 <- prune_taxa(taxa_sums(all9.22) > 9.4187, all9.22)
all9.29 <- prune_samples(sample_data(nosing.vst)$Date==”09.29.2015”,
nosing.vst)
all9.29 <- prune_taxa(taxa_sums(all9.29) > 9.4187, all9.29)
all.dateList.vst <- list(all7.14=all7.14, all8.4=all8.4,
all8.14=all8.14, all8.18=all8.18,
        all8.25=all8.25, all9.1=all9.1, all9.15=all9.15,
all9.22=all9.22,
        all9.29=all9.29)

# ordinations by date for all samples (urban + rural), VST counts
(untransformed counts don’t plot great)
dateListVST.ord <- llply(all.dateList.vst, function(x){
 ord <- ordinate(x, method=”PCoA”, distance=”horn”)
 plot_ordination(x, ord, type=”samples”, shape=”urbRur”, color=”Nearby_
Veg”, label=”SiteCode”,
         title=paste0(names(x))) +
  scale_color_manual(values=c(“turquoise3”, “darkgoldenrod”, “grey30”))
```

```
+ #"red", "purple")) + #scale_color_viridis(begin=0, end=.7) +
  geom_point(size=3) +
  theme(text=element_text(size=8))}
 )
dateListVST.ord

for (i in seq_along(dateListVST.ord)){
 ggsave(paste0(names(dateListVST.ord)[i], "_urbRur_byVeg_PCoAhorn_
vst.pdf"), dateListVST.ord[[i]], width=4, height=3, units="in",
useDingbats=FALSE)
}

forBart <- data.frame(unique(sample_data(nosing)$SiteCode))
colnames(forBart) <- "SiteCode"
forBart$Site <- sample_data(nosing)$Site[match(forBart$SiteCode,
sample_data(nosing)$SiteCode)]
forBart$VegType <- sample_data(nosing)$Other_Veg[match(forBart$SiteCode,
sample_data(nosing)$SiteCode)]
forBart$Context <- sample_data(nosing.
vst)$urbRur[match(forBart$SiteCode, sample_data(nosing.vst)$SiteCode)]
write.csv(forBart, "siteList.csv")

#clean up
garbage <- paste0("urb", c("7.14","8.4","8.14","8.25","9.1","9.15","9.2
2","9.29"))
rm(list = garbage)
garbage <- paste0("all", c("7.14","8.4","8.14","8.25","9.1","9.15","9.2
2","9.29"))
rm(list = garbage)

```

```{r mantel, include=TRUE, results="hide"}
urbvstTab <- getTab(urb.nosing.vst)
urbvstEnv <- data.frame(sample_data(urb.nosing.vst))
# convert variables to factors
urbvstEnv$Nearby_Veg <- as.factor(urbvstEnv$Nearby_Veg)
urbvstEnv$Date <- as.factor(urbvstEnv$Date)
urbvstEnv$SiteCode <- as.factor(urbvstEnv$SiteCode)
urbvstEnv$Context <- as.factor(urbvstEnv$Context)
urbvstEnv$Maintenance <- as.factor(urbvstEnv$Maintenance)
urbvstEnv$windDir <- as.factor(urbvstEnv$windDir)

# test spatial autocorrelation with Mantel test
urb.spat <- as.matrix(dist(cbind(urbvstEnv$Long, urbvstEnv$Lat)))
urb.horn <- as.matrix(vegdist(urbvstTab, method="horn"))
urbMan <- mantel(urb.horn, urb.spat, method="spearman",
permutations=99999)
str(urbMan)
urbMan$statistic
```

```
urbMan$signif
```


```{r hypTest, include=TRUE, results="hide", fig.width=6.5, fig.height=3,
fig.show="hold", fig.align="center", fig.cap="\\label{fig:hypTest}Effect
of increasing buffer size on strength of association between amount of
vegetation and community similarity."}
# PERMANOVA for variables of interest

urbvstTab <- getTab(urb.nosing.vst)
urbvstEnv <- data.frame(sample_data(urb.nosing.vst)[, c(5,6,32,33,45,61
,68,71,75,77)])
urbvstEnv$structDiv[is.na(urbvstEnv$structDiv)] <- 0
colnames(urbvstEnv)
#urbvstEnv <- urbvstEnv[, c(5,6,32,33,45,61,68,71,75,77)]
varlist <- c("SiteCode", "Collected_By", "Date", "Nearby_Veg",
"green50", "green800", "windDir", "tempAvg")
#, "Elevation", "windAvg", "windMax", "structDiv", "Date", "green50"
adonisRes <- list()
#try ps1.tab instead of urbvstTab urbvstEnv
#ps1.vstEnv <- data.frame(sample_data(ps1.vst))
#ps1.vstEnv$Nearby_Veg <- as.factor(ps1.vstEnv$Nearby_Veg)
#ps1.vstEnv$Date <- as.factor(ps1.vstEnv$Date)
#ps1.vstEnv$SiteCode <- as.factor(ps1.vstEnv$SiteCode)
#adonis(urbvstTab ~ SiteCode + Nearby_Veg, data=urbvstEnv, perm=9999,
method="horn")

for (i in varlist){
      form <- as.formula(paste("urbvstTab", i, sep="~"))
      adonisRes[[i]] <- adonis(form, data=urbvstEnv, perm=9999,
method="horn")
      }
#knitr::kable(adonisRes)
sink("adonisResults12.txt")
adonisRes
sink()

#PERMANOVA only on forest and grass
notPaved <- prune_samples(sample_data(urb.nosing.vst)$Nearby_Veg !=
"paved", urb.nosing.vst)
vegvstTab <- getTab(notPaved)
vegvstEnv <- data.frame(sample_data(notPaved))
adonisRes_structDiv <- adonis(vegvstTab ~ structDiv, data=vegvstEnv,
perm=9999, method="horn")
adonisRes_vegDiv <- adonis(vegvstTab ~ vegDiv, data=vegvstEnv,
perm=9999, method="horn")

#multivariate test with site location & collection date
adonisRes_sitedateInt <- adonis(urbvstTab ~ SiteCode*Date,
```

```
data=urbvstEnv, perm=9999, method="horn")
adonisRes_sitedate <- adonis(urbvstTab ~ SiteCode+Date, data=urbvstEnv,
perm=9999, method="horn")

# just for buffer distances and pull out R2 and P-value for each buffer
test
varlist <- c("veg50", "veg100", "veg200", "veg400", "veg800",
"veg1600")
buffRes <- list()
for (i in varlist){
      form <- as.formula(paste("urbvstTab", i, sep="~"))
      buffRes[[i]] <- adonis(form, data=urbvstEnv, perm=9999,
method="horn")
      }
buffAdon <- data.table(c(buffRes[[1]]$aov.tab$R2[1], buffRes[[2]]$aov.
tab$R2[1],
          buffRes[[3]]$aov.tab$R2[1], buffRes[[4]]$aov.tab$R2[1],
          buffRes[[5]]$aov.tab$R2[1], buffRes[[6]]$aov.tab$R2[1]),
         c(buffRes[[1]]$aov.tab$'Pr(>F)'[1], buffRes[[2]]$aov.
tab$'Pr(>F)'[1],
          buffRes[[3]]$aov.tab$'Pr(>F)'[1], buffRes[[4]]$aov.
tab$'Pr(>F)'[1],
          buffRes[[5]]$aov.tab$'Pr(>F)'[1], buffRes[[6]]$aov.
tab$'Pr(>F)'[1]),
         c(50, 100, 200, 400, 800, 1600))
colnames(buffAdon) <- c("R2", "Pval", "buffDist")
#plot(R2 ~ buffDist, data=buffAdon)
#model <- lm(R2 ~ buffDist, data=buffAdon)
#abline(model, col = "red")
#summary(model)
#coef(model)
ggplot(data=buffAdon, aes(x=buffDist, y=R2)) +
 geom_point() + labs(x="Buffer Distance")
ggsave("buffDist_R2.pdf", device="pdf", width=4, height=3, units="in")
ggplot(data=buffAdon, aes(x=buffDist, y=Pval)) +
 geom_point() + labs(x="Buffer Distance", y="p-value")
ggsave("buffDist_pval.pdf", device="pdf", width=4, height=3,
units="in")

# haven't figured out how to interpret simper output yet, should give
relative contribution of each taxa to dissimilarity
# same results as DESeq2?
(sim <- with(urbvstEnv, simper(urbvstTab, Nearby_Veg)))
SIMforest_paved <- t(do.call(rbind.data.frame, sim[[1]]))
write.csv(SIMforest_paved, "SIMforest_paved.csv")

varlist <- c("Nearby_Veg", "veg50", "tempAvg", "structDiv")

adonisRes <- list()
newRes <- list()
```

```
for (i in seq_along(urb.dateList.vst)){
 tab <- getTab(urb.dateList.vst[[i]])
 env <- data.frame(sample_data(urb.dateList.vst[[i]]))
 env$Nearby_Veg <- as.factor(env$Nearby_Veg)
 env$structDiv[is.na(env$structDiv)] <- 0
 for (h in varlist){
      form <- as.formula(paste("tab", h, sep="~"))
      adonisRes[[h]] <- adonis(form, data=env, perm=9999,
method="horn")
 }
 newRes[[i]] <- adonisRes
}

nvegRes <- cbind(unlist(lapply(1:9, function(i){newRes[[i]][[1]][[1]]
[[5]][1]})),
         unlist(lapply(1:9, function(i){newRes[[i]][[1]][[1]][[6]]
[1]})))
veg50Res <- cbind(unlist(lapply(1:9, function(i){newRes[[i]][[2]][[1]]
[[5]][1]})),
         unlist(lapply(1:9, function(i){newRes[[i]][[2]][[1]][[6]]
[1]})))
tavRes <- cbind(unlist(lapply(1:9, function(i){newRes[[i]][[3]][[1]]
[[5]][1]})),
        unlist(lapply(1:9, function(i){newRes[[i]][[3]][[1]][[6]]
[1]})))
structDiv <- cbind(unlist(lapply(1:9, function(i){newRes[[i]][[4]][[1]]
[[5]][1]})),
         unlist(lapply(1:9, function(i){newRes[[i]][[4]][[1]][[6]]
[1]})))
dateAdonis_df <- data.frame("nveg_R2"=nvegRes[,1], "nveg_
Pval"=nvegRes[,2], "veg50_R2"=veg50Res[,1], "veg50_Pval"=veg50Res[,2],
      "tav_R2"=tavRes[,1], "tav_Pval"=tavRes[,2], "structDiv_
R2"=structDiv[,1], "structDiv_Pval"=structDiv[,2])

row.names(dateAdonis_df) <- c("07.14.2015", "08.04.2015", "08.14.2015",
"08.18.2015", "08.25.2015",
               "09.01.2015", "09.15.2015", "09.22.2015", "09.29.2015")
write.csv(dateAdonis_df, "dateAdonis_df.csv")

# put into inline code chunk below:
# (R^2^=`r adonisRes$Nearby_Veg$aov.tab$R2[1]`, p<)
# `r adonisRes$SiteCode`
```

```{r structDiv, eval=TRUE, include=TRUE, results="hide"}

#plot LiDAR cylinder example
abf.las <- readLAS("/Users/gwynhwyfer/Documents/PhD work/dissertation/
chapter3_urbanMicrobiome/processed data/LiDARcyl/clip_ABF.las")
plot(abf.las, size=6)
```

```
#### test by assessing whether variability across time is less for
structurally
#### diverse sites (that block wind) than grassy sites
#### HOW TO QUANTIFY variability across time for each site, grouped by
cover type
noPaved <- prune_samples(sample_data(urb.nosing.vst)$Nearby_Veg !=
“paved”, urb.nosing.vst)
noPavedEnv <- data.frame(sample_data(noPaved))
noPaved <- getTab(noPaved)
# convert variables to factors
noPavedEnv$Nearby_Veg <- as.factor(noPavedEnv$Nearby_Veg)

alphaDiv <- diversity(noPaved, index=”shannon”)
cor.test(alphaDiv, noPavedEnv$structDiv)
struct.lm <- lm(alphaDiv ~ noPavedEnv$structDiv)
summary(struct.lm)

str(dateListVST.ord[[1]][[1]])
points.df <- data.frame(rbind(dateListVST.ord[[1]][[1]][,1:2],
dateListVST.ord[[2]][[1]][,1:2],
               dateListVST.ord[[3]][[1]][,1:2], dateListVST.ord[[4]]
[[1]][,1:2],
               dateListVST.ord[[5]][[1]][,1:2], dateListVST.ord[[6]]
[[1]][,1:2],
               dateListVST.ord[[7]][[1]][,1:2], dateListVST.ord[[8]]
[[1]][,1:2],
               dateListVST.ord[[9]][[1]][,1:2]))
str(points.df)
#list of SiteCode
#ldply
siteList <- unique(sample_data(urb.nosing.vst)$SiteCode)

dist1.2 <- ldply(siteList, function(x){
 sqrt((points.df$Axis.1[grep(paste(x), rownames(points.df))][1] -
points.df$Axis.1[grep(paste(x),
       rownames(points.df))][2])^2 + (points.df$Axis.2[grep(paste(x),
rownames(points.df))][1] -
       points.df$Axis.2[grep(paste(x), rownames(points.df))][2])^2)
})

dist2.3 <- ldply(siteList, function(x){
   sqrt((points.df$Axis.1[grep(paste(x), rownames(points.df))][2] -
points.df$Axis.1[grep(paste(x),
       rownames(points.df))][3])^2 + (points.df$Axis.2[grep(paste(x),
rownames(points.df))][2] -
       points.df$Axis.2[grep(paste(x), rownames(points.df))][3])^2)
})

dist3.4 <- ldply(siteList, function(x){
```

```
  sqrt((points.df$Axis.1[grep(paste(x), rownames(points.df))][3] -
points.df$Axis.1[grep(paste(x),
       rownames(points.df))][4])^2 + (points.df$Axis.2[grep(paste(x),
rownames(points.df))][3] -
       points.df$Axis.2[grep(paste(x), rownames(points.df))][4])^2)
 })

siteDist <- cbind(dist1.2, dist2.3, dist3.4)
row.names(siteDist) <- siteList
colnames(siteDist) <- c("dist1.2", "dist2.3", "dist3.4")
siteDist$totalDist <- siteDist[,1] + siteDist[,2] + siteDist[,3]
siteDist$Nearby_Veg <- sample_data(urb.nosing.vst)$Nearby_
Veg[match(row.names(siteDist), sample_data(urb.nosing.vst)$SiteCode)]

ggplot(siteDist) +
  geom_boxplot(aes(x=siteDist$Nearby_Veg, y=siteDist$totalDist,
color=siteDist$Nearby_Veg)) +
  # scale_y_continuous(limits=c(2,8)) +
  labs(x="", y="") +
  scale_color_manual(values=c("turquoise3", "darkgoldenrod", "grey30"))
+
  theme(legend.position="none")

kruskal.test(totalDist ~ as.factor(Nearby_Veg), data=siteDist)
#pairwise.wilcox.test(urb.nosing2.relTab$Actinobacteria, urb.nosing2.
relTab$Nearby_Veg,
 #        p.adjust.method = "bonferroni")

ggsave("struct_alphaDiv.pdf", device="pdf", width=6, height=4,
units="in")

```

```{r forestGrey, include=TRUE, results="hide"}
## split into pairwise land cover type comparisons
# try class.merge or order.merge or fam.merge or genus.merge instead of
urb.nosing
noGrass <- prune_samples(sample_data(class.merge)$Nearby_Veg !=
"grass", class.merge)
noGrass <- prune_taxa(taxa_sums(noGrass) != 0, noGrass)
hist(taxa_sums(noGrass))
hist(log10(apply(otu_table(noGrass), 1, var)))

noGrass.des <- phyloseq_to_deseq2(noGrass, ~ Date + Nearby_Veg)
noGrass.glm <- DESeq(noGrass.des, test="Wald", fitType="parametric")

res.noGrass.glm <- results(noGrass.glm, cooksCutoff=TRUE, alpha=0.05)
summary(res.noGrass.glm)
mcols(res.noGrass.glm, use.names=TRUE) #positive l2fc is paved
res.noGrass.glm.df <- subset(res.noGrass.glm, padj < 0.05)
```

```
res.noGrass.glm.df <- cbind(as(res.noGrass.glm.df, “data.frame”),
        as(tax_table(class.merge)[rownames(res.noGrass.glm.df),],
“matrix”),
        as(t(otu_table(class.mergeRel)[,rownames(res.noGrass.glm.df)]),
“matrix”),
        as(taxa_sums(class.mergeRel)[rownames(res.noGrass.glm.df)],
“matrix”),
        stringsAsFactors=FALSE
        )
colnames(res.noGrass.glm.df)[length(colnames(res.noGrass.glm.df))] <-
c(“relAbund”)
res.noGrass.glm.df$SequenceVariant <- rownames(res.noGrass.glm.df)
res.noGrass.glm.df$SiteType <- “forest”
res.noGrass.glm.df$SiteType[res.noGrass.glm.df$log2FoldChange >= 0] <-
“paved”
res.noGrass.glm.df$SiteType <- as.factor(res.noGrass.glm.df$SiteType)
res.noGrass.glm.df <- res.noGrass.glm.df[order(res.noGrass.glm.
df$log2FoldChange),]
res.noGrass.glm.df$Genus <- gsub(“Armatimonas/Armatimonadetes_gp1”,
“Armatimonas”, res.noGrass.glm.df$Genus)
#knitr::kable(res.noGrass.glm.df)
write.csv(res.noGrass.glm.df, file=”deseq_noGrass_class.csv”)
length(res.noGrass.glm.df[,1])
# create positive-negative barplot of differentially abundant taxa
sigplot_noGrass <- data.frame(row.names(res.noGrass.glm.df),
           res.noGrass.glm.df$Class,
           res.noGrass.glm.df$Order,
           res.noGrass.glm.df$Family,
           res.noGrass.glm.df$Genus,
           res.noGrass.glm.df$log2FoldChange,
           res.noGrass.glm.df$SiteType,
           stringsAsFactors = FALSE)
names(sigplot_noGrass) <- c(“SequenceVariant”, “Class”, “Order”,
“Family”, “Genus”, “log2FoldChange”, “SiteType”)
sigGenusList <- sigplot_noGrass$Genus
sigFamList <- sigplot_noGrass$Family
sigClassList <- sigplot_noGrass$Class
sigplot_noGrass$Genus <- gsub(“Clostridium_sensu_stricto”,
“Clostridium”, sigplot_noGrass$Genus)
sigplot_noGrass$Genus <- gsub(“Armatimonas/Armatimonadetes_gp1”,
“Armatimonadetes”, sigplot_noGrass$Genus)
sigplot_noGrass$Genus <- gsub(“Chthonomonas/Armatimonadetes_gp3”,
“Chthonomonas”, sigplot_noGrass$Genus)

ggplot(sigplot_noGrass, aes(x = reorder(SequenceVariant,
log2FoldChange), y = log2FoldChange,
          fill=Class, legend=””)) +
 geom_bar(stat = “identity”, position = “identity”) +
 #scale_fill_manual(values=famPal, drop=FALSE) +
 coord_flip() +
```

```
 scale_x_discrete(labels=sigplot_noGrass$Class) +
 scale_y_continuous(limits=c(-2,2), breaks=c(-2, -1, 0, 1, 2)) +
 labs(x="Class", y="l2FC") +
 theme(aspect.ratio=3, legend.position="none", axis.text=element_
text(size=10),
    axis.title=element_text(size=12)
 )

ggsave("difAbund_noGrass_class.pdf", device="pdf", width=5, height=9,
units="in")
ggsave("difAbund_noGrass_class.png", device="png", width=5, height=9,
units="in")

```

```{r grassForest, include=TRUE}
## split into pairwise land cover type comparisons
noGrey <- prune_samples(sample_data(class.merge)$Nearby_Veg != "paved",
class.merge)
noGrey <- prune_taxa(taxa_sums(noGrey) != 0, noGrey)
noGrey.des <- phyloseq_to_deseq2(noGrey, ~ Date + Nearby_Veg)
noGrey.glm <- DESeq(noGrey.des, test="Wald", fitType="parametric")

res.noGrey.glm <- results(noGrey.glm, cooksCutoff=TRUE, alpha=0.05)
summary(res.noGrey.glm)
mcols(res.noGrey.glm, use.names=TRUE) # to figure out which is
upregulated, should be positive for grass
res.noGrey.glm.df <- subset(res.noGrey.glm, padj < 0.05)
res.noGrey.glm.df <- cbind(as(res.noGrey.glm.df, "data.frame"),
        as(tax_table(class.merge)[rownames(res.noGrey.glm.df),],
"matrix"),
        as(t(otu_table(class.mergeRel)[,rownames(res.noGrey.glm.df)]),
"matrix"),
        as(taxa_sums(class.mergeRel)[rownames(res.noGrey.glm.df)],
"matrix"),
        stringsAsFactors=FALSE
        )
colnames(res.noGrey.glm.df)[length(colnames(res.noGrey.glm.df))] <-
c("relAbund")
res.noGrey.glm.df$SequenceVariant <- rownames(res.noGrey.glm.df)
res.noGrey.glm.df$SiteType <- "grass"
res.noGrey.glm.df$SiteType[res.noGrey.glm.df$log2FoldChange < 0] <-
"forest"
res.noGrey.glm.df$SiteType <- as.factor(res.noGrey.glm.df$SiteType)
res.noGrey.glm.df <- res.noGrey.glm.df[order(res.noGrey.glm.
df$log2FoldChange),] # order by l2fc
#knitr::kable(res.noGrey.glm.df)
write.csv(res.noGrey.glm.df, file="deseq_noGrey_class.csv")
length(res.noGrey.glm.df)
# create positive-negative barplot of differentially abundant taxa
```

```
sigplot_noGrey <- data.frame(row.names(res.noGrey.glm.df),
           as.character(res.noGrey.glm.df$Class),
           as.character(res.noGrey.glm.df$Order),
           as.character(res.noGrey.glm.df$Family),
           as.character(res.noGrey.glm.df$Genus),
           res.noGrey.glm.df$log2FoldChange,
           res.noGrey.glm.df$SiteType,
           stringsAsFactors = FALSE)
names(sigplot_noGrey) <- c("SequenceVariant", "Class", "Order",
"Family", "Genus", "log2FoldChange", "SiteType")
sigGenusList <- c(sigGenusList, sigplot_noGrey$Genus)
sigGenusList <- unique(sigGenusList)
sigFamList <- c(sigFamList, sigplot_noGrey$Family)
sigFamList <- unique(sigFamList)
sigClassList <- c(sigClassList, sigplot_noGrey$Class)
sigClassList <- unique(sigClassList)
sigplot_noGrey$Genus <- gsub("Clostridium_sensu_stricto",
"Clostridium", sigplot_noGrey$Genus)
sigplot_noGrey$Genus <- gsub("Armatimonas/Armatimonadetes_gp1",
"Armatimonadetes", sigplot_noGrey$Genus)
sigplot_noGrey$Genus <- gsub("Chthonomonas/Armatimonadetes_gp3",
"Chthonomonas", sigplot_noGrey$Genus)

ggplot(sigplot_noGrey, aes(x=reorder(SequenceVariant, log2FoldChange),
y=log2FoldChange, fill=Class)) +
 geom_bar(stat = "identity", position = "identity") +
 coord_flip() +
 #scale_fill_manual(values=famPal, drop=FALSE) +
 scale_x_discrete(labels=sigplot_noGrey$Class) +
 scale_y_continuous(limits=c(-2,2), breaks=c(-2, -1, 0, 1, 2)) +
 labs(x="Class", y="l2FC") +
 theme(aspect.ratio=3, legend.position="none", axis.text=element_
text(size=10),
    axis.title=element_text(size=12)
 )

ggsave("difAbund_noGrey_class.pdf", device="pdf", width=4, height=6,
units="in")
ggsave("difAbund_noGrey_class.png", device="png", width=4, height=6,
units="in")

```

```{r grassGrey, include=TRUE}
## split into pairwise land cover type comparisons
noForest <- prune_samples(sample_data(class.merge)$Nearby_Veg !=
"forest", class.merge)
noForest <- prune_taxa(taxa_sums(noForest) != 0, noForest)
sample_data(noForest)$Nearby_Veg
noForest.des <- phyloseq_to_deseq2(noForest, ~ Date + Nearby_Veg)
```

```
noForest.glm <- DESeq(noForest.des, test="Wald", fitType="parametric")

res.noForest.glm <- results(noForest.glm, cooksCutoff=FALSE, alpha=0.5)
summary(res.noForest.glm)
mcols(res.noForest.glm, use.names=TRUE)
res.noForest.glm.df <- subset(res.noForest.glm, padj < 0.5)
res.noForest.glm.df <- cbind(as(res.noForest.glm.df, "data.frame"),
        as(tax_table(class.merge)[rownames(res.noForest.glm.df),],
"matrix"),
        as(t(otu_table(class.mergeRel)[,rownames(res.noForest.glm.
df)]), "matrix"),
        as(taxa_sums(class.mergeRel)[rownames(res.noForest.glm.df)],
"matrix"),
        stringsAsFactors = FALSE
        )
colnames(res.noForest.glm.df)[length(colnames(res.noForest.glm.df))] <-
c("relAbund")
res.noForest.glm.df$SequenceVariant <- rownames(res.noForest.glm.df)
res.noForest.glm.df$SiteType <- "grass"
res.noForest.glm.df$SiteType[res.noForest.glm.df$log2FoldChange >= 0]
<- "paved"
res.noForest.glm.df$SiteType <- as.factor(res.noForest.glm.df$SiteType)
res.noForest.glm.df <- res.noForest.glm.df[order(res.noForest.glm.
df$log2FoldChange),] # order by l2fc
#knitr::kable(res.noForest.glm.df)
write.csv(res.noForest.glm.df, file="deseq_noForest.csv")
length(res.noForest.glm.df)
# create positive-negative barplot of differentially abundant taxa
sigplot_noForest <- data.frame(row.names(res.noForest.glm.df),
           as.character(res.noForest.glm.df$Class),
           as.character(res.noForest.glm.df$Order),
           as.character(res.noForest.glm.df$Family),
           as.character(res.noForest.glm.df$Class),
           res.noForest.glm.df$log2FoldChange,
           res.noForest.glm.df$SiteType,
           stringsAsFactors = FALSE)
names(sigplot_noForest) <- c("SequenceVariant", "Class", "Order",
"Family", "Genus", "log2FoldChange", "SiteType")
sigGenusList <- c(sigGenusList, sigplot_noForest$Genus)
sigGenusList <- unique(sigGenusList)
sigFamList <- c(sigFamList, sigplot_noForest$Family)
sigFamList <- unique(sigFamList)
sigClassList <- c(sigClassList, sigplot_noForest$Class)
sigClassList <- unique(sigClassList)
sigplot_noForest$Genus <- gsub("Clostridium_sensu_stricto",
"Clostridium", sigplot_noForest$Genus)
sigplot_noForest$Genus <- gsub("Armatimonas/Armatimonadetes_gp1",
"Armatimonadetes", sigplot_noForest$Genus)
sigplot_noForest$Genus <- gsub("Chthonomonas/Armatimonadetes_gp3",
"Chthonomonas", sigplot_noForest$Genus)
```

```
ggplot(sigplot_noForest, aes(x = reorder(SequenceVariant,
log2FoldChange), y = log2FoldChange,
          fill = Class, legend=””)) +
 geom_bar(stat = “identity”, position = “identity”) +
 #scale_fill_manual(values=famPal, drop=FALSE) +
 coord_flip() +
 labs(y=”l2FC”, x=”Class”) +
 scale_x_discrete(labels=sigplot_noForest$Class) +
 scale_y_continuous(limits=c(-2, 2), breaks=c(-2, -1, 0, 1, 2)) +
 theme(aspect.ratio=3, legend.position = “none”, axis.text=element_
text(size=10),
    axis.title=element_text(size=12)
 )

ggsave(“difAbund_noForest_class.pdf”, device=”pdf”, width=4, height=6,
units=”in”)
ggsave(“difAbund_noForest_class.png”, device=”png”, width=4, height=6,
units=”in”)

```

```{r pairwiseAdonis}
noGrass.tab <- getTab(noGrass)
noGrass.env <- data.frame(sample_data(noGrass))
adonis(noGrass.tab~Nearby_Veg, noGrass.env, permutations = 9999, method
= “horn”)

noGrey.tab <- getTab(noGrey)
noGrey.env <- data.frame(sample_data(noGrey))
adonis(noGrey.tab~Nearby_Veg, noGrey.env, permutations = 9999, method =
“horn”)

noForest.tab <- getTab(noForest)
noForest.env <- data.frame(sample_data(noForest))
adonis(noForest.tab~Nearby_Veg, noForest.env, permutations = 9999,
method = “horn”)

```

```{r classBar, eval=TRUE, include=TRUE, results=”hide”}
## stacked barplots to compare proportional composition for all urban
sites (class level) by Nearby_Veg
write.csv(unique(tax_table(urb.nosing)[,5]), “familyList.csv”)
sort.class = sort(tapply(taxa_sums(urb.nosing), tax_table(urb.nosing)[,
“Class”], sum), TRUE)
bottom.class <- sort.class[11:length(sort.class)]
#urb.nosing1 = subset_taxa(urb.nosing, Class %in% names(top.class))
#get top 10 most abundant Classes
urb.nosing2 = subset_taxa(urb.nosing, Class %in% sigClassList) #get all
```

```
other taxa
#urb.nosing2 <- merge_taxa(urb.nosing, taxa_names(urb.nosing2),
archetype=1) #merge all other taxa into Class “Other”
urb.nosing2 <- tax_glom(urb.nosing2, taxrank=”Class”)
get_taxa_unique(urb.nosing2, “Class”)
#tax_table(urb.nosing2)[,3] <- gsub(“NA”, “Other”, tax_table(urb.
nosing2)[,3])
#calculate standard error for errorbars
urb.nosing2.rel <- transform_sample_counts(urb.nosing2, function(x) 100
* x/sum(x))
# convert your processed phyloseq object into a dataframe
df <- psmelt(urb.nosing2.rel) #(urb.nosing2.rel) #class.mergeRel
# group by Nearby_Veg and Class, calculate mean abundance and standard
deviation
se <- function(x) sqrt(var(x)/length(x))
avgs <- ddply(df, .(Nearby_Veg, Class), plyr::summarize,
       mean = mean(Abundance),
       se = se(Abundance))
avgs$Nearby_Veg <- as.factor(avgs$Nearby_Veg)
# plot bar graph with standard deviation as error bars
ggplot(avgs, aes(fill=Nearby_Veg, x=Class, y=mean, group=Nearby_Veg)) +
 geom_bar(aes(color=Nearby_Veg, fill=Nearby_Veg), stat=”identity”,
position=position_dodge(width=0.8), width=0.6) +
 coord_flip() +
 geom_errorbar(aes(ymin=mean-se, ymax=mean+se), position=position_
dodge(width=0.8), width=0.5, size=0.15) +
 ylab(“Percentage of Sequences”) +
 scale_fill_manual(values=vegPal, na.value=”darkgrey”) +
 scale_color_manual(values=vegPal, na.value=”darkgrey”) +
 theme(axis.text = element_text(size=11))
ggsave(“bar_byClassVeg.pdf”, device=”pdf”, width=6.5, height=4,
units=”in”, useDingbats=FALSE)
ggsave(“bar_byClassVeg.png”, device=”png”, width=4.5, height=4.5,
units=”in”)

# test for statistical significance
class.merge.relTab <- data.frame(getTab(class.mergeRel))
colnames(class.merge.relTab) <- tax_table(class.mergeRel)[,3]
class.merge.relTab$Nearby_Veg <- as.factor(sample_data(class.
mergeRel)$Nearby_Veg[match(sample_data(class.mergeRel)$SampleID, row.
names(class.merge.relTab))])
# for each of the top 10 Classes
classNames <- colnames(class.merge.relTab)[1:51]

replications(Actinobacteria ~ Nearby_Veg, data=urb.nosing2.relTab)
kruskal.test(Actinobacteria ~ Nearby_Veg, data=class.merge.relTab)

wilcoxRes <- lapply(class.merge.relTab[1:51], function(x){pairwise.
wilcox.test(x, class.merge.relTab$Nearby_Veg, p.adjust.method =
“bonferroni”)})
```

```
sink("wilcoxRes.txt")
wilcoxRes
sink()

```

```{r famBar}
# dodged barplot comparing relative abundance of significant families
across land cover types

## split into pairwise land cover type comparisons
noForest <- prune_samples(sample_data(fam.merge)$Nearby_Veg !=
"forest", fam.merge)
noForest <- prune_taxa(taxa_sums(noForest) != 0, noForest)
sample_data(noForest)$Nearby_Veg # check that there are no forest sites
noForest.des <- phyloseq_to_deseq2(noForest, ~ Date + Nearby_Veg)
noForest.glm <- DESeq(noForest.des, test="Wald", fitType="parametric")
res.noForest.glm <- results(noForest.glm, cooksCutoff=TRUE, alpha=0.05)
summary(res.noForest.glm)
mcols(res.noForest.glm, use.names=TRUE)
res.noForest.glm.df <- subset(res.noForest.glm, padj < 0.05)
res.noForest.glm.df <- cbind(as(res.noForest.glm.df, "data.frame"),
        as(tax_table(fam.merge)[rownames(res.noForest.glm.df),],
"matrix"),
        as(t(otu_table(fam.mergeRel)[,rownames(res.noForest.glm.df)]),
"matrix"),
        stringsAsFactors = FALSE
        )
res.noForest.glm.df$SequenceVariant <- rownames(res.noForest.glm.df)
res.noForest.glm.df$SiteType <- "grass"
res.noForest.glm.df$SiteType[res.noForest.glm.df$log2FoldChange >= 0]
<- "paved"
res.noForest.glm.df$SiteType <- as.factor(res.noForest.glm.df$SiteType)
res.noForest.glm.df <- res.noForest.glm.df[order(res.noForest.glm.
df$log2FoldChange),] # order by l2fc
#knitr::kable(res.noForest.glm.df)
write.csv(res.noForest.glm.df, file="deseq_noForest_fam.csv")

noGrass <- prune_samples(sample_data(fam.merge)$Nearby_Veg != "grass",
fam.merge)
noGrass <- prune_taxa(taxa_sums(noGrass) != 0, noGrass)
sample_data(noGrass)$Nearby_Veg # check that there are no grass sites
noGrass.des <- phyloseq_to_deseq2(noGrass, ~ Date + Nearby_Veg)
noGrass.glm <- DESeq(noGrass.des, test="Wald", fitType="parametric")
res.noGrass.glm <- results(noGrass.glm, cooksCutoff=TRUE, alpha=0.05)
summary(res.noGrass.glm)
mcols(res.noGrass.glm, use.names=TRUE)
res.noGrass.glm.df <- subset(res.noGrass.glm, padj < 0.05)
res.noGrass.glm.df <- cbind(as(res.noGrass.glm.df, "data.frame"),
        as(tax_table(fam.merge)[rownames(res.noGrass.glm.df),],
```

```
“matrix”),
        as(t(otu_table(fam.mergeRel)[,rownames(res.noGrass.glm.df)]),
“matrix”),
        stringsAsFactors = FALSE
        )
res.noGrass.glm.df$SequenceVariant <- rownames(res.noGrass.glm.df)
res.noGrass.glm.df$SiteType <- “forest”
res.noGrass.glm.df$SiteType[res.noGrass.glm.df$log2FoldChange >= 0] <-
“paved”
res.noGrass.glm.df$SiteType <- as.factor(res.noGrass.glm.df$SiteType)
res.noGrass.glm.df <- res.noGrass.glm.df[order(res.noGrass.glm.
df$log2FoldChange),] # order by l2fc
#knitr::kable(res.noGrass.glm.df)
write.csv(res.noGrass.glm.df, file=”deseq_noGrass_fam.csv”)

noGrey <- prune_samples(sample_data(fam.merge)$Nearby_Veg != “paved”,
fam.merge)
noGrey <- prune_taxa(taxa_sums(noGrey) != 0, noGrey)
sample_data(noGrey)$Nearby_Veg # check that there are no paved sites
noGrey.des <- phyloseq_to_deseq2(noGrey, ~ Date + Nearby_Veg)
noGrey.glm <- DESeq(noGrey.des, test=”Wald”, fitType=”parametric”)
res.noGrey.glm <- results(noGrey.glm, cooksCutoff=TRUE, alpha=0.05)
summary(res.noGrey.glm)
mcols(res.noGrey.glm, use.names=TRUE)
res.noGrey.glm.df <- subset(res.noGrey.glm, padj < 0.05)
res.noGrey.glm.df <- cbind(as(res.noGrey.glm.df, “data.frame”),
        as(tax_table(fam.merge)[rownames(res.noGrey.glm.df),],
“matrix”),
        as(t(otu_table(fam.mergeRel)[,rownames(res.noGrey.glm.df)]),
“matrix”),
        stringsAsFactors = FALSE
        )
res.noGrey.glm.df$SequenceVariant <- rownames(res.noGrey.glm.df)
res.noGrey.glm.df$SiteType <- “grass”
res.noGrey.glm.df$SiteType[res.noGrey.glm.df$log2FoldChange >= 0] <-
“forest”
res.noGrey.glm.df$SiteType <- as.factor(res.noGrey.glm.df$SiteType)
res.noGrey.glm.df <- res.noGrey.glm.df[order(res.noGrey.glm.
df$log2FoldChange),] # order by l2fc
#knitr::kable(res.noGrey.glm.df)
write.csv(res.noGrey.glm.df, file=”deseq_noGrey_fam.csv”)

sigFamList <- data.frame(“Family”=res.noForest.glm.df$Family,
“l2fc”=res.noForest.glm.df$log2FoldChange)
sigFamList <- bind_rows(sigFamList, data.frame(“Family”=res.noGrass.
glm.df$Family, “l2fc”=res.noGrass.glm.df$log2FoldChange))
sigFamList <- bind_rows(sigFamList, data.frame(“Family”=res.noGrey.glm.
df$Family, “l2fc”=res.noGrey.glm.df$log2FoldChange))
sigFamList <- sigFamList[order(abs(sigFamList$l2fc)),] # order by
absolute value of l2fc
```

```
sigFamList <- unique(sigFamList) # remove duplicate rows
sigFamList <- sigFamList[abs(sigFamList$l2fc) > 0.25,]

sigFam <- subset_taxa(fam.merge, Family %in% sigFamList$Family) #get
differentially abundant Families identified by deSeq
sort.family <- sort(tapply(taxa_sums(sigFam), tax_table(sigFam)[,
"Family"], sum), TRUE)[1:30]
top.sigFam <- subset_taxa(sigFam, Family %in% names(sort.family)) #
take only Families > 9000 total reads

# convert your processed phyloseq object into a dataframe
df <- psmelt(top.sigFam)
# group by Nearby_Veg and Family, calculate mean abundance and standard
deviation
se <- function(x) sqrt(var(x)/length(x))
avgs <- ddply(df, .(Nearby_Veg, Family), plyr::summarize,
       mean = mean(Abundance),
       se = se(Abundance))
avgs$Nearby_Veg <- as.factor(avgs$Nearby_Veg)
# plot bar graph with standard deviation as error bars
ggplot(avgs, aes(fill=Nearby_Veg, x=Family, y=mean, group=Nearby_Veg)) +
 geom_bar(aes(color=Nearby_Veg, fill=Nearby_Veg), stat="identity",
position=position_dodge(width=0.8), width=0.5) +
 coord_flip() +
 geom_errorbar(aes(ymin=mean-se, ymax=mean+se), position=position_
dodge(width=0.8), width=0.5, size=0.15) +
 ylab("Percentage of Sequences") +
 scale_fill_manual(values=vegPal, na.value="darkgrey") +
 scale_color_manual(values=vegPal, na.value="darkgrey")
ggsave("bar_byFamVeg.pdf", device="pdf", width=5, height=6, units="in",
useDingbats=FALSE)
ggsave("bar_byFamVeg.png", device="png", width=4.5, height=4.5,
units="in")

# test for statistical significance
urb.nosing2.relTab <- data.frame(getTab(urb.nosing2.rel))
colnames(urb.nosing2.relTab) <- tax_table(urb.nosing2.rel)[,3]
urb.nosing2.relTab$Nearby_Veg <- as.factor(sample_data(urb.nosing2.
rel)$Nearby_Veg[match(sample_data(urb.nosing2.rel)$SampleID, row.
names(urb.nosing2.relTab))])
# for each of the top 10 Classes
replications(Actinobacteria ~ Nearby_Veg, data=urb.nosing2.relTab)
kruskal.test(Actinobacteria ~ Nearby_Veg, data=urb.nosing2.relTab)
pairwise.wilcox.test(urb.nosing2.relTab$Actinobacteria, urb.nosing2.
relTab$Nearby_Veg,
         p.adjust.method = "bonferroni")
```

### Health-relevant taxa

```{r healthTaxa, eval=TRUE}

barton <- subset_taxa(urb.nosing.vst, Family==”Bartonellaceae”) #good
barton.tab <- data.frame(getTab(barton))
barton.tab$Nearby_Veg <- sample_data(urb.nosing.vst)$Nearby_
Veg[match(row.names(barton.tab), row.names(sample_data(urb.nosing.
vst)))]
summary.lm(aov(barton.tab[,1] ~ Nearby_Veg, data=barton.tab))
barton.nograss <- subset(barton.tab, Nearby_Veg != “grass”)
summary.lm(aov(barton.nograss[,1] ~ Nearby_Veg, data=barton.nograss))
barton.noforest <- subset(barton.tab, Nearby_Veg != “forest”)
summary.lm(aov(barton.noforest[,1] ~ Nearby_Veg, data=barton.noforest))
# no significant differences

acinet <- subset_taxa(urb.nosing.vst, Genus==”Acinetobacter”) #good
acinet.tab <- data.frame(getTab(acinet))
acinet.tab$Nearby_Veg <- sample_data(urb.nosing.vst)$Nearby_
Veg[match(row.names(acinet.tab), row.names(sample_data(urb.nosing.
vst)))]
length(colnames(acinet.tab))
pairwise.wilcox.test(urb.nosing2.relTab$Actinobacteria, urb.nosing2.
relTab$Nearby_Veg,
         p.adjust.method = “bonferroni”)
summary.lm(aov(acinet.tab[,1] ~ Nearby_Veg, data=acinet.tab))
acinet.nograss <- subset(acinet.tab, Nearby_Veg != “grass”)
summary.lm(aov(acinet.nograss[,1] ~ Nearby_Veg, data=acinet.nograss))
# somewhat significant difference between forest and paved (R2=0.03,
p=0.031), not at all between grass and paved
acinet.noforest <- subset(acinet.tab, Nearby_Veg != “forest”)
summary.lm(aov(acinet.noforest[,1] ~ Nearby_Veg, data=acinet.noforest))
acinet.nopaved <- subset(acinet.tab, Nearby_Veg != “paved”)
summary.lm(aov(acinet.nopaved[,1] ~ Nearby_Veg, data=acinet.nopaved))
# more significant difference between forest and grass (R2=0.0787,
p=0.0033)
ggplot(acinet.tab) +
  geom_boxplot(aes(x=acinet.tab$Nearby_Veg, y=acinet.tab[1],
color=acinet.tab$Nearby_Veg)) +
  labs(x=””, y=””) +
  scale_color_manual(values=c(“turquoise3”, “darkgoldenrod”, “grey30”))
+
  theme(legend.position=”none”)

#diaphor <- subset_taxa(urb.nosing.vst, Genus==”Diaphorobacter”) #good
# none present

microbac <- subset_taxa(urb.nosing.vst, Genus==”Microbacterium”) #bad
microbac.tab <- data.frame(getTab(microbac))
microbac.tab$Nearby_Veg <- sample_data(urb.nosing.vst)$Nearby_
Veg[match(row.names(microbac.tab), row.names(sample_data(urb.nosing.
vst)))]
```

```
summary.lm(aov(microbac.tab[,1] ~ Nearby_Veg, data=microbac.tab))
microbac.nograss <- subset(microbac.tab, Nearby_Veg != “grass”)
summary.lm(aov(microbac.nograss[,1] ~ Nearby_Veg, data=microbac.
nograss))
microbac.noforest <- subset(microbac.tab, Nearby_Veg != “forest”)
summary.lm(aov(microbac.noforest[,1] ~ Nearby_Veg, data=microbac.
noforest))
# no significant differences

alcal <- subset_taxa(urb.nosing.vst, Genus==”Alcaligenes”) #bad
alcal.tab <- data.frame(getTab(alcal))
alcal.tab$Nearby_Veg <- sample_data(urb.nosing.vst)$Nearby_
Veg[match(row.names(alcal.tab), row.names(sample_data(urb.nosing.
vst)))]
summary.lm(aov(alcal.tab[,1] ~ Nearby_Veg, data=alcal.tab))
alcal.nograss <- subset(alcal.tab, Nearby_Veg != “grass”)
summary.lm(aov(alcal.nograss[,1] ~ Nearby_Veg, data=alcal.nograss))
alcal.noforest <- subset(alcal.tab, Nearby_Veg != “forest”)
summary.lm(aov(alcal.noforest[,1] ~ Nearby_Veg, data=alcal.noforest))
# no significant differences

#sciuri <- subset_taxa(urb.nosing.vst, Species==”sciuri”) #good
# none present

tax_table(urb.nosing.vst)@.Data[,7][“TACGTAGGTCCCGAGCGTTGTCCGGAT
TTATTGGGCGTAAAGCGAGCGCAGGTGGTTTATTAAGTCTGGTGTAAAAGGCAGTGGCTCAACC
ATTGTATGCATTGGAAACTGGTAGACTTGAGTGCAGGAGAGGAGAGTGGAATTCCATGTGTAGC
GGTGAAATGCGTAGATATATGGAGGAACACCGGGGGCGAAAGCGGCTCTCTGGCCTGTAACTGA
CACTGAGGCTCGAAAGCGTGGGGAGCAAACAGG”] <- “lactis”
lactis <- subset_taxa(urb.nosing.vst, Genus==”Lactococcus”)
lactis <- subset_taxa(lactis, Species==”lactis”) #good
lactis.tab <- data.frame(getTab(lactis))
lactis.tab$Nearby_Veg <- sample_data(urb.nosing.vst)$Nearby_
Veg[match(row.names(lactis.tab), row.names(sample_data(urb.nosing.
vst)))]
summary.lm(aov(lactis.tab[,1] ~ Nearby_Veg, data=lactis.tab))
lactis.nograss <- subset(lactis.tab, Nearby_Veg != “grass”)
summary.lm(aov(lactis.nograss[,1] ~ Nearby_Veg, data=lactis.nograss))
lactis.noforest <- subset(lactis.tab, Nearby_Veg != “forest”)
summary.lm(aov(lactis.noforest[,1] ~ Nearby_Veg, data=lactis.noforest))
# no significant differences

iners <- subset_taxa(urb.nosing.vst, Genus==”Lactobacillus”)
iners <- subset_taxa(iners, Species==”iners”) #good
iners.tab <- data.frame(getTab(iners))
iners.tab$Nearby_Veg <- sample_data(urb.nosing.vst)$Nearby_
Veg[match(row.names(iners.tab), row.names(sample_data(urb.nosing.
vst)))]
summary.lm(aov(iners.tab[,1] ~ Nearby_Veg, data=iners.tab))
iners.nograss <- subset(iners.tab, Nearby_Veg != “grass”)
```

```
summary.lm(aov(iners.nograss[,1] ~ Nearby_Veg, data=iners.nograss))
iners.noforest <- subset(iners.tab, Nearby_Veg != “forest”)
summary.lm(aov(iners.noforest[,1] ~ Nearby_Veg, data=iners.noforest))
# no significant differences

#morax <- subset_taxa(urb.nosing.vst, Genus==”Moraxella”) #bad
# none present

tax_table(urb.nosing.vst)@.Data[,7][“TACGTAGGGTCCGAGCGTTGTCCGGAA
TTACTGGGCGTAAAGAGCTCGTAGGTGGTTTGTCGCGTTGTTCGTGAAAACTCACAGCTCAACT
GTGGGCGTGCGGGCGATACGGGCAGACTAGAGTACTGCAGGGGAGACTGGAATTCCTGGTGTAG
CGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGGTCTCTGGGCAGTAACTG
ACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGG”] <- “vaccae”
vaccae <- subset_taxa(urb.nosing.vst, Species==”vaccae”) #good
tax_table(vaccae) # check that it is correct
vaccae.tab <- data.frame(getTab(vaccae))
vaccae.tab$Nearby_Veg <- sample_data(urb.nosing.vst)$Nearby_
Veg[match(row.names(vaccae.tab), row.names(sample_data(urb.nosing.
vst)))]
summary.lm(aov(vaccae.tab[,1] ~ Nearby_Veg, data=vaccae.tab))
vaccae.nograss <- subset(vaccae.tab, Nearby_Veg != “grass”)
summary.lm(aov(vaccae.nograss[,1] ~ Nearby_Veg, data=vaccae.nograss))
vaccae.noforest <- subset(vaccae.tab, Nearby_Veg != “forest”)
summary.lm(aov(vaccae.noforest[,1] ~ Nearby_Veg, data=vaccae.noforest))
# no significant differences

gammap <- subset_taxa(urb.nosing.vst, Class==”Gammaproteobacteria”)
#good
gammap.gen <- tax_glom(gammap, taxrank=”Genus”)
gammap.tab <- getTab(gammap.gen)
gammap.div <- data.frame(diversity(gammap.tab, index=”shannon”))
gammap.div$Nearby_Veg <- sample_data(urb.nosing.vst)$Nearby_
Veg[match(row.names(gammap.div), row.names(sample_data(urb.nosing.
vst)))]
colnames(gammap.div[1]) <- “Diversity”
gammap.div$veg50 <- sample_data(urb.nosing.vst)$veg50[match(row.
names(gammap.div), row.names(sample_data(urb.nosing.vst)))]
summary.lm(aov(gammap.div[,1] ~ veg50, data=gammap.div))
summary.lm(aov(gammap.div[,1] ~ Nearby_Veg, data=gammap.div))
ggplot(gammap.div) +
  geom_boxplot(aes(x=gammap.div$Nearby_Veg, y=gammap.div[1],
color=gammap.div$Nearby_Veg)) +
  scale_y_continuous(limits=c(3.6,3.8)) +
  labs(x=””, y=””) +
  scale_color_manual(values=c(“turquoise3”, “darkgoldenrod”, “grey30”))
+
  theme(legend.position=”none”)
gammap.nograss <- subset(gammap.div, Nearby_Veg != “grass”)
summary.lm(aov(gammap.nograss[,1] ~ Nearby_Veg, data=gammap.nograss))
gammap.noforest <- subset(gammap.div, Nearby_Veg != “forest”)
```

```
summary.lm(aov(gammap.noforest[,1] ~ Nearby_Veg, data=gammap.noforest))

```
We tested for associations between vegetation cover and the following health-relevant environmental taxa that have been identified:
* Bartonellaceae (Amish; Stein et al. 2016)

* Acinetobacter [lwoffii] (Ruokolainen, Hanski et al. 2012)

* Gammaproteobacteria, generic diversity rather than relative abundance (Hanski et al. 2012)

* Diaphorobacter [good], Microbacterium [bad], Alcaligenes [bad] (Fyhrquist et al. 2014)

* Staphylococcus sciuri [good], Lactococcus lactis [good], Lactobacillus iners [good], Moraxella [bad] (von Mutius 2016)

* Mycobacteria vaccae (Lowry)

### Contaminant relative abundance
```{r contamPlot, include=TRUE, results="hide"}
################## THIS IS FROM JAMES'S CELL PHONE PAPER
# plot relative abundances of potential contaminants in controls vs samples
plot(plotY ~ plotX,
   pch=21, bg=rgb(0,0,0,.3), cex=2, las=1,
   xlab='Rel Abundance in Experiment', ylab='Rel Abundance in Controls')
segments(0,0,1,1, lty=3, lwd=2, col='gray')
segments(0, .05, 1, .05, lty=1, lwd=2, col='tomato')
text(.12, .13, '1:1', font=3, col='gray30')
text(.1, .05, 'RA=0.05', font=3, pos=3, col='tomato')
text(plotX[names(cont3.otus)[1:3]], plotY[names(cont3.otus)[1:3]],
   rw.taxo.tmp[names(cont3.otus)[1:3], 'Genus'], pos=c(1))

```


### Community composition by sampling date
```{r dateBar, include=TRUE, results="hide", fig.width=6.5, fig.height=6.5, fig.show="hold", fig.align="center", fig.cap="\\label{fig:barSite}Barplot showing composition by individual site for each sampling date."}
dateList.rel <- llply(dateList, function(x){transform_sample_counts(x, function(x) 100 * x/sum(x))})

## stacked barplots to compare proportional composition of top 25 families (aggregated) for every site BY DATE
for (i in seq_along(dateList)){
 sort.class <- sort(tapply(taxa_sums(dateList[[i]]), tax_
```

```
table(dateList[[i]])[, “Family”], sum), TRUE)
 top.class <- sort.class[1:25] #what are the top 25 most abundant
Families?
 bottom.class <- sort.class[26:length(sort.class)]
 urb.nosing1 <- subset_taxa(dateList[[i]], Family %in% names(top.
class)) #get top 25 most abundant Family
 urb.nosing2 <- subset_taxa(dateList[[i]], Family %in% names(bottom.
class)) #get all other taxa
 urb.nosing2 <- merge_taxa(dateList[[i]], taxa_names(urb.nosing2),
archetype=1) #merge all other taxa into Family “Other”
 tax_table(urb.nosing2)[,5][is.na(tax_table(urb.nosing2)[,5])] <-
“AallOthers”
 urb.nosing2 <- tax_glom(urb.nosing2, taxrank=”Family”)
 urb.nosing2 <- transform_sample_counts(urb.nosing2, function(x) 100 *
x/sum(x))
 # reorder levels by Nearby_Veg
 sample_data(urb.nosing2)$SiteCode <- factor(
  sample_data(urb.nosing2)$SiteCode, levels=sample_data(urb.
nosing2)$SiteCode[order(sample_data(urb.nosing2)$Nearby_Veg)])
 plot_bar(urb.nosing2, x=”SiteCode”, fill = “Family”,
title=paste(names(dateList)[i])) +
  ylab(“Percentage of Sequences”) +
  geom_bar(aes(fill=Family, color=Family), stat=”identity”,
position=”stack”) +
  scale_fill_manual(values=famPal, na.value=”darkgrey”) +
  scale_color_manual(values=famPal, na.value=”darkgrey”)
 ggsave(paste0(“fam25bar_”, names(dateList)[i], “.pdf”), device=”pdf”,
width=7, height=4, units=”in”)
}

## stacked barplots to compare proportional composition of top 25 SVs
for every site BY DATE
for (i in seq_along(dateList)){
 top25 <- sort(taxa_sums(dateList.rel[[i]]), TRUE)[1:25]
 top25 <- prune_taxa(names(top25), dateList.rel[[i]])
 top25_sums <- data.frame(taxa_sums(top25))
 # get taxonomic Families of top 25
 top25tax <- tax_table(top25)[, c(“Phylum”, “Class”, “Order”, “Family”,
“Genus”)]
 top25matrix <- data.frame(as(top25tax, “matrix”))
 top25matrix <- cbind(top25matrix, top25_sums)
 write.csv(top25matrix, paste0(“top25_”, names(dateList.rel[i]),
“.csv”))
 top25rel <- transform_sample_counts(top25, function(x) 100 * x/sum(x))
 # reorder levels by Nearby_Veg
 sample_data(top25rel)$SiteCode <- factor(
  sample_data(top25rel)$SiteCode, levels=sample_
data(top25rel)$SiteCode[order(sample_data(top25rel)$Nearby_Veg)])
 plot_bar(top25rel, x=”SiteCode”, fill = “Family”,
title=paste(names(dateList)[i])) +
```

```
  ylab("Percentage of Sequences") +
  geom_bar(aes(fill=Family, color=Family), stat="identity",
position="stack") +
  scale_fill_manual(values=famPal, na.value="darkgrey") +
  scale_color_manual(values=famPal, na.value="darkgrey")
 ggsave(paste0("SV25bar_", names(dateList)[i], ".pdf"), device="pdf",
width=7, height=4, units="in")
}


for (i in seq_along(dateList.rel)){
 top25 <- sort(taxa_sums(dateList.rel[[i]]), TRUE)[1:25]
 top25 <- prune_taxa(names(top25), dateList.rel[[i]])
 top25_sums <- data.frame(taxa_sums(top25))
 # get taxonomic Families of top 25
 top25tax <- tax_table(top25)[, c("Phylum", "Class", "Order", "Family",
"Genus")]
 top25matrix <- data.frame(as(top25tax, "matrix"))
 top25matrix <- cbind(top25matrix, top25_sums)
 write.csv(top25matrix, paste0("top25_", names(dateList.rel[i]),
".csv"))
}

#clean up
rm(top25otus, top25_sums, top25tax, top25matrix)

```


### Relative abundances of individual taxa of interest across sites and
sampling dates
```{r barTaxa, include=TRUE, results="hide", fig.width=6.5, fig.
height=6.5, fig.show="hold", fig.align="center", fig.cap="\\
label{fig:barTaxa}Barplots showing distribution of individual taxa of
interest across sites and sampling dates."}

Sphingomonas <- prune_taxa("TACGGAGGGAGCTAGCGTTATTCGGAATTACTGGGC
GTAAAGCGCACGTAGGCGGCTTTGTAAGTAAGAGGTGAAAGCCCAGAGCTCAACTCTGGAATTG
CCTTTTAGACTGCATCGCTTGAATCATGGAGAGGTCAGTGGAATTCCGAGTGTAGAGGTGAAAT
TCGTAGATATTCGGAAGAACACCAGTGGCGAAGGCGGCTGACTGGACATGTATTGACGCTGAGG
TGCGAAAGCGTGGGGAGCAAACAGG", urb.nosing) # interesting taxa from top 25,
associated with hay dust and observed in pilot study
# reorder levels by Date then Nearby_Veg
sample_data(Sphingomonas)$SampleID <- factor(sample_
data(Sphingomonas)$SampleID,
                       levels = sample_
data(Sphingomonas)$SampleID[order(sample_data(Sphingomonas)$Date,
                                                sample_
data(Sphingomonas)$Nearby_Veg)],
                       ordered=TRUE)
plot_bar(Sphingomonas, x="SampleID", y="Abundance", fill="Nearby_Veg") +
```

```
 geom_bar(aes(color=Nearby_Veg, fill=Nearby_Veg), stat="identity",
position="stack") +
 scale_fill_manual(values=c("turquoise3", "darkgoldenrod", "grey30")) +
 scale_color_manual(values=c("turquoise3", "darkgoldenrod", "grey30"))
+
 theme(text=element_text(size=8), axis.text.x=element_text(size=6,
vjust=0.5))
ggsave("Sphingomonas_bar.pdf", device="pdf", width=12, height=4,
units="in")

A.multivorum <- prune_taxa("TACGAAGGGGGCTAGCGTTGCTCGGAATGACTGGGC
GTAAAGGGCGCGTAGGCGGATCGGACAGTCAGGCGTGAAATTCCTGGGCTTAACCTGGGGGCTG
CGTTTGAGACGTTGGGTCTTGAGTTTGGAAGAGGGTCGTGGAATTCCCAGTGTAGAGGTGAAAT
TCGTAGATATTGGGAAGAACACCGGTGGCGAAGGCGGCGACCTGGTCCTGGACTGACGCTGAGG
CGCGAAAGCGTGGGGAGCAAACAGG", urb.nosing) # interesting taxa from top 25,
breaks down heavy metals
# reorder levels by Date then Nearby_Veg
sample_data(A.multivorum)$SampleID <- factor(sample_
data(A.multivorum)$SampleID, levels = sample_data(A.
multivorum)$SampleID[order(sample_data(A.multivorum)$Date, sample_
data(A.multivorum)$Nearby_Veg)], ordered=TRUE)
plot_bar(A.multivorum, x="SampleID", y="Abundance", fill="Nearby_Veg") +
 geom_bar(aes(color=Nearby_Veg, fill=Nearby_Veg), stat="identity",
position="stack") +
 scale_fill_manual(values=c("turquoise3", "darkgoldenrod", "grey30")) +
 scale_color_manual(values=c("turquoise3", "darkgoldenrod", "grey30"))
+
 theme(text=element_text(size=8), axis.text.x=element_text(size=6,
vjust=0.5))
ggsave("A.multivorum_bar.pdf", device="pdf", width=12, height=4,
units="in")

#Erwinia <- prune_taxa("TACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAA
AGCGCACGCAGGCGGTCTGTCAAGTCAGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATT
TGAAACTGGCAGGCTAGAGTCTTGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGT
AGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAAGACTGACGCTCAGGTGCG
AAAGCGTGGGGAGCAAACAGG", urb.nosing) # interesting taxa from top 25,
associated with water and plant disease
# reorder levels by Date then Nearby_Veg
#sample_data(Erwinia)$SampleID <- factor(sample_data(Erwinia)$SampleID,
levels = sample_data(Erwinia)$SampleID[order(sample_
data(Erwinia)$Nearby_Veg, sample_data(Erwinia)$SiteCode)],
ordered=TRUE)
#plot_bar(Erwinia, x="SampleID", y="Abundance", fill="Nearby_Veg") +
# geom_bar(aes(color=Nearby_Veg, fill=Nearby_Veg), stat="identity",
position="stack") +
# scale_fill_manual(values=c("turquoise3", "darkgoldenrod", "grey30")) +
# scale_color_manual(values=c("turquoise3", "darkgoldenrod", "grey30"))
+
# theme(text=element_text(size=8), axis.text.x=element_text(size=6,
```

```
vjust=0.5))
#ggsave("Erwinia_bar.pdf", device="pdf", width=12, height=4,
units="in")

#M.radiotolerans <- prune_taxa("TACGAAGGGGGCTAGCGTTGCTCGGAATCACT
GGGCGTAAAGGGCGCGTAGGCGGCGTTTTAAGTCGGGGGTGAAAGCCTGTGGCTCAACCACAGA
ATGGCCTTCGATACTGGGACGCTTGAGTATGGTAGAGGTTGGTGGAACTGCGAGTGTAGAGGTG
AAATTCGTAGATATTCGCAAGAACACCGGTGGCGAAGGCGGCCAACTGGACCATTACTGACGCT
GAGGCGCGAAAGCGTGGGGAGCAAACAGG", urb.nosing) # interesting taxa from top
25,
# reorder levels by Date then Nearby_Veg
#sample_data(M.radiotolerans)$SampleID <- factor(sample_
data(M.radiotolerans)$SampleID, levels = sample_data(M.
radiotolerans)$SampleID[order(sample_data(M.radiotolerans)$Date,
sample_data(M.radiotolerans)$Nearby_Veg)], ordered=TRUE)

#Pantoea <- prune_taxa("TACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAA
AGCGCACGCAGGCGGTCTGTTAAGTCAGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATT
TGAAACTGGCAGGCTTGAGTCTTGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGT
AGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAAGACTGACGCTCAGGTGCG
AAAGCGTGGGGAGCAAACAGG", urb.nosing) # interesting taxa from top 25,
commonly isolated from grain dust
# reorder levels by Date then Nearby_Veg
#sample_data(Pantoea)$SampleID <- factor(sample_data(Pantoea)$SampleID,
levels = sample_data(Pantoea)$SampleID[order(sample_
data(Pantoea)$Nearby_Veg, sample_data(Pantoea)$SiteCode)],
ordered=TRUE)

# clean up
rm(urb.nosing.m, top25phy)

```

## REFERENCES CITED

### Chapter I


Bell, M.L., Ebisu, K., 2012. Environmental inequality in exposures to airborne particulate matter components in the United States. Environ Health Perspect. https://doi.org/10.1289/ehp.1205201

Benedict, M.A., McMahon, E.T., 2002. Green infrastructure: Smart conservation for the 21st century. Renew Resources J 20, 12–17.

Blaser, M.J., Falkow, S., 2009. What are the consequences of the disappearing human microbiota? Nat Rev Microbiol 7, 887–894.

Breuste, J., Niemelä, J., Snep, R.P.H., 2008. Applying landscape ecological principles in urban environments. Landscape Ecol 23, 1139–1142. https://doi.org/10.1007/s10980-008-9273-0

Brown, E.M., Arrieta, M.-C., Finlay, B.B., 2013. A fresh look at the hygiene hypothesis: How intestinal microbial exposure drives immune effector responses in atopic disease. Semin Immunol 25, 378–387. https://doi.org/10.1016/j.smim.2013.09.003

Crane, P., Kinzig, A., 2005. Nature in the metropolis. Science 308, 1225–1225. https://doi.org/10.1126/science.1114165

Dai, D., 2011. Racial/ethnic and socioeconomic disparities in urban green space accessibility: Where to intervene? Landsc Urban Plan 102, 234–244. https://doi.org/10.1016/j.landurbplan.2011.05.002

Dallimer, M., Tang, Z., Bibby, P.R., Brindley, P., Gaston, K.J., Davies, Z.G., 2011. Temporal changes in greenspace in a highly urbanized region. Biol Lett 7, 763–766. https://doi.org/10.1098/rsbl.2011.0025

Davis, K., 1955. The origin and growth of urbanization in the world. AJS 60, 429–437.

Eder, W., Ege, M.J., von Mutius, E., 2006. The asthma epidemic. N Engl J Med 355, 2226–2235. https://doi.org/10.1056/NEJMra054308

Ege, M.J., Mayer, M., Normand, A.-C., Genuneit, J., Cookson, W.O., Braun-Fahrländer, C., Heederik, D., Piarroux, R., von Mutius, E., 2011. Exposure to environmental microorganisms and childhood asthma. N Engl J Med 364, 701–709.

Ege, M.J., Mayer, M., Schwaiger, K., Mattes, J., Pershagen, G., van Hage, M., Scheynius, A., Bauer, J., von Mutius, E., 2012. Environmental bacteria and childhood asthma. Allergy 1565-1571. https://doi.org/10.1111/all.12028

Fitzpatrick, K., LaGory, M., 2000. Unhealthy places: The ecology of risk in the urban landscape. Routledge, New York, NY.

Fyhrquist, N., Ruokolainen, L., Suomalainen, A., Lehtimäki, S., Veckman, V., Vendelin, J., Karisola, P., Lehto, M., Savinko, T., Jarva, H., Kosunen, T.U., Corander, J., Auvinen, P., Paulin, L., von Hertzen, L., Laatikainen, T., Mäkelä, M., Haahtela, T., Greco, D., Hanski, I., Alenius, H., 2014. *Acinetobacter* species in the skin microbiota protect against allergic sensitization and inflammation. J Allergy Clin Immunol 134, 1301–1309.e11. https://doi.org/10.1016/j.jaci.2014.07.059

García, R., White, A., 2006. Healthy parks, schools, and communities: Mapping green access and equity for the Los Angeles region. The City Project Policy Report. Accessed online in May 29, 2007.

Gilbert, S.F., Sapp, J., Tauber, A.I., 2012. A symbiotic view of life: We have never been individuals. Q Rev Biol 87, 325–341. https://doi.org/10.1086/668166

Haahtela, T., Holgate, S., Pawankar, R., Akdis, C.A., Benjaponpitak, S., Caraballo, L., Demain, J., Portnoy, J., von Hertzen, L., 2013. The biodiversity hypothesis and allergic disease: World Allergy Organization position statement. World Allergy Organ J 6, 3. https://doi.org/10.1186/1939-4551-6-3

Hanski, I., von Hertzen, L., Fyhrquist, N., Koskinen, K., Torppa, K., Laatikainen, T., Karisola, P., Auvinen, P., Paulin, L., Makela, M.J., Vartiainen, E., Kosunen, T.U., Alenius, H., Haahtela, T., 2012. Environmental biodiversity, human microbiota, and allergy are interrelated. PNAS 109, 8334–8339. https://doi.org/10.1073/pnas.1205624109

Harlan, S.L., Brazel, A.J., Prashad, L., Stefanov, W.L., Larsen, L., 2006. Neighborhood microclimates and vulnerability to heat stress. Soc Sci Med 63, 2847–2863. https://doi.org/10.1016/j.socscimed.2006.07.030

Harnik, P., Simms, J., 2004. Parks: How far is too far? Planning 70.

Heederik, D., von Mutius, E., 2012. Does diversity of environmental microbial exposure matter for the occurrence of allergy and asthma? J Allergy Clin Immunol 130, 44–50.

Hope, D., Gries, C., Zhu, W., Fagan, W.F., Redman, C.L., Grimm, N.B., Nelson, A.L., Martin, C., Kinzig, A., 2003. Socioeconomics drive urban plant diversity. PNAS 100, 8788–8792.

Jackson, L.E., 2003. The relationship of urban design to human health and condition. Landsc Urban Plan 64, 191–200.

Jennings, V., Johnson Gaither, C., Gragg, R.S., 2012. Promoting environmental justice through urban green space access: A synopsis. Environ Justice 5, 1–7. https://doi.org/10.1089/env.2011.0007

Kinzig, A.P., Warren, P., Martin, C., Hope, D., Katti, M., 2005. The effects of human

socioeconomic status and cultural characteristics on urban patterns of biodiversity. Ecol Soc 10, 23.

Kowarik, I., 2011. Novel urban ecosystems, biodiversity, and conservation. Environ Pollut 159, 1974–1983. https://doi.org/10.1016/j.envpol.2011.02.022

Landry, S.M., Chakraborty, J., 2009. Street trees and equity: Evaluating the spatial distribution of an urban amenity. Environ Plan A 41, 2651–2670. https://doi.org/10.1068/a41236

McKinney, M.L., 2006. Urbanization as a major cause of biotic homogenization. Biol Conserv 127, 247–260. https://doi.org/10.1016/j.biocon.2005.09.005

Meyerson, F.A., Merino, L., Durand, J., 2007. Migration and environment in the context of globalization. Front Ecol Environ 5, 182–190. https://doi.org/10.1890/1540-9295(2007)5[182:MAEITC]2.0.CO;2

Raison, C.L., Lowry, C.A., Rook, G.A., 2010. Inflammation, sanitation, and consternation: loss of contact with coevolved, tolerogenic microorganisms and the pathophysiology and treatment of major depression. Arch Gen Psychiatry 67, 1211.

Raison, C.L., Miller, A.H., 2013. Malaise, melancholia and madness: The evolutionary legacy of an inflammatory bias. Brain Behav Immun 31, 1–8. https://doi.org/10.1016/j.bbi.2013.04.009

Rook, G.A., 2013. Regulation of the immune system by biodiversity from the natural environment: An ecosystem service essential to health. PNAS USA. 110, 18360–18367. https://doi.org/10.1073/pnas.1313731110

Rook, G.A.W., 2010. 99th Dahlem Conference on Infection, Inflammation and Chronic Inflammatory Disorders: Darwinian medicine and the "hygiene" or "old friends" hypothesis. Am J Clin Exp Immunol 160, 70–79. https://doi.org/10.1111/j.1365-2249.2010.04133.x

Rook, G.A.W., Adams, V., Palmer, R., Brunet, L.R., Hunt, J., Martinelli, R., 2004. Mycobacteria and other environmental organisms as immunomodulators for immunoregulatory disorders. Springer Semin Immunopathol 25, 237–255. https://doi.org/10.1007/s00281-003-0148-9

Rook, G.A.W., Lowry, C.A., Raison, C.L., 2013. Microbial "Old Friends", immunoregulation and stress resilience. EMPH 2013, 46–64. https://doi.org/10.1093/emph/eot004

Ruokolainen, L., von Hertzen, L., Fyhrquist, N., Laatikainen, T., Lehtomäki, J., Auvinen, P., Karvonen, A., Hyvärinen, A.M., Tillmann, V., Niemelä, O., Knip, M., Haahtela, T., Pekkanen, J., Hanski, I., 2015. Green areas around homes reduce atopic sensitization in children. Allergy 195-202. https://doi.org/10.1111/all.12545

Russell, S.L., Gold, M.J., Hartmann, M., Willing, B.P., Thorson, L., Wlodarska, M., Gill, N., Blanchet, M.-R., Mohn, W.W., McNagny, K.M., Finlay, B.B., 2012. Early life antibiotic-driven changes in microbiota enhance susceptibility to allergic asthma. EMBO Rep13,

440–447. https://doi.org/10.1038/embor.2012.32

Sachs, J.S., 2007. Good germs, bad germs: health and survival in a bacterial world. Hill and Wang, New York.

Strohbach, M.W., Haase, D., Kabisch, N., 2009. Birds and the city: Urban biodiversity, land use, and socioeconomics. Ecol Soc 14, 31.

The Coalition for a Livable Future, 2007. The regional equity atlas: Metropolitan Portland's geography of opportunity. Portland, Oregon.

United Nations, 2012. World urbanization prospects: The 2011 revision. New York.

von Hertzen, L., Haahtela, T., 2006. Disconnection of man and the soil: Reason for the asthma and atopy epidemic? J Allergy Clin Immunol 117, 334–344. https://doi.org/10.1016/j.jaci.2005.11.013

von Hertzen, L., Hanski, I., Haahtela, T., 2011. Natural immunity. EMBO reports.

Young, C., Jarvis, P., Hooper, I., Trueman, I.C., 2009. Urban landscape ecology and its evaluation: a review, in: Dupont, A., Jacobs, H. (Eds.), Landscape Ecology Research Trends. Nova Science Publishers, New York, NY, pp. 45–69.

**Chapter II**

Adams, R.I., Miletto, M., Lindow, S.E., Taylor, J.W., Bruns, T.D., 2014. Airborne bacterial communities in residences: Similarities and differences with fungi. PLOS ONE 9, e91283. https://doi.org/10.1371/journal.pone.0091283

Adams, R.I., Miletto, M., Taylor, J.W., Bruns, T.D., 2013. Dispersal in microbes: Fungi in indoor air are dominated by outdoor air and show dispersal limitation at short distances. The ISME J 7, 1262–1273. https://doi.org/10.1038/ismej.2013.28

Barberán, A., Ladau, J., Leff, J.W., Pollard, K.S., Menninger, H.L., Dunn, R.R., Fierer, N., 2015. Continental-scale distributions of dust-associated bacteria and fungi. PNAS 112, 5756–5761. https://doi.org/10.1073/pnas.1420815112

Bell, K.S., Philp, J.C., Aw, D.W.J., Christofi, N., 1998. The genus *Rhodococcus*. J Appl Microbiol 85, 195–210.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 57, 289–300.

Blaser, M.J., Falkow, S., 2009. What are the consequences of the disappearing human microbiota? Nat Rev Microbiol 7, 887–894.

Bowers, R.M., McLetchie, S., Knight, R., Fierer, N., 2010. Spatial variability in airborne bacterial communities across land-use types and their relationship to the bacterial communities of potential source environments. The ISME J 5, 601–612. https://doi.org/10.1038/ismej.2010.167

Bowers, R.M., Sullivan, A.P., Costello, E.K., Collett, J.L., Knight, R., Fierer, N., 2011. Sources of bacteria in outdoor air across cities in the midwestern United States. Appl Environ Microbiol 77, 6350–6356. https://doi.org/10.1128/AEM.05498-11

Burrows, S.M., Elbert, W., Lawrence, M.G., Pöschl, U., 2009. Bacteria in the global atmosphere–Part 1: Review and synthesis of literature data for different ecosystems. Atmos Chem Phys 9, 9263–9280.

Busse, H.-J., 2003. *Sphingomonas aurantiaca* sp. nov., *Sphingomonas aerolata* sp. nov. and *Sphingomonas faeni* sp. nov., air- and dustborne and Antarctic, orange-pigmented, psychrotolerant bacteria, and emended description of the genus *Sphingomonas*. Int J Syst Evol Microbiol 53, 1253–1260. https://doi.org/10.1099/ijs.0.02461-0

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., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Meth 7, 335–336. https://doi.org/10.1038/nmeth.f.303

Claesson, M.J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J.R., Ross, R.P., O'Toole, P.W., 2010. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. Nucl Acids Res 38, e200–e200. https://doi.org/10.1093/nar/gkq873

Dadvand, P., de Nazelle, A., Figueras, F., Basagaña, X., Su, J., Amoly, E., Jerrett, M., Vrijheid, M., Sunyer, J., Nieuwenhuijsen, M.J., 2012. Green space, health inequality and pregnancy. Environ Int 40, 110–115. https://doi.org/10.1016/j.envint.2011.07.004

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

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., Andersen, G.L., 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72, 5069–5072. https://doi.org/10.1128/AEM.03006-05

Donovan, G.H., Butry, D.T., Michael, Y.L., Prestemon, J.P., Liebhold, A.M., Gatziolis, D., Mao, M.Y., 2013. The relationship between trees and human health. Am J Prev Med 44, 139–145. https://doi.org/10.1016/j.amepre.2012.09.066

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Ege, M.J., Mayer, M., Schwaiger, K., Mattes, J., Pershagen, G., van Hage, M., Scheynius, A., Bauer, J., von Mutius, E., 2012. Environmental bacteria and childhood asthma. Allergy 1565-1571. https://doi.org/10.1111/all.12028

Erb-Downward, J.R., Sadighi Akha, A.A., Wang, J., Shen, N., He, B., Martinez, F.J., Gyetko, M.R., Curtis, J.L., Huffnagle, G.B., 2012. Use of direct gradient analysis to uncover biological hypotheses in 16S survey data and beyond. Sci Rep 2. https://doi.org/10.1038/srep00774

Fall, T., Lundholm, C., Örtqvist, A.K., Fall, K., Fang, F., Hedhammar, Å., Kämpe, O., Ingelsson, E., Almqvist, C., 2015. Early exposure to dogs and farm animals and the risk of childhood asthma. JAMA Pediatr 169, e153219. https://doi.org/10.1001/jamapediatrics.2015.3219

Hanski, I., von Hertzen, L., Fyhrquist, N., Koskinen, K., Torppa, K., Laatikainen, T., Karisola, P., Auvinen, P., Paulin, L., Makela, M.J., Vartiainen, E., Kosunen, T.U., Alenius, H., Haahtela, T., 2012. Environmental biodiversity, human microbiota, and allergy are interrelated. PNAS 109, 8334–8339. https://doi.org/10.1073/pnas.1205624109

Huse, S., Welch, D.B.M., Sogin, M.L., 2012. Sequencing errors, diversity estimates, and the rare biosphere, in: The science and applications of microbial genomics. National Academies Press, Washington, D.C.

Innerebner, G., Knief, C., Vorholt, J.A., 2011. Protection of *Arabidopsis thaliana* against leaf-pathogenic *Pseudomonas syringae* by *Sphingomonas* strains in a controlled model system. Appl Environ Microbiol 77, 3202–3210. https://doi.org/10.1128/AEM.00133-11

Jumpponen, A., Jones, K.L., 2010. Seasonally dynamic fungal communities in the *Quercus macrocarpa* phyllosphere differ between urban and nonurban environments. New Phytol 186, 496–513. https://doi.org/10.1111/j.1469-8137.2010.03197.x

Kim, H., Nishiyama, M., Kunito, T., Senoo, K., Kawahara, K., Murakami, K., Oyaizu, H., 1998. High population of *Sphingomonas* species on plant surface. J Appl Microbiol 85, 731–736. https://doi.org/10.1111/j.1365-2672.1998.00586.x

Kinkel, L.L., 1997. Microbial population dynamics on leaves. Annu Rev Phytopathol 35, 327–347. https://doi.org/10.1146/annurev.phyto.35.1.327

Lambais, M.R., Lucheta, A.R., Crowley, D.E., 2014. Bacterial community assemblages associated with the phyllosphere, dermosphere, and rhizosphere of tree species of the Atlantic forest are host taxon dependent. Microb Ecol 68, 567–574. https://doi.org/10.1007/s00248-014-0433-2

Lighthart, B., 1984. Microbial aerosols: Estimated contribution of combine harvesting to an airshed. Appl Environ Microbiol 47, 430–432.

Lighthart, B., Shaffer, B.T., Frisch, A.S., Paterno, D., 2009. Atmospheric culturable bacteria associated with meteorological conditions at a summer-time site in the mid-Willamette Valley, Oregon. Aerobiologia 25, 285–295. https://doi.org/10.1007/s10453-009-9133-7

Lindemann, J., Constantinidou, H.A., Barchet, W.R., Upper, C.D., 1982. Plants as sources of airborne bacteria, including ice nucleation-active bacteria. Appl Environ Microbiol 44, 1059–1063.

Lindemann, J., Upper, C.D., 1985. Aerial dispersal of epiphytic bacteria over bean plants. Appl Environ Microbiol 50, 1229–1232.

Lindow, S.E., Brandl, M.T., 2003. Microbiology of the phyllosphere. Appl Environ Microbiol 69, 1875–1883. https://doi.org/10.1128/AEM.69.4.1875-1883.2003

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 1–21. https://doi.org/10.1186/s13059-014-0550-8

Lymperopoulou, D.S., Adams, R.I., Lindow, S.E., 2016. Contribution of vegetation to the microbial composition of nearby outdoor air. Appl Environ Microbiol 82, 3822–3833. https://doi.org/10.1128/AEM.00610-16

Maas, J., 2006. Green space, urbanity, and health: How strong is the relation? J Epidemiol Community Health 60, 587–592. https://doi.org/10.1136/jech.2005.043125

Maas, J., Verheij, R.A., de Vries, S., Spreeuwenberg, P., Schellevis, F.G., Groenewegen, P.P., 2009. Morbidity is related to a green living environment. J Epidemiol Community Health 63, 967–973. https://doi.org/10.1136/jech.2008.079038

Matthysse, A.G., 2006. The Genus Agrobacterium, in: Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., Stackebrandt, E. (Eds.), The Prokaryotes. Springer New York, New York, NY, pp. 91–114.

McMurdie, P.J., Holmes, S., 2013. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLOS ONE 8, e61217. https://doi.org/10.1371/journal.pone.0061217

Meadow, J.F., Altrichter, A.E., Kembel, S.W., Kline, J., Mhuireach, G., Moriyama, M., Northcutt, D., O'Connor, T.K., Womack, A.M., Brown, G.Z., Green, J.L.., Bohannan, B.J.M., 2013. Indoor airborne bacterial communities are influenced by ventilation, occupancy, and outdoor air source. Indoor Air 24, 41–48. https://doi.org/10.1111/ina.12047

Mitchell, R., Popham, F., 2007. Greenspace, urbanity and health: Relationships in England. J Epidemiol Community Health 61, 681–683. https://doi.org/10.1136/jech.2006.053553

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H., Wagner, H., 2015. vegan: Community ecology package.

Oren, A., 2006. The Genera Rhodothermus, Thermonema, Hymenobacter and Salinibacter, in: Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., Stackebrandt, E. (Eds.), The Prokaryotes. Springer New York, New York, NY, pp. 712–738.

Peñuelas, J., Terradas, J., 2014. The foliar microbiome. Trends Plant Sci. https://doi.org/10.1016/j.tplants.2013.12.007

R Development Core Team, 2010. R: A language and environment for statistical computing.

Rook, G.A., 2013. Regulation of the immune system by biodiversity from the natural environment: An ecosystem service essential to health. PNAS USA. 110, 18360–18367. https://doi.org/10.1073/pnas.1313731110

Ruokolainen, L., von Hertzen, L., Fyhrquist, N., Laatikainen, T., Lehtomäki, J., Auvinen, P., Karvonen, A., Hyvärinen, A.M., Tillmann, V., Niemelä, O., Knip, M., Haahtela, T., Pekkanen, J., Hanski, I., 2015. Green areas around homes reduce atopic sensitization in children. Allergy 195-202. https://doi.org/10.1111/all.12545

Russell, S.L., Gold, M.J., Hartmann, M., Willing, B.P., Thorson, L., Wlodarska, M., Gill, N., Blanchet, M.-R., Mohn, W.W., McNagny, K.M., Finlay, B.B., 2012. Early life antibiotic-driven changes in microbiota enhance susceptibility to allergic asthma. EMBO Rep13, 440–447. https://doi.org/10.1038/embor.2012.32

Shade, A., Jones, S.E., Caporaso, J.G., Handelsman, J., Knight, R., Fierer, N., Gilbert, J.A., 2014. Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. mBio 5, e01371-14-e01371-14. https://doi.org/10.1128/mBio.01371-14

Shaffer, B.T., Lighthart, B., 1997. Survey of culturable airborne bacteria at four diverse locations in Oregon: Urban, rural, forest, and coastal. Microbial ecology 34, 167–177.

Steyn, P.L., Segers, P., Vancanneyt, M., Sandra, P., Kersters, K., Joubert, J.J., 1998. Classification of heparinolytic bacteria into a new genus, *Pedobacter*, comprising four species: *Pedobacter heparinus* comb. nov., *Pedobacter piscium* comb. nov., *Pedobacter africanus* sp. nov. and *Pedobacter saltans* sp. nov. Proposal of the family Sphingobacteriaceae fam. nov. Int J Syst Evol Microbiol 48, 165–177. https://doi.org/10.1099/00207713-48-1-165

United Nations, 2012. World urbanization prospects: The 2011 revision. New York.

Villeneuve, P.J., Jerrett, M., G. Su, J., Burnett, R.T., Chen, H., Wheeler, A.J., Goldberg, M.S., 2012. A cohort study relating urban green space with mortality in Ontario, Canada. Environ Res 115, 51–58. https://doi.org/10.1016/j.envres.2012.03.003

Vokou, D., Vareli, K., Zarali, E., Karamanoli, K., Constantinidou, H.-I.A., Monokrousos, N., Halley, J.M., Sainis, I., 2012. Exploring biodiversity in the bacterial community of the Mediterranean phyllosphere and its relationship with airborne bacteria. Microb Ecol 64, 714–724. https://doi.org/10.1007/s00248-012-0053-7

von Hertzen, L., Haahtela, T., 2006. Disconnection of man and the soil: Reason for the asthma and atopy epidemic? J Allergy Clin Immunol 117, 334–344. https://doi.org/10.1016/j.jaci.2005.11.013

Vorholt, J.A., 2012. Microbial life in the phyllosphere. Nat Rev Microbiol 10, 828–840. https://doi.org/10.1038/nrmicro2910

Whipps, J.M., Hand, P., Pink, D., Bending, G.D., 2008. Phyllosphere microbiology with special reference to diversity and plant genotype. J Appl Microbiol 105, 1744–1755. https://doi.org/10.1111/j.1365-2672.2008.03906.x

Womack, A.M., Bohannan, B.J.M., Green, J.L., 2010. Biodiversity and biogeography of the atmosphere. Philos Trans R Soc Lond B Biol Sci 365, 3645–3653. https://doi.org/10.1098/rstb.2010.0283

**Chapter III**

Andersson, A.M., Weiss, N., Rainey, F., Salkinoja-Salonen, M.S., 1999. Dust-borne bacteria in animal sheds, schools and children's day care centres. J Appl Microbiol 86, 622–634.

Bowers, R.M., Clements, N., Emerson, J.B., Wiedinmyer, C., Hannigan, M.P., Fierer, N., 2013. Seasonal variability in bacterial and fungal diversity of the near-surface atmosphere. Environ Sci Technol 47, 12097–12106. https://doi.org/10.1021/es402970s

Bowers, R.M., McLetchie, S., Knight, R., Fierer, N., 2010. Spatial variability in airborne bacterial communities across land-use types and their relationship to the bacterial communities of potential source environments. The ISME J 5, 601–612. https://doi.org/10.1038/ismej.2010.167

Bowers, R.M., Sullivan, A.P., Costello, E.K., Collett, J.L., Knight, R., Fierer, N., 2011. Sources of bacteria in outdoor air across cities in the Midwestern United States. Appl Environ Microbiol 77, 6350–6356. https://doi.org/10.1128/AEM.05498-11

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Meth 13, 581–583. https://doi.org/10.1038/nmeth.3869

Ege, M.J., Mayer, M., Normand, A.-C., Genuneit, J., Cookson, W.O., Braun-Fahrländer, C., Heederik, D., Piarroux, R., von Mutius, E., 2011. Exposure to environmental microorganisms and childhood asthma. N Engl J Med 364, 701–709.

Erb-Downward, J.R., Sadighi Akha, A.A., Wang, J., Shen, N., He, B., Martinez, F.J., Gyetko, M.R., Curtis, J.L., Huffnagle, G.B., 2012. Use of direct gradient analysis to uncover biological hypotheses in 16S survey data and beyond. Sci Rep 2. https://doi.org/10.1038/srep00774

Fierer, N., Liu, Z., Rodriguez-Hernandez, M., Knight, R., Henn, M., Hernandez, M.T., 2008. Short-term temporal variability in airborne bacterial and fungal populations. Appl Environ Microbiol 74, 200–207. https://doi.org/10.1128/AEM.01467-07

Gandolfi, I., Bertolini, V., Ambrosini, R., Bestetti, G., Franzetti, A., 2013. Unravelling the bacterial diversity in the atmosphere. Appl Microbiol Biotechnol 97, 4727–4736. https://doi.org/10.1007/s00253-013-4901-2

Gandolfi, I., Bertolini, V., Bestetti, G., Ambrosini, R., Innocente, E., Rampazzo, G., Papacchini, M., Franzetti, A., 2015. Spatio-temporal variability of airborne bacterial communities and their correlation with particulate matter chemical composition across two urban areas. Appl Microbiol Biotechnol 99, 4867–4877. https://doi.org/10.1007/s00253-014-6348-5

Genitsaris, S., Stefanidou, N., Katsiapi, M., Kormas, K.A., Sommer, U., Moustaka-Gouni, M., 2017. Variability of airborne bacteria in an urban Mediterranean area (Thessaloniki, Greece). Atmos Environ 157, 101–110. https://doi.org/10.1016/j.atmosenv.2017.03.018

Gihring, T.M., Green, S.J., Schadt, C.W., 2012. Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. Environ Microbiol 14, 285–290. https://doi.org/10.1111/j.1462-2920.2011.02550.x

Glassing, A., Dowd, S.E., Galandiuk, S., Davis, B., Chiodini, R.J., 2016. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. Gut Pathog 8. https://doi.org/10.1186/s13099-016-0103-7

Haahtela, T., Holgate, S., Pawankar, R., Akdis, C.A., Benjaponpitak, S., Caraballo, L., Demain, J., Portnoy, J., von Hertzen, L., 2013. The biodiversity hypothesis and allergic disease: World Allergy Organization position statement. World Allergy Organ J 6, 3. https://doi.org/10.1186/1939-4551-6-3

Hanski, I., von Hertzen, L., Fyhrquist, N., Koskinen, K., Torppa, K., Laatikainen, T., Karisola, P., Auvinen, P., Paulin, L., Makela, M.J., Vartiainen, E., Kosunen, T.U., Alenius, H., Haahtela, T., 2012. Environmental biodiversity, human microbiota, and allergy are interrelated. PNAS 109, 8334–8339. https://doi.org/10.1073/pnas.1205624109

Huse, S., Welch, D.B.M., Sogin, M.L., 2012. Sequencing errors, diversity estimates, and the rare biosphere, in: The science and applications of microbial genomics. National Academies Press, Washington, D.C.

Innocente, E., Squizzato, S., Visin, F., Facca, C., Rampazzo, G., Bertolini, V., Gandolfi, I., Franzetti, A., Ambrosini, R., Bestetti, G., 2017. Influence of seasonality, air mass origin and particulate matter chemical composition on airborne bacterial community structure in the Po Valley, Italy. Sci Total Environ 593–594, 677–687. https://doi.org/10.1016/j.scitotenv.2017.03.199

Kraemer, J.G., Ramette, A., Aebi, S., Oppliger, A., Hilty, M., 2018. Influence of pig farming on

the human's nasal microbiota: The key role of the airborne microbial communities. Appl Environ Microbiol https://doi.org/10.1128/AEM.02470-17

Kube, M., Migdoll, A.M., Gehring, I., Heitmann, K., Mayer, Y., Kuhl, H., Knaust, F., Geider, K., Reinhardt, R., 2010. Genome comparison of the epiphytic bacteria *Erwinia billingiae* and *E. tasmaniensis* with the pear pathogen *E. pyrifoliae*. BMC Genomics 11, 393. https://doi.org/10.1186/1471-2164-11-393

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

Lee, J.Y., Park, E.H., Lee, S., Ko, G., Honda, Y., Hashizume, M., Deng, F., Yi, S., Kim, H., 2017. Airborne bacterial communities in three east Asian cities of China, South Korea, and Japan. Sci Rep 7, 5545. https://doi.org/10.1038/s41598-017-05862-4

Leppänen, H.K., Täubel, M., Jayaprakash, B., Vepsäläinen, A., Pasanen, P., Hyvärinen, A., 2017. Quantitative assessment of microbes from samples of indoor air and dust. J Expo Sci Environ Epidemiol. https://doi.org/10.1038/jes.2017.24

Lighthart, B., 1984. Microbial aerosols: Estimated contribution of combine harvesting to an airshed. Appl Environ Microbiol 47, 430–432.

Lindemann, J., Upper, C.D., 1985. Aerial dispersal of epiphytic bacteria over bean plants. Appl Environ Microbiol 50, 1229–1232.

Lymperopoulou, D.S., Adams, R.I., Lindow, S.E., 2016. Contribution of vegetation to the microbial composition of nearby outdoor air. Appl Environ Microbiol 82, 3822–3833. https://doi.org/10.1128/AEM.00610-16

Mahapatra N.R., Banerjee P.C., 2008. Extreme tolerance to cadmium and high resistance to copper, nickel and zinc in different Acidiphilium strains. Lett Appl Microbiol 23, 393–397. https://doi.org/10.1111/j.1472-765X.1996.tb01344.x

Martinez Arbizu, P., 2017. pairwiseAdonis: Pairwise multilevel comparison using adonis. R package version 0.0.1.

McGaughey, R.J., 2016. FUSION/LDV: Software for LIDAR data analysis and visualization.

McMurdie, P.J., Holmes, S., 2013. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLOS ONE 8, e61217. https://doi.org/10.1371/journal.pone.0061217

Meadow, J.F., Altrichter, A.E., Kembel, S.W., Kline, J., Mhuireach, G., Moriyama, M., Northcutt, D., O'Connor, T.K., Womack, A.M., Brown, G.Z., Green, J.L.., Bohannan, B.J.M., 2013. Indoor airborne bacterial communities are influenced by ventilation, occupancy, and outdoor air source. Indoor Air 24, 41–48. https://doi.org/10.1111/ina.12047

Mhuireach, G., Johnson, B.R., Altrichter, A.E., Ladau, J., Meadow, J.F., Pollard, K.S., Green, J.L., 2016. Urban greenness influences airborne bacterial community composition. Sci Total Environ 571, 680–687. https://doi.org/10.1016/j.scitotenv.2016.07.037

Mills, J.G., Weinstein, P., Gellie, N.J.C., Weyrich, L.S., Lowe, A.J., Breed, M.F., 2017. Urban habitat restoration provides a human health benefit through microbiome rewilding: The Microbiome Rewilding Hypothesis. Restoration Ecol 25, 866–872. https://doi.org/10.1111/rec.12610

Parks, D.H., Beiko, R.G., 2013. Measures of phylogenetic differentiation provide robust and complementary insights into microbial communities. ISME J 7, 173–183. https://doi.org/10.1038/ismej.2012.88

Prescott, S.L., Larcombe, D.-L., Logan, A.C., West, C., Burks, W., Caraballo, L., Levin, M., Etten, E.V., Horwitz, P., Kozyrskyj, A., Campbell, D.E., 2017. The skin microbiome: Impact of modern environments on skin ecology, barrier integrity, and systemic immune programming. World Allergy Organ J 10, 29. https://doi.org/10.1186/s40413-017-0160-5

Prussin, A.J., Marr, L.C., 2015. Sources of airborne microorganisms in the built environment. Microbiome 3, 78. https://doi.org/10.1186/s40168-015-0144-z

Rook, G., Bäckhed, F., Levin, B.R., McFall-Ngai, M.J., McLean, A.R., 2017. Evolution, human-microbe interactions, and life history plasticity. Lancet 390, 521–530. https://doi.org/10.1016/S0140-6736(17)30566-4

Rook, G.A., 2013. Regulation of the immune system by biodiversity from the natural environment: An ecosystem service essential to health. PNAS USA. 110, 18360–18367. https://doi.org/10.1073/pnas.1313731110

Ruokolainen, L., von Hertzen, L., Fyhrquist, N., Laatikainen, T., Lehtomäki, J., Auvinen, P., Karvonen, A., Hyvärinen, A.M., Tillmann, V., Niemelä, O., Knip, M., Haahtela, T., Pekkanen, J., Hanski, I., 2015. Green areas around homes reduce atopic sensitization in children. Allergy 195-202. https://doi.org/10.1111/all.12545

Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P., Parkhill, J., Loman, N.J., Walker, A.W., 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol 12, 87. https://doi.org/10.1186/s12915-014-0087-z

Seifried, J.S., Wichels, A., Gerdts, G., 2015. Spatial distribution of marine airborne bacterial communities. Microbiologyopen 4, 475–490. https://doi.org/10.1002/mbo3.253

Smith, D.J., Jaffe, D.A., Birmele, M.N., Griffin, D.W., Schuerger, A.C., Hee, J., Roberts, M.S., 2012. Free tropospheric transport of microorganisms from Asia to North America. Microb Ecol 64, 973–985. https://doi.org/10.1007/s00248-012-0088-9

Sorci, G., Guivier, E., Lippens, C., Faivre, B., 2016. Microbes, parasites and immune diseases,

in: Evolutionary thinking in medicine, advances in the evolutionary analysis of human behaviour. Springer, Cham, pp. 211–223. https://doi.org/10.1007/978-3-319-29716-3_15

United Nations, 2012. World urbanization prospects: The 2011 revision. New York.

Wakao, N., Nagasawa, N., Matsuura, T., Matsukura, H., Matsumoto, T., Hiraishi, A., Sakurai, Y., Shiota, H., 1994. *Acidiphilium multivorum* sp. nov., an acidophilic chemoorganotrophic bacterium from pyritic acid mine drainage. J Gen Appl Microbiol 40, 143–159. https://doi.org/10.2323/jgam.40.143

Wolda, H., 1981. Similarity indices, sample size and diversity. Oecologia 50, 296–302. https://doi.org/10.1007/BF00344966

Wuyts, K., Verheyen, K., De Schrijver, A., Cornelis, W.M., Gabriels, D., 2008. The impact of forest edge structure on longitudinal patterns of deposition, wind speed, and turbulence. Atmos Environ 42, 8651–8660. https://doi.org/10.1016/j.atmosenv.2008.08.010

**Chapter IV**

Adams, W.C., 1993. Measurement of breathing rate and volume in routinely performed daily activities.

Afifi, A.A., Breslow, L., 1994. The maturing paradigm of public health. Annu Rev Public Health 15, 223–235. https://doi.org/10.1146/annurev.pu.15.050194.001255

Alcock, I., White, M., Cherrie, M., Wheeler, B., Taylor, J., McInnes, R., Otte Im Kampe, E., Vardoulakis, S., Sarran, C., Soyiri, I., Fleming, L., 2017. Land cover and air pollution are associated with asthma hospitalisations: A cross-sectional study. Environ Int 109, 29–41. https://doi.org/10.1016/j.envint.2017.08.009

Andersson, A.M., Weiss, N., Rainey, F., Salkinoja-Salonen, M.S., 1999. Dust-borne bacteria in animal sheds, schools and children's day care centres. J Appl Microbiol 86, 622–634.

Auld, R.R., Myre, M., Mykytczuk, N.C.S., Leduc, L.G., Merritt, T.J.S., 2013. Characterization of the microbial acid mine drainage microbial community using culturing and direct sequencing techniques. J Microbiol Methods 93, 108–115. https://doi.org/10.1016/j.mimet.2013.01.023

Belkaid, Y., Hand, T.W., 2014. Role of the microbiota in immunity and inflammation. Cell 157, 121–141. https://doi.org/10.1016/j.cell.2014.03.011

Bell, M.L., Ebisu, K., 2012. Environmental inequality in exposures to airborne particulate matter components in the United States. Environ Health Perspect. https://doi.org/10.1289/ehp.1205201

Bharwani, A., Mian, M.F., Foster, J.A., Surette, M.G., Bienenstock, J., Forsythe, P., 2016.

Structural & functional consequences of chronic psychosocial stress on the microbiome & host. Psychoneuroendocrinology 63, 217–227. https://doi.org/10.1016/j.psyneuen.2015.10.001

Boston (Mass ) Department of Parks, 1881. Sixth Annual Report of the Board of Commisioners of the Department of Parks, for the City of Boston, for the Year 1880. Rockwell and Churchill, City Printers, Boston, MA.

Bowers, R.M., Clements, N., Emerson, J.B., Wiedinmyer, C., Hannigan, M.P., Fierer, N., 2013. Seasonal variability in bacterial and fungal diversity of the near-surface atmosphere. Environ Sci Technol 47, 12097–12106. https://doi.org/10.1021/es402970s

Bowers, R.M., McLetchie, S., Knight, R., Fierer, N., 2010. Spatial variability in airborne bacterial communities across land-use types and their relationship to the bacterial communities of potential source environments. The ISME J 5, 601–612. https://doi.org/10.1038/ismej.2010.167

Bowers, R.M., Sullivan, A.P., Costello, E.K., Collett, J.L., Knight, R., Fierer, N., 2011. Sources of bacteria in outdoor air across cities in the Midwestern United States. Appl Environ Microbiol 77, 6350–6356. https://doi.org/10.1128/AEM.05498-11

Britter, R.E., Hanna, S.R., 2003. Flow and dispersion in urban areas. Annu Rev Fluid Mech 35, 469–496. https://doi.org/10.1146/annurev.fluid.35.101101.161147

Brooks, B., Firek, B.A., Miller, C.S., Sharon, I., Thomas, B.C., Baker, R., Morowitz, M.J., Banfield, J.F., 2014. Microbes in the neonatal intensive care unit resemble those found in the gut of premature infants. Microbiome 2, 1. https://doi.org/10.1186/2049-2618-2-1

Buccolieri, R., Sandberg, M., Di Sabatino, S., 2010. City breathability and its link to pollutant concentration distribution within urban-like geometries. Atmos Environ 44, 1894–1903. https://doi.org/10.1016/j.atmosenv.2010.02.022

Burrows, S.M., Elbert, W., Lawrence, M.G., Pöschl, U., 2009. Bacteria in the global atmosphere–art 1: Review and synthesis of literature data for different ecosystems. Atmos Chem Phys 9, 9263–9280.

Calderón-Garcidueñas, L., Solt, A.C., Henríquez-Roldán, C., Torres-Jardón, R., Nuse, B., Herritt, L., Villarreal-Calderón, R., Osnaya, N., Stone, I., García, R., Brooks, D.M., González-Maciel, A., Reynoso-Robles, R., Delgado-Chávez, R., Reed, W., 2008. Long-term air pollution exposure is associated with neuroinflammation, an altered innate immune response, disruption of the blood-brain barrier, ultrafine particulate deposition, and accumulation of amyloid β-42 and α-synuclein in children and young adults. Toxicol Pathol 36, 289–310. https://doi.org/10.1177/0192623307313011

Camarinha-Silva, A., Jáuregui, R., Chaves-Moreno, D., Oxley, A.P.A., Schaumburg, F., Becker, K., Wos-Oxley, M.L., Pieper, D.H., 2014. Comparing the anterior nare bacterial community of two discrete human populations using Illumina amplicon sequencing. Environ Microbiol

16, 2939–2952. https://doi.org/10.1111/1462-2920.12362

Casey, J.A., James, P., Cushing, L., Jesdale, B.M., Morello-Frosch, R., 2017. Race, ethnicity, income concentration and 10-year change in urban greenness in the United States. Int J Environ Res Public Health 14. https://doi.org/10.3390/ijerph14121546

Chen, B., Nie, Z., Chen, Z., Xu, B., 2017. Quantitative estimation of 21st-century urban greenspace changes in Chinese populous cities. Sci Total Environ 609, 956–965. https://doi.org/10.1016/j.scitotenv.2017.07.238

Cook, W.M., Casagrande, D.G., Hope, D., Groffman, P.M., Collins, S.L., 2004. Learning to roll with the punches: Adaptive experimentation in human-dominated systems. Front Ecol Environ 2, 467–474. https://doi.org/10.1890/1540-9295(2004)002[0467:LTRWTP]2.0.CO;2

Corburn, J., 2004. Confronting the challenges in reconnecting urban planning and public health. Am J Public Health 94, 541.

Cox, D.T.C., Shanahan, D.F., Hudson, H.L., Fuller, R.A., Anderson, K., Hancock, S., Gaston, K.J., 2017. Doses of nearby nature simultaneously associated with multiple health benefits. Int J Environ Res Public Health 14. https://doi.org/10.3390/ijerph14020172

Cryan, J.F., Dinan, T.G., 2015. More than a gut feeling: The microbiota regulates neurodevelopment and behavior. Neuropsychopharmacology 40, 241–242. https://doi.org/10.1038/npp.2014.224

Dadvand, P., Bartoll, X., Basagaña, X., Dalmau-Bueno, A., Martinez, D., Ambros, A., Cirach, M., Triguero-Mas, M., Gascon, M., Borrell, C., Nieuwenhuijsen, M.J., 2016. Green spaces and general health: Roles of mental health status, social support, and physical activity. Environ Int 91, 161–167. https://doi.org/10.1016/j.envint.2016.02.029

Dadvand, P., de Nazelle, A., Figueras, F., Basagaña, X., Su, J., Amoly, E., Jerrett, M., Vrijheid, M., Sunyer, J., Nieuwenhuijsen, M.J., 2012. Green space, health inequality and pregnancy. Environ Int 40, 110–115. https://doi.org/10.1016/j.envint.2011.07.004

Dadvand, P., Nieuwenhuijsen, M.J., Esnaola, M., Forns, J., Basagaña, X., Alvarez-Pedrerol, M., Rivas, I., López-Vicente, M., Pascual, M.D.C., Su, J., Jerrett, M., Querol, X., Sunyer, J., 2015. Green spaces and cognitive development in primary schoolchildren. PNAS 112, 7937–7942. https://doi.org/10.1073/pnas.1503402112

Dadvand, P., Villanueva, C.M., Font-Ribera, L., Martinez, D., Basagaña, X., Belmonte, J., Vrijheid, M., Gražulevičienė, R., Kogevinas, M., Nieuwenhuijsen, M.J., 2014. Risks and benefits of green spaces for children: a cross-sectional study of associations with sedentary behavior, obesity, asthma, and allergy. Environ Health Perspect 122, 1329–1335. https://doi.org/10.1289/ehp.1308038

Dai, D., 2011. Racial/ethnic and socioeconomic disparities in urban green space accessibility:

Where to intervene? Landsc Urban Plan 102, 234–244. https://doi.org/10.1016/j.landurbplan.2011.05.002

Dallimer, M., Tang, Z., Bibby, P.R., Brindley, P., Gaston, K.J., Davies, Z.G., 2011. Temporal changes in greenspace in a highly urbanized region. Biol Lett 7, 763–766. https://doi.org/10.1098/rsbl.2011.0025

Dalton, A.M., Jones, A.P., Sharp, S.J., Cooper, A.J.M., Griffin, S., Wareham, N.J., 2016. Residential neighbourhood greenspace is associated with reduced risk of incident diabetes in older people: A prospective cohort study. BMC Public Health 16, 1171. https://doi.org/10.1186/s12889-016-3833-z

de Vries, S., van Dillen, S.M.E., Groenewegen, P.P., Spreeuwenberg, P., 2013. Streetscape greenery and health: Stress, social cohesion and physical activity as mediators. Soc Sci Med 94, 26–33. https://doi.org/10.1016/j.socscimed.2013.06.030

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

Dendup, T., Feng, X., Clingan, S., Astell-Burt, T., 2018. Environmental risk factors for developing type 2 diabetes mellitus: A systematic review. Int J Environ Res Public Health 15. https://doi.org/10.3390/ijerph15010078

Dewey, J., 1958. Experience and nature. Courier Corporation.

Donovan, G.H., Butry, D.T., Michael, Y.L., Prestemon, J.P., Liebhold, A.M., Gatziolis, D., Mao, M.Y., 2013. The relationship between trees and human health. Am J Prev Med 44, 139–145. https://doi.org/10.1016/j.amepre.2012.09.066

Donovan, G.H., Michael, Y.L., Butry, D.T., Sullivan, A.D., Chase, J.M., 2011. Urban trees and the risk of poor birth outcomes. Health Place 17, 390–393. https://doi.org/10.1016/j.healthplace.2010.11.004

Dooris, M., 2009. Holistic and sustainable health improvement: The contribution of the settings-based approach to health promotion. Perspect Public Health 129, 29–36. https://doi.org/10.1177/1757913908098881

Dubos, R., 1968. So human an animal: How we are shaped by surroundings and events. Transaction Publishers.

Dubos, R., 1966. Hippocrates in modern dress. Proc Inst Med Chic 9, 275–288.

Ege, M.J., Mayer, M., Normand, A.-C., Genuneit, J., Cookson, W.O., Braun-Fahrländer, C., Heederik, D., Piarroux, R., von Mutius, E., 2011. Exposure to environmental microorganisms and childhood asthma. N Engl J Med 364, 701–709.

Ege, M.J., Mayer, M., Schwaiger, K., Mattes, J., Pershagen, G., van Hage, M., Scheynius, A., Bauer, J., von Mutius, E., 2012. Environmental bacteria and childhood asthma. Allergy 1565-1571. https://doi.org/10.1111/all.12028

Eisenman, T.S., 2013. Frederick Law Olmsted, green infrastructure, and the evolving city. J Plan Hist 12, 287–311. https://doi.org/10.1177/1538513212474227

ESRI 2011. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.

Eyerich, S., Eyerich, K., Traidl-Hoffmann, C., Biedermann, T., 2018. Cutaneous barriers and skin immunity: Differentiating a connected network. Trends Immunol, Special Tissues: Local Immune Responses 39, 315–327. https://doi.org/10.1016/j.it.2018.02.004

Fahimipour, A.K., Vandegrift, R., Bateman, A., Bohannan, B.J.M., in prep. The transient dynamics of microbial colonization on human skin. Written for mSphere.

Fairchild, A.L., Rosner, D., Colgrove, J., Bayer, R., Fried, L.P., 2010. The exodus of public health: What history can tell us about the future. Am J Public Health 100, 54–63. https://doi.org/10.2105/AJPH.2009.163956

Faith, J.J., Guruge, J.L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A.L., Clemente, J.C., Knight, R., Heath, A.C., Leibel, R.L., Rosenbaum, M., Gordon, J.I., 2013. The long-term stability of the human gut microbiota. Science 341, 1237439. https://doi.org/10.1126/science.1237439

Felson, A.J., Pickett, S.T., 2005. Designed experiments: New approaches to studying urban ecosystems. Front Ecol Environ 3, 549–556. https://doi.org/10.1890/1540-9295(2005)003[0549:DENATS]2.0.CO;2

Feng, X., Astell-Burt, T., 2017. Do greener areas promote more equitable child health? Health Place 46, 267–273. https://doi.org/10.1016/j.healthplace.2017.05.006

Fessler, D.M.T., Abrams, E.T., 2004. Infant mouthing behavior: The immunocalibration hypothesis. Med Hypotheses 63, 925–932. https://doi.org/10.1016/j.mehy.2004.08.004

Fierer, N., 2017. Embracing the unknown: Disentangling the complexities of the soil microbiome. Nat Rev Microbiol 15, 579–590. https://doi.org/10.1038/nrmicro.2017.87

Fitzpatrick, K., LaGory, M., 2000. Unhealthy places: The ecology of risk in the urban landscape. Routledge, New York, NY.

Fong, K.C., Hart, J.E., James, P., 2018. A review of epidemiologic studies on greenness and health: Updated literature through 2017. Curr Environ Health Rep 5, 77–87. https://doi.org/10.1007/s40572-018-0179-y

Frank, D.N., Feazel, L.M., Bessesen, M.T., Price, C.S., Janoff, E.N., Pace, N.R., 2010. The human

nasal microbiota and *Staphylococcus aureus* carriage. PLOS ONE 5, e10598. https://doi.org/10.1371/journal.pone.0010598

Franzetti, A., Gandolfi, I., Gaspari, E., Ambrosini, R., Bestetti, G., 2010. Seasonal variability of bacteria in fine and coarse urban air particulate matter. Appl Microbiol Biotechnol 90, 745–753. https://doi.org/10.1007/s00253-010-3048-7

Fröhlich-Nowoisky, J., Kampf, C.J., Weber, B., Huffman, J.A., Pöhlker, C., Andreae, M.O., Lang-Yona, N., Burrows, S.M., Gunthe, S.S., Elbert, W., Su, H., Hoor, P., Thines, E., Hoffmann, T., Després, V.R., Pöschl, U., 2016. Bioaerosols in the Earth system: Climate, health, and ecosystem interactions. Atmos Research 182, 346–376. https://doi.org/10.1016/j.atmosres.2016.07.018

Frumkin, H., 2003. Healthy places: Exploring the evidence. Am J Public Health 93, 1451.

Frumkin, H., Bratman, G.N., Breslow, S.J., Cochran, B., Kahn, P.H., Lawler, J.J., Levin, P.S., Tandon, P.S., Varanasi, U., Wolf, K.L., Wood, S.A., 2017. Nature contact and human health: A research agenda. Environ Health Perspect 125, 75001. https://doi.org/10.1289/EHP1663

Fyhrquist, N., Ruokolainen, L., Suomalainen, A., Lehtimäki, S., Veckman, V., Vendelin, J., Karisola, P., Lehto, M., Savinko, T., Jarva, H., Kosunen, T.U., Corander, J., Auvinen, P., Paulin, L., von Hertzen, L., Laatikainen, T., Mäkelä, M., Haahtela, T., Greco, D., Hanski, I., Alenius, H., 2014. *Acinetobacter* species in the skin microbiota protect against allergic sensitization and inflammation. J Allergy Clin Immunol 134, 1301–1309.e11. https://doi.org/10.1016/j.jaci.2014.07.059

Gan, M., Deng, J., Zheng, X., Hong, Y., Wang, K., 2014. Monitoring urban greenness dynamics using multiple endmember spectral mixture analysis. PLOS ONE 9, e112202. https://doi.org/10.1371/journal.pone.0112202

García, R., White, A., 2006. Healthy parks, schools, and communities: Mapping green access and equity for the Los Angeles region. The City Project Policy Report. Accessed online in May 29, 2007.

Gilbert, J.A., Neufeld, J.D., 2014. Life in a world without microbes. PLOS Biology 12, e1002020. https://doi.org/10.1371/journal.pbio.1002020

Gilbert, S.F., Tauber, A.I., 2016. Rethinking individuality: The dialectics of the holobiont. Biol Philos 31, 839–853. https://doi.org/10.1007/s10539-016-9541-3

Grahn, P., Stigsdotter, U.A., 2003. Landscape planning and stress. Urban For Urban Green 2, 1–18.

Grimmond, C.S.B., Oke, T.R., 1999. Aerodynamic properties of urban areas derived from analysis of surface form. J Appl Meteor 38, 1262–1292. https://doi.org/10.1175/1520-0450(1999)038<1262:APOUAD>2.0.CO;2

Groenewegen, P.P., van den Berg, A.E., Maas, J., Verheij, R.A., de Vries, S., 2012. Is a green residential environment better for health? If so, why? Ann Assoc Am Geogr 102, 996–1003. https://doi.org/10.1080/00045608.2012.674899

Gur, T.L., Worly, B.L., Bailey, M.T., 2015. Stress and the commensal microbiota: Importance in parturition and infant neurodevelopment. Front Psychiatry 6. https://doi.org/10.3389/fpsyt.2015.00005

Haines, M.R., 2001. The urban mortality transition in the United States, 1800-1940, Summary. Annales de démographie historique no 101, 33–64. https://doi.org/10.3917/adh.101.0033

Hanski, I., von Hertzen, L., Fyhrquist, N., Koskinen, K., Torppa, K., Laatikainen, T., Karisola, P., Auvinen, P., Paulin, L., Makela, M.J., Vartiainen, E., Kosunen, T.U., Alenius, H., Haahtela, T., 2012. Environmental biodiversity, human microbiota, and allergy are interrelated. PNAS 109, 8334–8339. https://doi.org/10.1073/pnas.1205624109

Harlan, S.L., Brazel, A.J., Prashad, L., Stefanov, W.L., Larsen, L., 2006. Neighborhood microclimates and vulnerability to heat stress. Soc Sci Med 63, 2847–2863. https://doi.org/10.1016/j.socscimed.2006.07.030

Harris, B., 2004. Public health, nutrition, and the decline of mortality: The McKeown thesis revisited. Soc Hist Med 17, 379–407. https://doi.org/10.1093/shm/17.3.379

Hartig, T., Mitchell, R., de Vries, S., Frumkin, H., 2014. Nature and health. Annu Rev Public Health 35, 207–228. https://doi.org/10.1146/annurev-publhealth-032013-182443

Hill, H.W., 1913. The new public health. Press of the Journal-Lancet.

Hippocrates, n.d. On airs, waters and places. The Internet Classics Archive.

Hoisington, A.J., Brenner, L.A., Kinney, K.A., Postolache, T.T., Lowry, C.A., 2015. The microbiome of the built environment and mental health. Microbiome 3, 60. https://doi.org/10.1186/s40168-015-0127-0

Hope, D., Gries, C., Zhu, W., Fagan, W.F., Redman, C.L., Grimm, N.B., Nelson, A.L., Martin, C., Kinzig, A., 2003. Socioeconomics drive urban plant diversity. PNAS 100, 8788–8792.

Hystad, P., Davies, H.W., Frank, L., Van Loon, J., Gehring, U., Tamburic, L., Brauer, M., 2014. Residential greenness and birth outcomes: Evaluating the influence of spatially correlated built-environment factors. Environ Health Perspect 122, 1095–1102. https://doi.org/10.1289/ehp.1308049

Innocente, E., Squizzato, S., Visin, F., Facca, C., Rampazzo, G., Bertolini, V., Gandolfi, I., Franzetti, A., Ambrosini, R., Bestetti, G., 2017. Influence of seasonality, air mass origin and particulate matter chemical composition on airborne bacterial community structure in the Po Valley, Italy. Sci Total Environ 593–594, 677–687. https://doi.org/10.1016/j.scitotenv.2017.03.199

Izhaki, I., Fridman, S., Gerchman, Y., Halpern, M., 2013. Variability of bacterial community composition on leaves between and within plant species. Curr Microbiol 66, 227–235. https://doi.org/10.1007/s00284-012-0261-x

James, P., Hart, J.E., Banay, R.F., Laden, F., 2016. Exposure to greenness and mortality in a nationwide prospective cohort study of women. Environ Health Perspect 124, 1344–1352. https://doi.org/10.1289/ehp.1510363

Jatzlauk, G., Bartel, S., Heine, H., Schloter, M., Krauss-Etschmann, S., 2017. Influences of environmental bacteria and their metabolites on allergies, asthma, and host microbiota. Allergy 72, 1859–1867. https://doi.org/10.1111/all.13220

Jennings, V., Johnson Gaither, C., Gragg, R.S., 2012. Promoting environmental justice through urban green space access: A synopsis. Environ Justice 5, 1–7. https://doi.org/10.1089/env.2011.0007

Kabisch, N., Qureshi, S., Haase, D., 2015. Human–environment interactions in urban green spaces—A systematic review of contemporary issues and prospects for future research. Environ Impact Assess 50, 25–34. https://doi.org/10.1016/j.eiar.2014.08.007

Kearns, G., 1988. The urban penalty and the population history of England, in: Brändström, A., Tedebrand, L.-G. (Eds.), Society, health and population during the demographic transition. Almqvist and Wiksell International, Stockholm, pp. 213–236.

Kembel, S.W., O'Connor, T.K., Arnold, H.K., Hubbell, S.P., Wright, S.J., Green, J.L., 2014. Relationships between phyllosphere bacterial communities and plant functional traits in a neotropical forest. PNAS 111, 13715–13720.

Kinkel, L.L., 1997. Microbial population dynamics on leaves. Annu Rev Phytopathol 35, 327–347. https://doi.org/10.1146/annurev.phyto.35.1.327

Kinzig, A.P., Warren, P., Martin, C., Hope, D., Katti, M., 2005. The effects of human socioeconomic status and cultural characteristics on urban patterns of biodiversity. Ecol Soc 10, 23.

Kish, L., Hotte, N., Kaplan, G.G., Vincent, R., Tso, R., Gänzle, M., Rioux, K.P., Thiesen, A., Barkema, H.W., Wine, E., Madsen, K.L., 2013. Environmental particulate matter induces murine intestinal inflammatory responses and alters the gut microbiome. PLOS ONE 8, e62220. https://doi.org/10.1371/journal.pone.0062220

Klepeis, N.E., Nelson, W.C., Ott, W.R., Robinson, J.P., Tsang, A.M., Switzer, P., Behar, J.V., Hern, S.C., Engelmann, W.H., 2001. The National Human Activity Pattern Survey (NHAPS): A resource for assessing exposure to environmental pollutants. J Expo Sci Environ Epidemiol 11, 231–252.

Knief, C., Ramette, A., Frances, L., Alonso-Blanco, C., Vorholt, J.A., 2010. Site and plant species are important determinants of the *Methylobacterium* community composition in the

plant phyllosphere. The ISME J 4, 719–728. https://doi.org/10.1038/ismej.2010.9

Kowalski, W.J., Bahnfleth, W., 1998. Airborne respiratory diseases and mechanical systems for control of microbes. HPAC.

Kraemer, J.G., Ramette, A., Aebi, S., Oppliger, A., Hilty, M., 2018. Influence of pig farming on the human's nasal microbiota: The key role of the airborne microbial communities. Appl Environ Microbiol https://doi.org/10.1128/AEM.02470-17

Kuo, F.E., 2001. Coping with poverty: Impacts of environment and attention in the inner city. Environ Behav 33, 5–34. https://doi.org/10.1177/00139160121972846

Kuo, M., 2015. How might contact with nature promote human health? Promising mechanisms and a possible central pathway. Front Psychol 6. https://doi.org/10.3389/fpsyg.2015.01093

Laforest-Lapointe, I., Messier, C., Kembel, S.W., 2017. Tree leaf bacterial community structure and diversity differ along a gradient of urban intensity. mSystems 2, e00087-17. https://doi.org/10.1128/mSystems.00087-17

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

Landry, S.M., Chakraborty, J., 2009. Street trees and equity: Evaluating the spatial distribution of an urban amenity. Environ Plan A 41, 2651–2670. https://doi.org/10.1068/a41236

Lee, A.C.K., Maheswaran, R., 2010. The health benefits of urban green spaces: A review of the evidence. J Public Health 33, 212–222. https://doi.org/10.1093/pubmed/fdq068

Lehtimäki, J., Karkman, A., Laatikainen, T., Paalanen, L., Hertzen, L. von, Haahtela, T., Hanski, I., Ruokolainen, L., 2017. Patterns in the skin microbiota differ in children and teenagers between rural and urban environments. Sci Rep 7, 45651. https://doi.org/10.1038/srep45651

Leung, M., Chan, A.H.S., 2006. Control and management of hospital indoor air quality. Med Sci Monit 12, SR17-23.

Li, Q., 2010. Effect of forest bathing trips on human immune function. Environ Health Prev Med 15, 9–17. https://doi.org/10.1007/s12199-008-0068-3

Liddicoat, C., Bi, P., Waycott, M., Glover, J., Lowe, A.J., Weinstein, P., 2018. Landscape biodiversity correlates with respiratory health in Australia. J Environ Manage 206, 113–122. https://doi.org/10.1016/j.jenvman.2017.10.007

Liddicoat, C., Waycott, M., Weinstein, P., 2016. Environmental change and human health: Can environmental proxies inform the biodiversity hypothesis for protective microbial–human contact? BioScience 66, 1023–1034. https://doi.org/10.1093/biosci/biw127

Lighthart, B., 1984. Microbial aerosols: Estimated contribution of combine harvesting to an airshed. Appl Environ Microbiol 47, 430–432.

Lighthart, B., Shaffer, B.T., Frisch, A.S., Paterno, D., 2009. Atmospheric culturable bacteria associated with meteorological conditions at a summer-time site in the mid-Willamette Valley, Oregon. Aerobiologia 25, 285–295. https://doi.org/10.1007/s10453-009-9133-7

Lindemann, J., Constantinidou, H.A., Barchet, W.R., Upper, C.D., 1982. Plants as sources of airborne bacteria, including ice nucleation-active bacteria. Appl Environ Microbiol 44, 1059–1063.

Lindemann, J., Upper, C.D., 1985. Aerial dispersal of epiphytic bacteria over bean plants. Appl Environ Microbiol 50, 1229–1232.

Lindow, S.E., Brandl, M.T., 2003. Microbiology of the phyllosphere. Appl Environ Microbiol 69, 1875–1883. https://doi.org/10.1128/AEM.69.4.1875-1883.2003

Linehan, J.L., Harrison, O.J., Han, S.-J., Byrd, A.L., Vujkovic-Cvijin, I., Villarino, A.V., Sen, S.K., Shaik, J., Smelkinson, M., Tamoutounour, S., Collins, N., Bouladoux, N., Dzutsev, A., Rosshart, S.P., Arbuckle, J.H., Wang, C.-R., Kristie, T.M., Rehermann, B., Trinchieri, G., Brenchley, J.M., O'Shea, J.J., Belkaid, Y., 2018. Non-classical immunity controls microbiota impact on skin immunity and tissue repair. Cell 172, 784–796.e18. https://doi.org/10.1016/j.cell.2017.12.033

Liu, D.H.F., Liptak, B.G., 1997. Environmental engineers' handbook, Second Edition. CRC Press.

Logan, A.C., Jacka, F.N., Prescott, S.L., 2016. Immune-microbiota interactions: Dysbiosis as a global health issue. Curr Allergy Asthma Rep 16, 13. https://doi.org/10.1007/s11882-015-0590-5

Lowry, C.A., Hollis, J.H., de Vries, A., Pan, B., Brunet, L.R., Hunt, J.R.F., Paton, J.F.R., van Kampen, E., Knight, D.M., Evans, A.K., Rook, G.A.W., Lightman, S.L., 2007. Identification of an immune-responsive mesolimbocortical serotonergic system: Potential role in regulation of emotional behavior. Neuroscience 146, 756–772. https://doi.org/10.1016/j.neuroscience.2007.01.067

Lymperopoulou, D.S., Adams, R.I., Lindow, S.E., 2016. Contribution of vegetation to the microbial composition of nearby outdoor air. Appl Environ Microbiol 82, 3822–3833. https://doi.org/10.1128/AEM.00610-16

Maas, J., 2006. Green space, urbanity, and health: How strong is the relation? J Epidemiol Community Health 60, 587–592. https://doi.org/10.1136/jech.2005.043125

Maas, J., Verheij, R.A., de Vries, S., Spreeuwenberg, P., Schellevis, F.G., Groenewegen, P.P., 2009. Morbidity is related to a green living environment. J Epidemiol Community Health 63, 967–973. https://doi.org/10.1136/jech.2008.079038

Macintyre, S., Ellaway, A., Cummins, S., 2002. Place effects on health: How can we conceptualise, operationalise and measure them? Soc Sci Med, Selected papers from the 9th International Symposium on Medical Geography 55, 125–139. https://doi.org/10.1016/S0277-9536(01)00214-3

MacKerron, G., Mourato, S., 2013. Happiness is greater in natural environments. Glob Environ Change 23, 992–1000. https://doi.org/10.1016/j.gloenvcha.2013.03.010

Mao, G.X., Lan, X.G., Cao, Y.B., Chen, Z.M., He, Z.H., Lv, Y.D., Wang, Y.Z., Hu, X.L., Wang, G.F., Yan, J., 2012. Effects of short-term forest bathing on human health in a broad-leaved evergreen forest in Zhejiang Province, China. Biomed Environ Sci 25, 317–324. https://doi.org/10.3967/0895-3988.2012.03.010

Markevych, I., Schoierer, J., Hartig, T., Chudnovsky, A., Hystad, P., Dzhambov, A.M., de Vries, S., Triguero-Mas, M., Brauer, M., Nieuwenhuijsen, M.J., Lupp, G., Richardson, E.A., Astell-Burt, T., Dimitrova, D., Feng, X., Sadeh, M., Standl, M., Heinrich, J., Fuertes, E., 2017. Exploring pathways linking greenspace to health: Theoretical and methodological guidance. Environ Res 158, 301–317. https://doi.org/10.1016/j.envres.2017.06.028

Martensen, R., 2009. Landscape designers, doctors, and the making of healthy urban spaces in 19th century America, in: Campbell, L.K., Wiesen, A. (Eds.), Restorative commons: Creating health and well-being through urban landscapes. Government Printing Office, pp. 26–37.

Matthews, D.M., Jenks, S.M., 2013. Ingestion of *Mycobacterium vaccae* decreases anxiety-related behavior and improves learning in mice. Behav Processes 96, 27–35. https://doi.org/10.1016/j.beproc.2013.02.007

McDade, T.W., 2005. Life history, maintenance, and the early origins of immune function. Am J Hum Biol 17, 81–94. https://doi.org/10.1002/ajhb.20095

McEachan, R.R.C., Prady, S.L., Smith, G., Fairley, L., Cabieses, B., Gidlow, C., Wright, J., Dadvand, P., van Gent, D., Nieuwenhuijsen, M.J., 2016. The association between green space and depressive symptoms in pregnant women: Moderating roles of socioeconomic status and physical activity. J Epidemiol Community Health 70, 253–259. https://doi.org/10.1136/jech-2015-205954

McMahan, E.A., Estes, D., 2015. The effect of contact with natural environments on positive and negative affect: A meta-analysis. J Posit Psychol 10, 507–519. https://doi.org/10.1080/17439760.2014.994224

Meisel, J.S., Sfyroera, G., Bartow-McKenney, C., Gimblet, C., Bugayev, J., Horwinski, J., Kim, B., Brestoff, J.R., Tyldsley, A.S., Zheng, Q., Hodkinson, B.P., Artis, D., Grice, E.A., 2018. Commensal microbiota modulate gene expression in the skin. Microbiome 6, 20. https://doi.org/10.1186/s40168-018-0404-9

Mhuireach, G., Johnson, B.R., Altrichter, A.E., Ladau, J., Meadow, J.F., Pollard, K.S., Green,

J.L., 2016. Urban greenness influences airborne bacterial community composition. Sci Total Environ 571, 680–687. https://doi.org/10.1016/j.scitotenv.2016.07.037

Mills, J.G., Weinstein, P., Gellie, N.J.C., Weyrich, L.S., Lowe, A.J., Breed, M.F., 2017. Urban habitat restoration provides a human health benefit through microbiome rewilding: The Microbiome Rewilding Hypothesis. Restoration Ecol 25, 866–872. https://doi.org/10.1111/rec.12610

Milne, G., 2015. The evolutionary determinants of health programme: Urban living in the 21st century from a human evolutionary perspective. Archaeology International 18. https://doi.org/10.5334/ai.1809

Milne, G., 2017. Uncivilized genes: How a greater understanding of the evolutionary determinants of health could improve our urban wellbeing. Arch Mol Med & Gen 1, 9-14. https://doi.org/10.4172/2153-0645-C1-015

Mitchell, R., Popham, F., 2007. Greenspace, urbanity and health: Relationships in England. J Epidemiol Community Health 61, 681–683. https://doi.org/10.1136/jech.2006.053553

Monteil, M.A., 2008. Saharan dust clouds and human health in the English-speaking Caribbean: What we know and don't know. Environ Geochem Health 30, 339–343. https://doi.org/10.1007/s10653-008-9162-0

Mulder, I.E., Schmidt, B., Stokes, C.R., Lewis, M., Bailey, M., Aminov, R.I., Prosser, J.I., Gill, B.P., Pluske, J.R., Mayer, C.-D., Musk, C.C., Kelly, D., 2009. Environmentally-acquired bacteria influence microbial diversity and natural innate immune responses at gut surfaces. BMC Biol 7, 79. https://doi.org/10.1186/1741-7007-7-79

Naik, S., Bouladoux, N., Wilhelm, C., Molloy, M.J., Salcedo, R., Kastenmuller, W., Deming, C., Quinones, M., Koo, L., Conlan, S., Spencer, S., Hall, J.A., Dzutsev, A., Kong, H., Campbell, D.J., Trinchieri, G., Segre, J.A., Belkaid, Y., 2012. Compartmentalized control of skin immunity by resident commensals. Science 337, 1115–1119. https://doi.org/10.1126/science.1225152

Nassauer, J.I., Opdam, P., 2008. Design in science: Extending the landscape ecology paradigm. Landscape Ecol 23, 633–644. https://doi.org/10.1007/s10980-008-9226-7

Nieuwenhuijsen, M.J., Khreis, H., Triguero-Mas, M., Gascon, M., Dadvand, P., 2017. Fifty shades of green: Pathway to healthy urban living. Epidemiology 28, 63–71. https://doi.org/10.1097/EDE.0000000000000549

Nowak, D.J., Greenfield, E.J., 2012. Tree and impervious cover change in US. Urban For Urban Green, 11(1), pp.21-30. https://doi.org/10.1016/j.ufug.2011.11.005

Oh, J., Byrd, A.L., Park, M., Kong, H.H., Segre, J.A., 2016. Temporal stability of the human skin microbiome. Cell 165, 854–866. https://doi.org/10.1016/j.cell.2016.04.008

Padgett, D.A., Glaser, R., 2003. How stress influences the immune response. Trends Immunol 24, 444–448.

Parajuli, A., Grönroos, M., Siter, N., Puhakka, R., Vari, H.K., Roslund, M.I., Jumpponen, A., Nurminen, N., Laitinen, O.H., Hyöty, H., Rajaniemi, J., Sinkkonen, A., 2018. Urbanization reduces transfer of diverse environmental microbiota indoors. Front Microbiol 9. https://doi.org/10.3389/fmicb.2018.00084

Park, B.-J., Tsunetsugu, Y., Kasetani, T., Hirano, H., Kagawa, T., Sato, M., Miyazaki, Y., 2007. Physiological effects of shinrin-yoku (taking in the atmosphere of the forest)—using salivary cortisol and cerebral activity as indicators. J Physiol Anthropol 26, 123–128.

Pedersen, B.K., Hoffman-Goetz, L., 2000. Exercise and the immune system: Regulation, integration, and adaptation. Physiol Rev 80, 1055–1081. https://doi.org/10.1152/physrev.2000.80.3.1055

Porter, D., 1999. Health, civilization and the state: A history of public health from ancient to modern times. Taylor & Francis, London, UK.

Prescott, S.L., Larcombe, D.-L., Logan, A.C., West, C., Burks, W., Caraballo, L., Levin, M., Etten, E.V., Horwitz, P., Kozyrskyj, A., Campbell, D.E., 2017. The skin microbiome: Impact of modern environments on skin ecology, barrier integrity, and systemic immune programming. World Allergy Organ J 10, 29. https://doi.org/10.1186/s40413-017-0160-5

Redford, A.J., Fierer, N., 2009. Bacterial succession on the leaf surface: A novel system for studying successional dynamics. Microb Ecol 58, 189–198. https://doi.org/10.1007/s00248-009-9495-y

Reher, D.S., 2001. In search of the "urban penalty": Exploring urban and rural mortality patterns in Spain during the demographic transition. International Journal of Population Geography 7, 105–127. https://doi.org/10.1002/ijpg.212

Richardson, E.A., Pearce, J., Mitchell, R., Kingham, S., 2013. Role of physical activity in the relationship between urban green space and health. Public Health 127, 318–324. https://doi.org/10.1016/j.puhe.2013.01.004

Roe, J.J., Thompson, C.W., Aspinall, P.A., Brewer, M.J., Duff, E.I., Miller, D., Mitchell, R., Clow, A., 2013. Green space and stress: Evidence from cortisol measures in deprived urban communities. Int J Environ Res Public Health 10, 4086–4103. https://doi.org/10.3390/ijerph10094086

Rook, G., Bäckhed, F., Levin, B.R., McFall-Ngai, M.J., McLean, A.R., 2017. Evolution, human-microbe interactions, and life history plasticity. Lancet 390, 521–530. https://doi.org/10.1016/S0140-6736(17)30566-4

Rook, G.A., 2013. Regulation of the immune system by biodiversity from the natural environment: An ecosystem service essential to health. PNAS USA. 110, 18360–18367.

https://doi.org/10.1073/pnas.1313731110

Rook, G.A.W., 2008. Review series on helminths, immune modulation and the hygiene hypothesis: The broader implications of the hygiene hypothesis. Immunology 126, 3–11. https://doi.org/10.1111/j.1365-2567.2008.03007.x

Ruokolainen, L., von Hertzen, L., Fyhrquist, N., Laatikainen, T., Lehtomäki, J., Auvinen, P., Karvonen, A., Hyvärinen, A.M., Tillmann, V., Niemelä, O., Knip, M., Haahtela, T., Pekkanen, J., Hanski, I., 2015. Green areas around homes reduce atopic sensitization in children. Allergy 195-202. https://doi.org/10.1111/all.12545

Seedorf, H., Griffin, N.W., Ridaura, V.K., Reyes, A., Cheng, J., Rey, F.E., Smith, M.I., Simon, G.M., Scheffrahn, R.H., Woebken, D., Spormann, A.M., Van Treuren, W., Ursell, L.K., Pirrung, M., Robbins-Pianka, A., Cantarel, B.L., Lombard, V., Henrissat, B., Knight, R., Gordon, J.I., 2014. Bacteria from diverse habitats colonize and compete in the mouse gut. Cell 159, 253–266. https://doi.org/10.1016/j.cell.2014.09.008

Seifried, J.S., Wichels, A., Gerdts, G., 2015. Spatial distribution of marine airborne bacterial communities. Microbiologyopen 4, 475–490. https://doi.org/10.1002/mbo3.253

Shaffer, B.T., Lighthart, B., 1997. Survey of culturable airborne bacteria at four diverse locations in Oregon: Urban, rural, forest, and coastal. Microb Ecol 34, 167–177.

Shanahan, D.F., Lin, B.B., Bush, R., Gaston, K.J., Dean, J.H., Barber, E., Fuller, R.A., 2015. Toward improved public health outcomes from urban nature. Am J Public Health 105, 470–477. https://doi.org/10.2105/AJPH.2014.302324

Smets, W., Moretti, S., Denys, S., Lebeer, S., 2016. Airborne bacteria in the atmosphere: Presence, purpose, and potential. Atmos Environ 139, 214–221. https://doi.org/10.1016/j.atmosenv.2016.05.038

Smith, D.J., Jaffe, D.A., Birmele, M.N., Griffin, D.W., Schuerger, A.C., Hee, J., Roberts, M.S., 2012. Free tropospheric transport of microorganisms from Asia to North America. Microb Ecol 64, 973–985. https://doi.org/10.1007/s00248-012-0088-9

Sorci, G., Guivier, E., Lippens, C., Faivre, B., 2016. Microbes, parasites and immune diseases, in: Evolutionary thinking in medicine, advances in the evolutionary analysis of human behaviour. Springer, Cham, pp. 211–223. https://doi.org/10.1007/978-3-319-29716-3_15

Spirn, A.W., 1988. The poetics of city and nature: Towards a new aesthetic for urban design. Landscape Jrnl 7, 108–126. https://doi.org/10.3368/lj.7.2.108

Sprockett, D., Fukami, T., Relman, D.A., 2018. Role of priority effects in the early-life assembly of the gut microbiota. Nat Rev Gastroenterol Hepatol 15, 197–205. https://doi.org/10.1038/nrgastro.2017.173

Stigsdotter, U.K., Ekholm, O., Schipperijn, J., Toftager, M., Kamper-Jorgensen, F., Randrup,

T.B., 2010. Health promoting outdoor environments - Associations between green space, and health, health-related quality of life and stress based on a Danish national representative survey. Scandinavian J Public Health 38, 411–417. https://doi.org/10.1177/1403494810367468

Strachan, D.P., 1989. Hay fever, hygiene, and household size. BMJ 299, 1259–1260.

Strohbach, M.W., Haase, D., Kabisch, N., 2009. Birds and the city: Urban biodiversity, land use, and socioeconomics. Ecol Soc 14, 31.

Szreter, S., 1988. The importance of social intervention in Britain's mortality decline c.1850–1914: A re-interpretation of the role of public health. Soc History Med 1, 1–38.

Tang, J.W., 2009. The effect of environmental parameters on the survival of airborne infectious agents. J R Soc Interface 6 Suppl. 6, S737-746. https://doi.org/10.1098/rsif.2009.0227.focus

The Coalition for a Livable Future, 2007. The regional equity atlas: Metropolitan Portland's geography of opportunity. Portland, Oregon.

Timonen, H., Wigder, N., Jaffe, D., 2013. Influence of background particulate matter (PM) on urban air quality in the Pacific Northwest. J Environ Manage 129, 333–340. https://doi.org/10.1016/j.jenvman.2013.07.023

Tong, Y., Lighthart, B., 1997. Solar radiation has a lethal effect on natural populations of culturable outdoor atmospheric bacteria. Atmos Environ 31, 897–900.

Triguero-Mas, M., Gidlow, C.J., Martínez, D., de Bont, J., Carrasco-Turigas, G., Martínez-Íñiguez, T., Hurst, G., Masterson, D., Donaire-Gonzalez, D., Seto, E., Jones, M.V., Nieuwenhuijsen, M.J., 2017. The effect of randomised exposure to different types of natural outdoor environments compared to exposure to an urban environment on people with indications of psychological distress in Catalonia. PLOS ONE 12, e0172200. https://doi.org/10.1371/journal.pone.0172200

Tzoulas, K., Korpela, K., Venn, S., Yli-Pelkonen, V., Kaźmierczak, A., Niemela, J., James, P., 2007. Promoting ecosystem and human health in urban areas using Green Infrastructure: A literature review. Landsc Urban Plan 81, 167–178. https://doi.org/10.1016/j.landurbplan.2007.02.001

Uchino, B.N., 2006. Social support and health: A review of physiological processes potentially underlying links to disease outcomes. J Behav Med 29, 377–387. https://doi.org/10.1007/s10865-006-9056-5

Ulmer, J.M., Wolf, K.L., Backman, D.R., Tretheway, R.L., Blain, C.J., O'Neil-Dunne, J.P., Frank, L.D., 2016. Multiple health benefits of urban tree canopy: The mounting evidence for a green prescription. Health Place 42, 54–62. https://doi.org/10.1016/j.healthplace.2016.08.011

United Nations, D. of E. and S.A., Population Division, 2017. World population prospects: The 2017 revision, key findings and advance tables., Working Paper No. ESA/P/WP/248.

United Nations, D. of E. and S.A., Population Division, 2014. World urbanization prospects: The 2014 revision, Highlights.

Vienneau, D., de Hoogh, K., Faeh, D., Kaufmann, M., Wunderli, J.M., Röösli, M., SNC Study Group, 2017. More than clean air and tranquility: Residential green is independently associated with decreasing mortality. Environ Int 108, 176–184. https://doi.org/10.1016/j.envint.2017.08.012

Villeneuve, P.J., Jerrett, M., G. Su, J., Burnett, R.T., Chen, H., Wheeler, A.J., Goldberg, M.S., 2012. A cohort study relating urban green space with mortality in Ontario, Canada. Environ Res 115, 51–58. https://doi.org/10.1016/j.envres.2012.03.003

Vitruvius, n.d. De Architectura.

von Hertzen, L., Haahtela, T., 2006. Disconnection of man and the soil: Reason for the asthma and atopy epidemic? J Allergy Clin Immunol 117, 334–344. https://doi.org/10.1016/j.jaci.2005.11.013

von Hertzen, L., Hanski, I., Haahtela, T., 2011. Natural immunity. EMBO Rep 12, 1089–1093.

von Mutius, E., 2016. The microbial environment and its influence on asthma prevention in early life. J Allergy Clin Immunol 137, 680–689. https://doi.org/10.1016/j.jaci.2015.12.1301

Vorholt, J.A., 2012. Microbial life in the phyllosphere. Nat Rev Microbiol 10, 828–840. https://doi.org/10.1038/nrmicro2910

Ward Thompson, C., 2011. Linking landscape and health: The recurring theme. Landsc Urban Plan 99, 187–195. https://doi.org/10.1016/j.landurbplan.2010.10.006

Ward Thompson, C., de Oliveira, S., 2016. Evidence on health benefits of urban green spaces, in: Egorov, A., Mudu, P., Braubach, M., Martuzzi, M. (Eds.), Urban green spaces and health: A review of evidence. World Health Organization Regional Office for Europe, Copenhagen, pp. 3–20.

Ward Thompson, C., Roe, J., Aspinall, P., Mitchell, R., Clow, A., Miller, D., 2012. More green space is linked to less stress in deprived communities: Evidence from salivary cortisol patterns. Landsc Urban Plan 105, 221–229. https://doi.org/10.1016/j.landurbplan.2011.12.015

Whipps, J.M., Hand, P., Pink, D., Bending, G.D., 2008. Phyllosphere microbiology with special reference to diversity and plant genotype. J Appl Microbiol 105, 1744–1755. https://doi.org/10.1111/j.1365-2672.2008.03906.x

White, M.P., Alcock, I., Wheeler, B.W., Depledge, M.H., 2013. Would you be happier living in a greener urban area? A fixed-effects analysis of panel data. Psychol Sci 24, 920–928. https://doi.org/10.1177/0956797612464659

Whitman, W.B., Coleman, D.C., Wiebe, W.J., 1998. Prokaryotes: The unseen majority. PNAS 95, 6578–6583.

Womack, A.M., Bohannan, B.J.M., Green, J.L., 2010. Biodiversity and biogeography of the atmosphere. Philos Trans R Soc Lond B Biol Sci 365, 3645–3653. https://doi.org/10.1098/rstb.2010.0283

Wood, L., Hooper, P., Foster, S., Bull, F., 2017. Public green spaces and positive mental health—Investigating the relationship between access, quantity and types of parks and mental wellbeing. Health Place 48, 63–71. https://doi.org/10.1016/j.healthplace.2017.09.002

Wu, J., Jackson, L., 2017. Inverse relationship between urban green space and childhood autism in California elementary school districts. Environ Int 107, 140–146. https://doi.org/10.1016/j.envint.2017.07.010

Wuyts, K., Verheyen, K., De Schrijver, A., Cornelis, W.M., Gabriels, D., 2008. The impact of forest edge structure on longitudinal patterns of deposition, wind speed, and turbulence. Atmos Environ 42, 8651–8660. https://doi.org/10.1016/j.atmosenv.2008.08.010

Ying, S., Zeng, D.-N., Chi, L., Tan, Y., Galzote, C., Cardona, C., Lax, S., Gilbert, J., Quan, Z.-X., 2015. The influence of age and gender on skin-associated microbial communities in urban and rural human populations. PLOS ONE 10, e0141842. https://doi.org/10.1371/journal.pone.0141842