Summer 8-15-2017

Understanding the Relationship Between Hosts and their Microbiome

Boahemaa Adu-Oppong
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS
Division of Biology and Biomedical Sciences
Evolution, Ecology and Population Biology

Dissertation Examination Committee:
Gautam Dantas, Chair
Carey-Ann Burnham
Justin Fay
Andrew Kau
Barbara Kunkel
Scott Mangan
Jonathan A. Myers
Ting Wang

Understanding the Relationship Between Hosts and their Microbiome

by
Boahemaa Adu-Oppong

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2017
St. Louis, Missouri

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

REB – Root Endophytic Bacterial
HGT – Horizontal Gene Transfer
UTI – Urinary Tract Infection
FDR – False Discovery Rate
AMR – Anti-microbial Resistance
MDR – Multi-drug resistance
BAP – Blood Agar Plate
MAC – MacConkey

ACKNOWLEDGMENTS

This work was funded in part by a National Science Foundation graduate research fellowship (award number DGE-1143945). This work was funded by NIH Director's New Innovator Award, the National Institute of Diabetes and Digestive and Kidney Diseases (DP2DK098089), Barnes Jewish Hospital Foundation-Washington University Institute of Clinical and Translational Sciences, and National Institute of General Medicine (R01GM099538) awarded to Gautam Dantas.

I would like to thank my advisor, **Gautam Dantas**, for his guidance and unwavering support throughout my Ph.D. I never once left a mentoring meeting feeling hopeless but rather determined to complete the task at hand. His abilityto create a sense of community within the lab produced a work environment that was conducive to not only great science but great long-lasting relationships. I am forever indebted to Gautam for his appreciation for outreach and diversity. He encouraged my endeavors with the Young Scientist Program, even though it took away time from research. However, because of his support I was inducted into the Edwards A. Bouchet Fellowship, awarded the Gerry and Bob Virgil Ethic of Service Award, awarded the Distinguished Individual Leader Award and named Volunteer Choice Outstanding YSP Director.

I am appreciative to those who served on my thesis committee: **Carey-Ann Burnham**, **Scott Mangan**, **Jonathan Myers**, **Justin Fay**, **Ting Wang**, **Barbara Kunkel** and **Andrew Kau**. After each thesis update, I left more excited about my project than before. Thank you for your critical advice and feedback. I would like to thank **Scott Mangan** for his willingness to spend hours discussing my project and most importantly increasing my knowledge on statistics. **Carey-Ann Burnham** for consistent solid feedback and help with obtaining samples.

I would like to thank my colleagues in the **Dantas Lab** for their feedback and support throughout my graduate school education. I had the privileged of working with Dantas Lab Cohort 1 and 2 and it has been a wonderful experience. I had a lot of help from collaborators across campuses. I would like to thank **Meghan Wallace** for her help with processing many of the urine samples. **Christopher Catano** for help with the null-modeling of the bacterial communities. Many thanks to **Mike Dryer**, **Darlene Branson**, **Khadine Gomez**, **Max Herzog**, **Kevin Stiles** and the **undergraduates of the Mangan Lab** for their hard work in the greenhouse. It was a privileged to work within the **Center of Genome Sciences** and I want to especially thank **Eric Martin** and **Brian Koebbe** for their technical support on the cluster and **Jessica Hoisington-Lopez** for her technical support with sequencing.

I would like to thank the support of my friends who became family in St. Louis. Special thanks to **Jill Sander-Chali**, **Kalaba Sander-Chali** and **Mapalo Sander-Chali** for making me feel at home in St. Louis. Many thanks to **Samantha Lurie** for being there for me during the tough times and allowing me to be part of The Show Me Costa Rica Project. A special thank you to **Damon Williams** who never doubted my abilities and for his continual support and love. Thank you to the EEPB girls **Jennifer Gruhn**, **Cassie Vernier**, and **Erika Schumacher** for a great friendship that I will always cherish. In particular, to **Kimberley Sukhum** for always being a great sounding board and an incredible friend. The Young Scientist Program has been a major part of my graduate student career and without **Reyka** or **Sohini** it would have been harder to run the organization as we did. They not only became my partners in crime but genuine friends. A big thank you to **DeAngela Hayes** for first and foremost choosing WashU to complete her Master's in Business so that we could continue our great friendship and make our bond stronger. Thank you for the amazing dinners and time we spent with each other. To **LaMar Bonham**,

thank you for bringing nothing but pure happiness, joy and sense of stability into my life. I can't wait for what the future holds for us. I would also like to thank my friends from Houston and abroad for their words of advice and support. In particular, to **Sarah Lyons** and **Sarah Agara** for their love, support and words of encouragement throughout the years.

The one woman who made this all possible is **Joan Strassmann**. I never would have thought about attaining a PhD if I had not stumbled upon a Biology discussion session lead by **Chandra Jack**. No words can express the magnitude of thanks I have for both amazing women. Thank **Joan** for being my cheerleader and for your continual support throughout the years.

Last but not least, I would like to thank my family near and far. My immediate family, **Romeal Phoquo**, **Alhassan Adu-Oppong**, **Fabeah Adu-Oppong**, **Twenewaa Adu-Oppong**, and **Aiden Adu-Oppong** have been there for me since day one and have never let me feel discouraged or defeated. They have always supported my dreams and aspirations and continued to shower me with words of affirmation even when I didn't deserve them. I don't know if I would have been able to do this without such a loving and supporting family. A special thank you to **my mother**, even though we didn't have much, you always found a way to make us feel like we had it all. Thank you.

Boahemaa Adu-Oppong

Washington University in St. Louis

August 2017

Dedicated to **my family**,
those who are here with us
and those who are watching over us.

ABSTRACT OF THE DISSERTATION

Understanding the Interplay Between Microbial Communities and Their Hosts

by

Boahemaa Adu-Oppong

Doctor of Philosophy in Biology and Biomedical Sciences

Evolution, Ecology and Population Biology

Washington University in St. Louis, 2017

Professor Gautam Dantas, Chair

Microbes are bountiful and associated with every animal and plant kingdom. Furthermore, microbes can alter host phenotype, development, health and functioning. However, this is not a one-way interaction, hosts can structure microbial communities by changing the environment to be suitable for certain microbial species. Several studies have characterized microbial communities associated with hosts to answer two main questions in ecology: who's there, and what are they doing? However, two questions from the field of community ecology are often ignored (1) what forces are structuring the microbial communities (how was the community formed) and (2) how stable are these communities. Vellend synthesized that all communities are governed by four main processes: drift, selection, speciation and dispersal. These processes can be grouped into 2 components of assembly, either deterministic (selection, speciation, dispersal) or stochastic (drift, dispersal limitation). The goal of my thesis was to (1) understand the relative contribution of these processes on microbial communities and (2) how stable is the assemblage of microbial community over time and during an infection.

In order to determine if microbial communities are structured deterministically or stochastically, I studied the root endophytic microbiome, which has been shown to directly impact plant physiology. By analyzing 252 root endophytic bacterial (REB) communities, which

had been perturbed using antibiotics and sterilization, I show the communities are assembled deterministically. The strongest selective force structuring the REB communities was plant identity even in a perturbed state. I demonstrate the interplay between REB communities and plant phenotype by linking the variation in the reduction of biomass in autoclaved soils to changes in the abundance of bacterial species. This suggests hosts can selectively increase or decrease the abundance of bacterial species that will increase the plant's fitness. Consequently, this allows plants to co-exist by specializing on different bacterial species.

To determine the stability of microbial community structure, I studied the urine microbiome of individuals who are do not have urinary symptoms and those who are suspected to have a Urinary Tract Infection (UTI). By analyzing the urine microbiome of 220 urine samples, I show that the urine microbiome is in an altered state during an infection and is stable over time in asymptomatic women. Asymptomatic individuals are enriched with *Lactobacillus crispatus* and *L. iners* while individuals with suspected UTIs are enriched with *Ruminococcus torques, Propionibacterium acnes* and *Escherichia coli*. There is a plethora of putative pathogens uncovered only with non-conventional culturing methods. Roughly 21% of individuals with suspected UTIs did not have the putative cultured pathogen at high relative abundance but a different known UTI pathogen when direct sequencing was utilized. This suggests that UTIs could be caused by a dysbiosis of the urine microbiome rather than direct inoculation of an organism from the gastrointestinal tract.

Collectively these studies show that microbial communities can be structured by the host and host state, and are deterministically assembled. Further work to investigate how the host can structure the microbial community possibly through changing environmental conditions or through immune response is warranted.

Chapter 1 Introduction to Microbiota and Host Dynamics

Bacteria are one of the most ancient and abundant extant organisms on our planet with 4 x 10^{30} individuals [1]; these microbes live in diverse habitats [2]. Due to their large population size, short generation time, small size, ability to disperse passively [3-6], ability to dormant for decades [7] and phenotypic plasticity [8, 9] it has been a challenge implementing a macro-ecological framework of community assembly to micro-organisms [10]. However, we have discovered that bacteria do display biogeographic patterns [2, 11-14] and can be limited by dispersal [15]. Therefore, we can reject the simplistic notion which impacted microbial ecology for decades that "*Everything is everywhere, but the environment selects*" [16]. This is extremely important regarding the microbiota of humans and plants which have been evolving with eukaryotes for thousands and millions of years, respectively. The environment is not the only force driving the assemblage of microbes associated with hosts.

Vellend produced a unified framework for community assembly [17] and has incorporated into microbial ecology [10, 18-21]. This framework is centered on 4 main processes which shape community assembly (1) drift, (2) selection, (3) dispersal and (4) speciation. Drift is random changes in species relative abundance [17]. For microbial communities, drift becomes important when communities are under weak selection and have low alpha diversity and observed species richness [10]. This process can cause extinction of low abundant species; therefore, it is important to estimate drift to protect focal species. The ideal approach to determining if a community is assembled stochastically is by applying a 'null model' which randomizes community composition data [22, 23]. Deviations from the null model are used to

quantify the relative influence of stochastic and deterministic processes [24, 25]. Therefore, great consideration should be taken when choosing a null model since they can lead to different interpretations of the observed data [26, 27]. Many studies provide evidence that drift has a strong influence on microbial communities [28-31]. For example, the structure of microbial communities in zebrafish were explained by neutral processes [30]. Therefore, random process should not be ignored when determining what processes drive community assembly in microbial communities.

The second process, selection, is a force that is directly impacting the relative abundance of species in a community [10]. Many studies have quantified the effect of abiotic conditions [32-37] and biotic interactions [38-43] on structuring microbial communities which suggests selection can play a large role. However, few have quantified the amount of variation which can be explained by solely selection [37, 44, 45]. For example, the bacterial communities on aquifers found in the Hanford formation range (coarse-grained) have weaker selection forces structuring the microbial communities compared to those found in the Ringold range (finer-grained) [37]. This implies that selection can explain some of the variation in microbial community composition but is rarely acting alone. Therefore, studies that solely show that some of the variation in microbial community composition is explained by an abiotic or biotic interaction does not prove that selection is the only force shaping the community.

The third process, dispersal, is the movement of microbes within space and time [18]. Many assume dispersal for microbes is a stochastic process since many disperse passively. Consequently, passive dispersal (dispersal limitation) is not enough to cause variation in microbial community composition [37]. The combination of passive dispersal and drift can lead to differences in composition between communities [37, 46]. However, each microbial species

can differ in dispersal probability which could cause the assembly process to be non-random [10]. If dispersal is active and frequent, then communities would have similar compositions [37]. For example, the dominant process which shapes the presence or absence of bacterial taxa in the gut microbiome of adults in Papua New Guinea is high dispersal while in the United States it is a combination of high dispersal and selection [44]. This further proves that bacteria can actively disperse which could lead to the homogenization of microbial community composition.

The last process, speciation or local diversification, is the creation of new species in the environment. For microbial communities, this can happen over a short period of time due to horizontal gene transfer (HGT) and over decades due to the ability for microbes to remain dormant for thousands of years [10, 18]. Mutation is another form of diversification for microbial communities. For example, a bacterial community may rapidly evolve to become resistant in the presence of an antibiotic [47]. Applying these four processes on microbial community assembly can shed light on the variation seen between host microbiota [18]. The four processes can further be grouped into two processes: stochastic and deterministic [25]. Stochastic processes are unpredictable disturbance, probabilistic dispersal and random births-deaths, while deterministic processes are abiotic environment (‘environmental filtering') and both antagonistic and synergistic species interactions [25, 31]. In Chapter 2, we will focus only on the first three processes (Figure 1), and in Chapter 3 we will focus on selection and speciation.

With current sequencing technology such as targeted marker gene sequencing (ex. 16S rRNA), whole genome sequencing, and shotgun metagenomic sequencing, we can interrogate microbial communities and understand not only “who is there”, but also elucidate function [19]. Marker gene sequencing, such as sequencing the 16S rRNA gene, amplifies housekeeping genes which are used to create phylogenetic species trees [48]. This technology has allowed us to

characterize the 99% that is not readily culturable with conventional culture media [49]. Whole genome sequencing has allowed us to sequence entire isolate genomes to study population dynamins which we commonly use to understand disease outbreaks [50-53]. Lastly, shotgun metagenomics has allowed us to answer both questions “who is there” and “what are they doing” using a community ecology perspective. Bacterial communities can be tracked over time so we can better understand the relationship between host and their microbiome [54-57].

The evolution and diversity of animals and plants have been impacted through symbiotic

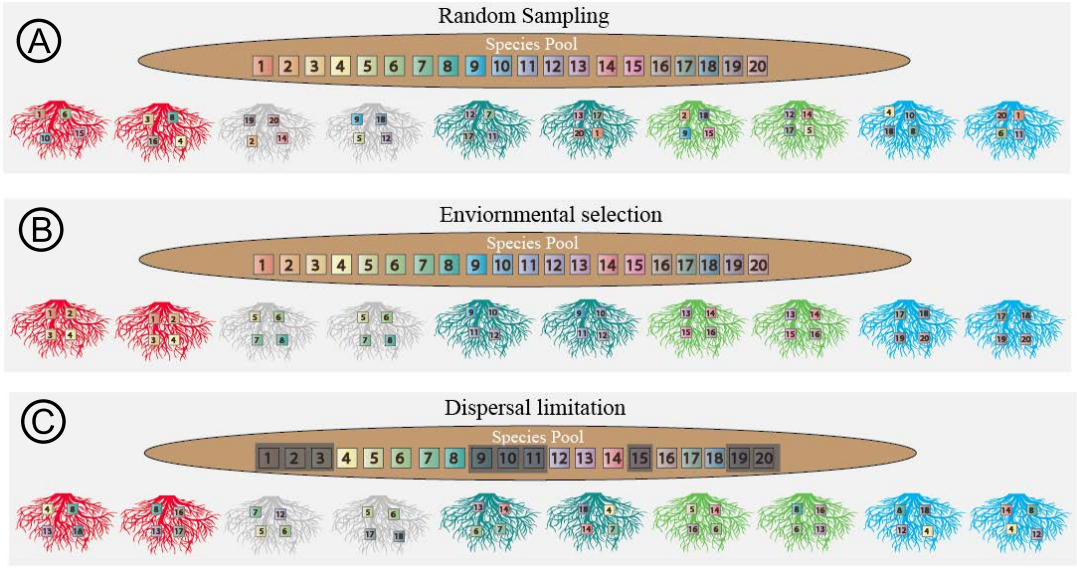


Figure 1-1. Community assembly processes which could give rise to different REB communities. Each panel is the result of one of the three process which could affect local assembly in 2 individuals from 5 plant species (red, grey, dark green, light green, and blue) from the same regional species pool. The regional species pool is comprised of 20 different bacterial species. (A) The REB communities are assembled randomly and differ among individuals of the same plant species. (B) the REB communities are assembled deterministically and individuals of the same plant species have identical communities. In (C) the REB communities are assembled deterministically since the species not greyed out are the only ones able to disperse into plant roots and the communities across plant species are heterogeneous due to random processes.

relationships with microbes [58]. Two relationships I explore in my thesis are the microbiome harbored in the human urinary bladder and prairie plant roots. The root is an organ which facilitates in nutrient uptake such as nitrogen of phosphorous from the soil [59]. The urinary

bladder is an organ which stores the filtrates from blood in the human body. The concept of the urine microbiome is in its infancy; therefore, the function of microbes in the bladders is still unknown [60]. Many have speculated that the microbes are able to outcompete pathogens and stimulate the immune system [60]. The root microbiome harbors a diverse community of microbes which are regarded as the host's extended phenotype [59, 61, 62] and the same can be speculated about the urine microbiome. Therefore, changes in composition or functionality of the microbiome, will affect the host. To combat microbial diseases such as Urinary Tract Infections (UTIs) we need to understand the ecology of the disease and its impact on the native microbiome of the urinary tract.

Many studies within microbial ecology have focused on characterizing and cataloguing microbial communities across various habitats. However, experimental manipulation is necessary to link patterns and processes [63]. With the ability to perturb communities we can begin to understand their stableness [64]. If the composition of the microbial community is unchanged, the community is resistant to disturbance. If the composition is altered but after time returns to the original composition, the community is resilient. If the composition is altered but performs the same functions as the previous community, the community has functional redundancy [64]. This can then be used in conjunction with disease state and community assembly theories to understand the interplay between host phenotype and the microbiome.

My first objective was to determine if microbiomes associated with hosts are assembled deterministically and shaped primarily by the host. To evaluate this, we conducted a greenhouse experiment where plant roots are grown in different soil treatments for the duration of four months. We sequenced the 16S rRNA gene of bacteria that resided inside the roots of the plants. To determine the strengths of non-random and random processes on the assembly of REB

communities, we implemented a null model. To determine which selective force could explain the most variation in the REB communities, we conducted a multi-variate analysis which partitioned the variance by our treatments. I collaborated with the Mangan Lab to conduct the greenhouse experiment. I performed the sequencing analysis, as well as generating figures and will be the primary author of this chapter.

My next objective was to determine if the microbiome can shift into an altered state due to host state (diseased vs health) and if the microbiome of the urine is comprised of clonal or diverged bacteria. We collected remnant banked urine samples from patients suspected to have UTI patients and urine from asymptomatic women. We cultured and sequenced isolates from patients with suspected UTIs to determine if the putative pathogens were clonal or non-clonal. We sequenced culturable bacteria on conventional media to determine if the population was similar between diseased and non-diseased state. To determine if there is a difference in the urine microbiome between patients presumed to have urinary tract infection and asymptomatic patients, we performed shotgun sequencing on the urine of both populations. I collaborated with the Burnham lab to collect, culture and extract DNA from cultured isolates. I performed the sequencing analysis, as well as generating figures and will be the primary author of this chapter.

Chapter 2 Prairie plants host distinct root endophytic bacterial communities assembled by non-random processes

Boahemaa Adu-Oppong[1], Scott Mangan[2], Claudia Stein[2], Christopher P. Catano[2], Jonathan A. Myers[2], Gautam Dantas[1,3,4,5*]

Affiliations:

[1]Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, Missouri, United States of America
[2]Department of Biology and Tyson Research Center, Washington University in St. Louis, Missouri, United States of America
[3]Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO, USA.
[4]Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO, USA.
[5]Department of Biomedical Engineering, Washington University, St. Louis, MO, USA.

2.1 ABSTRACT

Plant associated microbes can influence community assembly, the maintenance of biodiversity and stability of ecosystems. However, we know little about the relative strength of forces (host-based selection or environmental-based selection) which can contribute to the assembly of plant associated microbes and how they are assembled (deterministically or stochastically). Even less is known about how the composition of the microbes can directly impact plant fitness. We grew five prairie species in perturbed soils to test for the relative strength of selection on the assembly of root endophytic bacterial communities. Despite soil perturbations, root endophytic bacterial communities assembled deterministically structured by host identity which explained most of the variation in the difference of composition between root endophytic bacterial communities. Additionally, biomasses correlated with turnover of bacterial

community composition and individual bacterial taxa. These results suggest plants co-exist due to stabilizing niche differences by controlling the assemblage of root endophytic bacterial communities.

2.2 INTRODUCTION

Elucidating mechanisms that structure communities and affect ecosystem processes is a long-standing goal in plant ecology. Plant community composition, diversity and stability are driven by a multitude of abiotic and biotic factors: climate, age, environmental harshness, area, isolation, disturbance, environmental heterogeneity and plant-soil feedback [65, 66]. Over the past few years, plant-microbe interactions have been identified as a mechanism driving plant structure and affecting ecosystem processes [67-73]. Although there has been an emphasis in understanding plant-microbe interactions, there is a dearth of studies focusing on the driving forces structuring microbial communities.

Understanding the mechanisms which lead to the divergence of microbial communities is essential to understanding how plants and microbes interact. The microbial community is an extension of the plant phenotype by increasing uptake of nutrients from the environment [61]. If microbial communities are structured mainly by plant hosts, then different plant species can co-exist due to differences in acquisition of microbes from the same environment. Optimization on assemblage of microbes could lead to different abilities in resource uptake which can cause stabilizing niche differences [74]. Therefore, if microbes can influence plant community composition and diversity then plant biomass should correlate with the divergence in composition of microbes.

Plants interact with microbes mainly through the soil and reside in three niches: bulk soil, endosphere, and rhizosphere. Prior root-associated microbiome studies in both model plants (e.g.

Arabidopsis thaliana), agricultural plants (e.g. corn, rice) and non-agricultural plants (e.g. eastern cottonwood, agave) have established that the microbiome inside the root endophytic compartment (endosphere) is distinct from both bulk soil microbiomes and the microbes associated on the outside surface of plants roots (rhizosphere) [75-82]. Bacterial endophytes are defined as bacteria that can be isolated from surface-sterilized plant tissue and do not visibly harm the plant [83]. We focused on the endophytic bacteria because they can influence the growth and development of plants [84-86] and provide greater beneficial effects than rhizosphere-colonizing bacteria [87].

Endophytic microbes are thought to be structured by a two-step selection process [88]. The first selection is host rhizodeposition and cell wall features which promote growth of organotrophic bacteria [88]. The second selection is host genotype factors which fine-tunes the microbial community [88]. The theory of the two-step selection process has been hypothesized using observational data from studies rather than experimentally manipulating conditions to test the strengths of deterministic factors. It neglects other deterministic factors which could lead to the same divergence pattern. Most importantly, the two-step selection process assumes that divergence of microbial communities is deterministic rather than stochastic.

Our *first hypothesis* is root endophytic bacterial (REB) communities are influenced by host, if most of the variation in the differences in composition of REB communities after perturbations is explained by host. Perturbations provides insight to the key drivers of community dynamics [89]. After perturbation, we can test the strength of deterministic factors on the assembly of the REB communities. Rhizosphere bacterial communities are heavily influenced by plant hosts [90-92]. Multiple studies have shown that soils trained by one plant species can affect the growth of conspecifics and heterospecifics [66, 93, 94]. However, it has

yet to be determined whether the resulting REB communities are assembled due to strong host selection or due to dispersal limitation. If dispersal is not limited then when plants are grown in the same trained soils, the REB communities would be homogenous. Alternatively, if dispersal is limited then when plant hosts are grown in the same trained soils, the REB communities would be heterogeneous. We also introduced two other types of perturbations: autoclaving of soils and application of antibiotics. If the communities are resilient, then after perturbations we would expect the divergence in microbial communities to be largely still explained by deterministic factors.

Our *second hypothesis* is that root bacterial endophytic communities are structured by deterministic processes. Divergence in microbial communities can be influenced by either stochastic or deterministic processes. Variation in microbial communities can arise through stochastic processes such as dispersal limitation, diversification, mass effects and random demographics [10, 11, 18, 95]. Deterministic processes can also shape microbial communities through environmental heterogeneity, species interaction and niche partitioning [10, 11, 18, 95]. Distinguishing between the two processes can be done by creating a null model which produces a pattern that would be expected in the absence of an ecological mechanism (i.e. selection) [27].

Our *third hypothesis* is that if REB communities can influence plant community composition, then REB can influence plant fitness. Plant-soil interactions have largely been studied through the framework of the influence of soil communities on plant fitness [96-100]. However, most of these studies have either treated soil microbes as an undefined, “black-boxed” variable, or have used culture-based methods which interrogate less than 1% of known soil microbes [101-104]. Therefore, we have focused on a subset of microbes known to influence plant fitness and have not been investigated before.

We conducted a greenhouse experiment using 5 prairie species because the prairie is one of the most endangered ecosystems in the world [105]. We perturbed soil communities by disturbance (antibiotics and autoclaving) and soil history (soil was trained by plant hosts). We conducted a null model in the absence of selection to test for the strength of determinism or stochasticity on the resulting microbial community. Additionally, we correlated plant fitness to the diversity of endophytic communities to elucidate the influence of REB communities on plant fitness. We found that most of the variation in REB composition was explained by host and assembled deterministically. Correlations between growth and composition of root endophytic bacterial community were evident for 4 of the 5 plant hosts tested. This suggest that plants co-exist due to their ability to structure the REB communities resulting in stabilizing niche differences.

2.3 MATERIALS AND METHODS

2.3.1 PLANT HOSTS AND SEED COLLECTION

We chose 5 prairie species: *Monarda fistulosa* (Wild Bergamot, *Ratibida pinnata* (Grey-head coneflower), *Carduus nutans* (Musk Thistle), *Conyza canadensis* (Horseweeed) and *Heliopsis helianthoides* (Smooth oxeye). These species were chosen because they are highly abundant in the prairie; therefore, we could collect enough soil to conduct the greenhouse experiment. We began collecting seeds June 2013. We purchased all seeds from Prairie Moon Nursery (Winona, Minnesota, USA) except *Conyza canadensis* which were donated from Mike Dryer from the Greenhouse Facility at Washington University in St. Louis and *Carduus nutans* was collected at Tyson Research Center.

2.3.2 DETERMINING STRENGTH OF DETERMINISTIC FACTORS

Greenhouse experimental set up

To ensure that roots were colonized by microbes in the collected inoculum, we surface sterilized and germinated seeds in autoclaved (gravity cycle for 65 min twice) Propagation Mix (Sungro horticulture Agawam, MA, USA).

We conducted a semi-full factorial design to investigate if dispersal limitation or host selection was a driving force in structuring REB communities. To link changes in biomass to the soil biota, we controlled for abiotic soil effects by filling pots with 6% inoculum and 94% background soil [99]. The inoculum comprised of rhizosphere collected from each species in the experimental prairie site. The background soil was an autoclaved (gravity cycle for 65 min twice) mixture of aggregated field soil-sand mixture (2:1).

Fourteen replicates of each plant host received heterospecific inoculum. Twenty-four replicates of each plant host received conspecific inoculum. Six replicates for each plant host received conspecific and heterospecific autoclaved inoculum. Half of all replicates were subjected to an antibiotic treatment. This was our third perturbation which would allow us to further test the strength of deterministic factors on REB community composition. This resulted in 5 (plant hosts) x [4 (heterospecific inoculum) x 2 (antibiotic treatment) x 7 replicates + [1 (conspecific inoculum) x 2 (antibiotic treatment) x 12 replicates]] + [5 (plant hosts) x 5 (autoclaved inoculum) x 2 (antibiotic treatment) x 3 replicates] = 550 experimental units in a semi-full factorial design.

Perturbations: Autoclaving and Antibiotics

Autoclaving soil perturbs the microbial community by reducing the number of bacterial species in a community. We autoclaved half of the collected inoculum (gravity cycle for 65 min followed by a second gravity cycle for 65 min 24 hours later). After perturbations, we calculated the strength of deterministic factors in the structuring of the altered REB communities.

Antibiotics were chosen as a perturbation due to its ability to directly affect microbial communities by eliminating species from the communities without directly impacting plant growth. We chose four antibiotics: chloramphenicol (8mg/L), oxolinic acid (0.2 μg/mL), gentamicin (32mg/L or 4mg/L), streptomycin (512mg/L). Chloramphenicol and gentamicin are used in agar plates when isolating fungi to decrease the presence of bacteria. Oxolinic acid, gentamicin and streptomycin are used in the plant-agriculture community to target bacterial pathogens that affect crops. Chloramphenicol is a broad range antibiotic that is bacteriostatic and inhibits protein synthesis by binding to the 50S ribosomal subunit (Sigma Product Information). Oxolinic acid is effective against gram-negatives and is a quinolone compound. It inhibits the DNA gyrases (Sigma Product Information). Gentamicin is a broad range antibiotic that inhibits bacterial protein synthesis by binding to the 30S subunit of the ribosome (Sigma Product Information). Streptomycin is a broad range antibiotic but has been known to be less effective against gram-negative aerobes. It blocks protein synthesis by targeting the 70S ribosome. The concentrations of the antibiotics were determined from EuCast2 or from searching the literature.

Pots not treated with antibiotics were administered 10ml of autoclaved DI water. The first treatment was given July 12, 2013; we administered 10 mL of the antibiotic cocktail. For the other treatments (July 20th, July 29th, August 5th, August 22nd, September 13th) we administered 15 mL of the antibiotic cocktail.

Plant care and trait measurement

The experimented started July 2013 and ended October 2013. The duration was chosen to give all plants optimal time for growth. Plants were top watered as needed with RO water. All pots arranged twice into randomized blocks and maintained in controlled greenhouse conditions for the duration of the experiment. Dropped leaves were collected and included in total biomass

for the individual. To minimize insects traveling from pot to pot, yellow sticky traps were set up throughout the bays. At the end of the experiment, we harvested both above and below ground biomass and placed biomass in envelopes. We measured dried biomass.

Characterization of REB communities

To characterize the REB communities, we weighed approximately a gram of belowground biomass for microbial extraction and stored in -80^{o} C. To accurately measure belowground biomass, total belowground biomass was weighed before and after removal of portion used for microbial extraction. The estimated loss was calculated and added to the dried biomass weight.

Belowground biomass was resuspended in 15 mL of filter sterilized PBS-S buffer (130 mM NaCl, 7 mM Na2HPO4, 3 mM NaH2PO4, pH 7.0, 0.02% Silwet L-77) and sonicated (Fisher Scientific Sonic Dismembrator Model 500, Pittsburgh, PA, USA) at low frequency for 5 min with 5 30 sec bursts followed by 5 30 sec rests for 252 root samples. We collected 14 samples (After Sonication) after this stage and submitted them for sequencing. Then roots were resuspended in 15 mL of filter sterilized PBS-S buffer and centrifuged at 1,500 *g* for 20 minutes. We collected another 14 samples (After Wash) after this stage and submitted them for sequencing. The roots were aseptically transferred to a new 15 mL conical tube and freeze dried overnight. Microbial community was extracted from roots per manufacture's protocol using the PowerSoil DNA Isolation Kit (Mo-Bio Laboratories, Carlsbad, CA, USA). We performed PCRs in triplicates to control for bias in PCR reactions and amplified the 16s rRNA gene V4 region (http://www.earthmicrobiome.org/emp-standard-protocols/16s/) using the barcodes designed in [106]. Before sequencing, we visualized the bands on gels. After a positive confirmation, we combined all samples and sequenced on Illumina MiSeq (Illumnia Inc., San Diego, CA, USA)

with 2x250 bp paired-end reads at the Center for Genome Sciences at Washington University in St. Louis. Sequences were demultiplexed using QIIME [107]. Paired-end reads were truncated at the first base with a quality score of <Q4 and then merged with usearch [108], with a 100% identity in overlap region and a combined length of 253±5 bp. The merged reads were then quality filtered by usearch with a maximum expected error of 0.5. Operational taxonomic units (OTUs) were picked using the usearch pipeline [108] and known chimera OTUs were filtered from the list. Reads were matched to OTUs at 100% sequence identity. Representative sequences from each OTU were aligned using PyNAST and assigned taxonomy using RDP Classifier using QIIME version 1.5.0-dev. Any sample with fewer than 30 OTUs were not dropped from the study. Additionally, OTUs which were not found in at least one sample or had fewer than 30 individuals were removed from the dataset for a total of 595 OTUs.

Microbial community count data was transformed using the DESeq2 package in R based on previous recommendations [109]. All analyses were performed using the package 'vegan' v.2.4.1 [110], 'RVAideMemoire' v.0.9.61 [111] and 'phyloseq' v.1.18.1 [112] in R version 3.2.2. Principal coordinates (PCoA) of Bray-Curtis pairwise dissimilarities were identified using the vegan function 'capscale'. To explain the difference in dissimilarity of microbial communities, we tested the effect of host, soil history, autoclaving of field soil and exposure to antibiotics in a full model using the non-parametric permutation test ADONIS II in package 'RVAideMemoire' with 999 permutations. We corrected for multiple comparisons with the False Discovery Rate post-hoc test to determine which pairs were significantly different.

2.3.3 DETERMINISTIC VS STOCHASTIC PROCESSES

We wanted to study the effect of our treatments on the assembly of the REB communities. Measures of β diversity can be used to determine whether communities are

assembled deterministically or stochastically. However, because β diversity is dependent on α, any effect of our treatments on α can alter β simply through numerical sampling effects. Therefore, we implemented the Raup-Crick (RC) null model described in [113] to estimate effects on β diversity not simply due to changes in alpha that are stochastic with respect to species identity. Changes in RC beta-diversity can be used to infer the strength of underlying assembly mechanisms (deterministic vs neutral) [113-115], with some caveats. Both low alpha diversity and demographic stochasticity can limit inferences from this metric [116]. However, alpha diversity and total microbial abundance is high in all our treatments (Figure 2-S1).

We determined the species pool as the total number of species and their observed occupancy across the plant host by soil inoculum. Species were randomly sampled from the pool in proportion to their occupancy, and assigned to local communities to create a null distribution of the expected number of shared species between pairwise plant hosts. The RC value will indicate whether the REB communities is less similar (approaching 1), as similar (approaching 0.5) or more similar (approaching 0), than expected by chance. The higher the deviance from 0.5 (purely stochastic), the more deterministic is the community assembly.

2.3.4 LINKING BELOWGROUND SPECIES COMPOSITION TO PLANT FITNESS

To understand if differences in composition of REB communities could affect plant diversity, we first characterized differences in fitness which could be explained by the different perturbations. We log transformed biomass to compare fairly treatment effect on biomass for different species [99] and conducted an ANOVA to test for the effect of autoclaving of field soil, exposure to antibiotics, plant host and soil history. We also tested for the effect of interactions between plant host and location of soil collection to ensure soils collected in different plots did not affect biomass. To link fitness to acquisition of REB communities in perturbed states, we

tested for an interaction between plant host and autoclaving of field soil and plant host and soil history. To test whether those differences in fitness could result in turnover in REB communities, we correlated composition of REB communities and biomass. We used a Mantel test with 999 permutations in the package 'ade4' v.1.7-4. To test whether a taxon of bacteria could affect biomass, we correlated biomass and abundance of bacteria taxon. We used cort.test with pearson correlations in the package 'stats' v.3.3.2. P values were adjusted using Bonferroni. All results were graphed using 'ggplot2' v.2.2.0 [117] in R version 3.3.2.

2.4 RESULTS

2.4.1 PLANT HOST CONTROLS ASSEMBLAGE OF ROOT BACTERIAL ENDOPHYTIC COMMUNITIES

The variation in composition of REB communities are largely influenced by plant host ($R^2_{ADONIS} = 0.073$, $P < 0.001$) and autoclave treatment ($R^2_{ADONIS} = 0.078$, $P < 0.001$) (Table 2-S1). However, soil history ($R^2_{ADONIS} = 0.019$, $P < 0.008$) and antibiotic treatment ($R^2_{ADONIS} = 0.005$, $P < 0.026$) also explained variation in composition of root microbial communities (Table 2-S1). Therefore, other deterministic factors can shape the REB community. These results are supported by CAP analysis, in which samples clustered by host identity (Figure 2-S2a), soil history (Figure 2-S2b), antibiotic treatment (Figure 2-S2c) and autoclave treatment (Figure 2-S2d).

The REB community was heavily perturbed by the autoclave treatment (Figure 2-S3); therefore, we tested for the strength of determinism in the live and autoclaved field soils. The variation in composition of REB communities were still largely explained by plant host in the live ($R^2_{ADONIS} = 0.11$, $P < 0.001$, Figure 2-1) and autoclaved ($R^2_{ADONIS} = 0.23$, $P < 0.001$, Figure 2-S4b) field soil treatments. Despite complete turnover of REB community, the community was

largely influenced by plant host suggesting that the REB community is resilience to perturbations. Additionally, soil history (R^2_{ADONIS} = 0.03, P <0.001, Figure 2-S5a) and antibiotic treatment (R^2_{ADONIS} = 0.008, P < 0.006, Figure 2-S5b) did explain some of the variation between root microbial communities.

2.4.2 Root bacterial endophytic communities are deterministic

Our perturbation treatments influenced changes in alpha diversity (Figure 2-S3); therefore, we implemented a null model which removed selection and controlled for stochastic changes which could be due to sampling effects. All REB communities independent of antibiotic and soil history perturbations were highly convergent (low variation) and deterministically assembled (values of RC approached 0) (Figure 2-2). Therefore, we reject the null hypothesis that the REB communities are not under selection. This provides evidence that the deterministic force shaping the REB communities is the host.

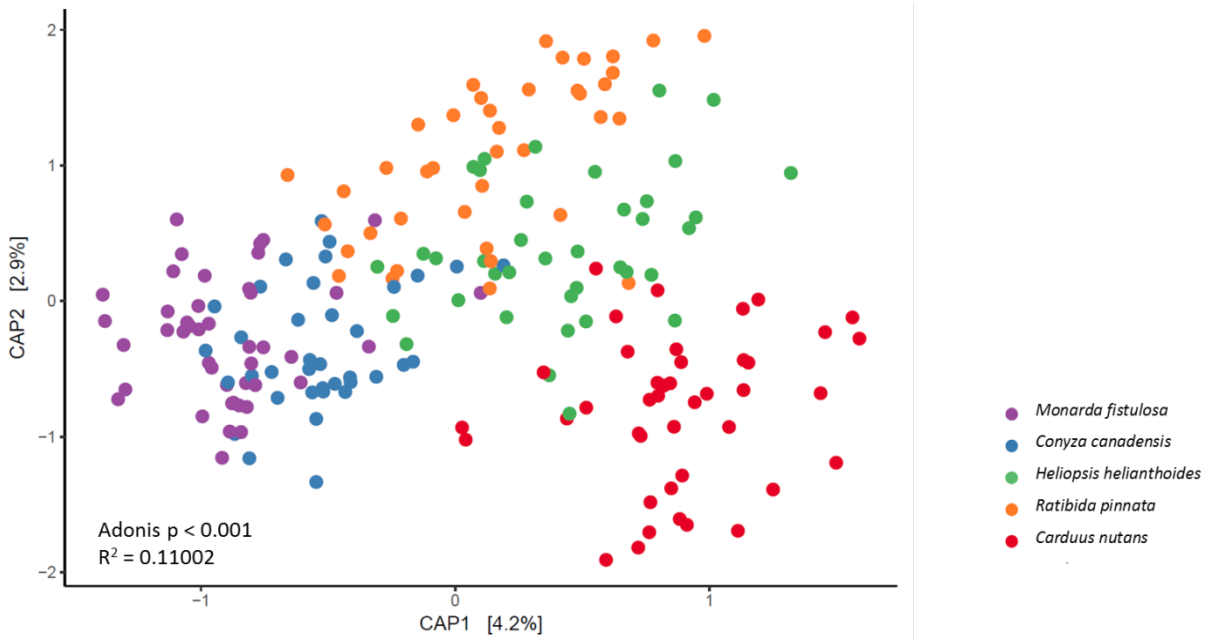


Figure 2-1: REB communities stratified by host identity. Plant host explains more of the variation than soil history and antibiotic treatment in the live soil. Ordination of Bray-Cutis dissimilarities shows clustering of root endophytic bacterial communities by plant host.

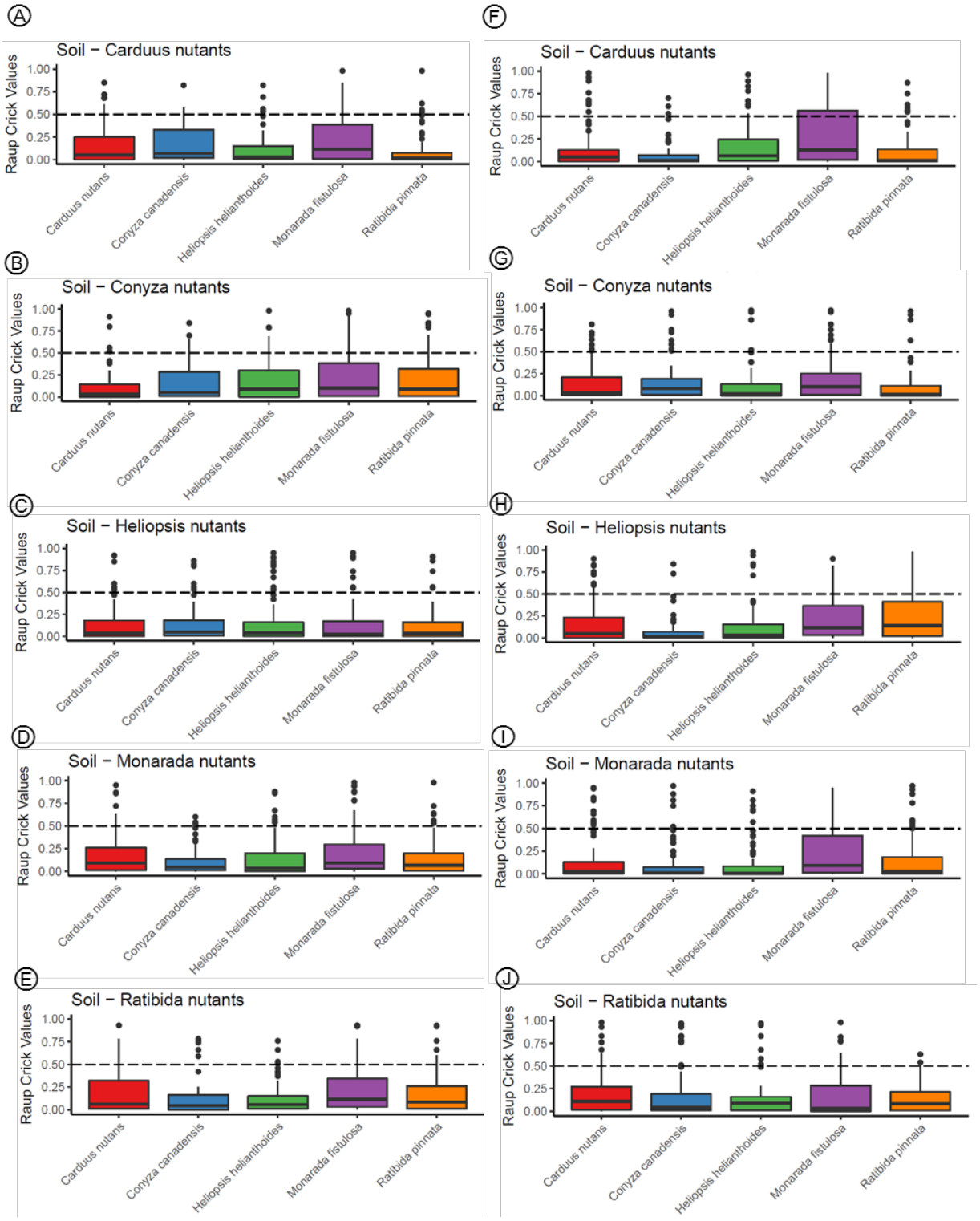


Figure 2-2: REB communities are structured by non-random processes. Raup Crick values generated from the null model for each plant host grown in soils trained by each plant host (A-E) without antibiotics (F-J) with antibiotics.

2.4.3 DIFFERENCES IN BIOMASS CORRELATE WITH DIFFERENCES IN MICROBIAL COMMUNITY COMPOSITION

Overall, host responded to the autoclave treatment in a species dependent manner. Total biomass was affected by autoclave treatment (*Anova* $p < .0001$, Figure 2-3, Table 2-S2). Three plant hosts had lower biomasses in autoclaved field soils, *M. fistulosa, H. helianthodies* and *R. pinnata* (Table 2-S2). *C. nutans* and *C. canadensis* had equivalent fitness in field soils and autoclaved field soils (Table 2-S2).

We then tested whether differences in biomass could correlate with composition of REB communities. *C. nutans* (Mantel $r=0.2$, $p < 0.013$, Figure 2-4a), *H. helianthodies* (Mantel $r=0.4$, $p<0.001$, Figure 2-4c), *M. fistulosa* (Mantel $r =0.2$, $p < 0.01$, Figure 2-4d), and *R. pinnata* (Mantel $r=0.7$, $p < 0.0001$, Figure 2-4e) all demonstrated strong correlation between biomass and community similarity except *C. canadensis* (Mantel $r = -0.04$, $p = 0.7$, Figure 2-4b).

We tested if biomass was correlated with abundance of bacteria taxa to demonstrate that presence of a bacteria taxa could affect plant fitness. There were positive and negative significant correlations found for *H. helianthoides* (Figure 2-5a), *M. fistulosa* (Figure 2-5b) and *R. pinnata* (Figure 2-5c).

2.5 DISCUSSION

Plant roots have been hypothesized to harbor bacteria that are not randomly assembled from the soil but deterministically assembled via a two-step process: edaphic and host factors [88, 118]. Many studies have mainly characterized the communities rather than experimentally manipulating conditions to quantify the strengths of edaphic and host factors on the assembly of REB communities and to create a null hypothesis to understand what the composition of the community would be under no selection [75, 77-79, 82]. In this study, we perturb the bacterial

communities to test whether they are deterministically assembled and determine the strengths of edaphic and host on the assembly of the REB communities. Additionally, we investigated whether if plants with similar fitness (biomass) had similar REB communities. If so, individuals of the same plant species would be competing for similar resources which could impact growth of conspecifics negatively while allowing for co-existence with heterospecifics through niche stabilizing mechanisms. This study goes beyond simply characterizing the composition of REB communities to understanding the mechanisms driving assembly alongside building links between composition and plant fitness.

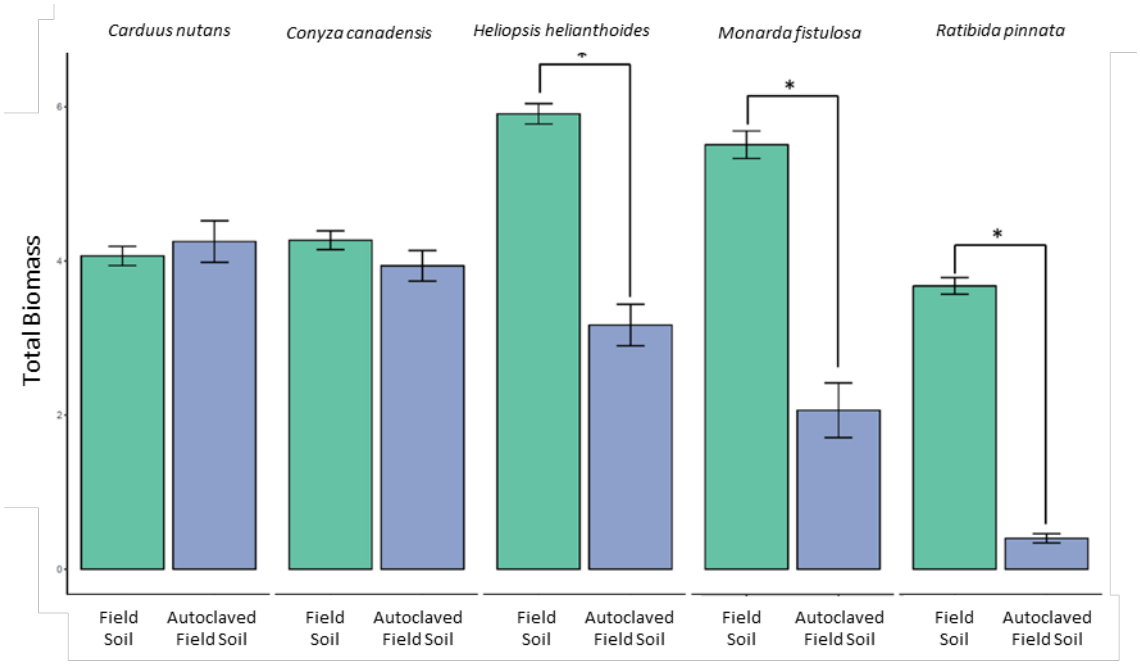


Figure 2-3: Total biomass is effected by microbial disturbance differentially by plant identity. Total biomass (above + belowground biomass) is dramatically reduced in field autoclaved soils compared to field soil for R. pinnata, H. helianthoides, and M. fistulosa. Total biomass remained unchanged for C. nutans and C. canadensis.

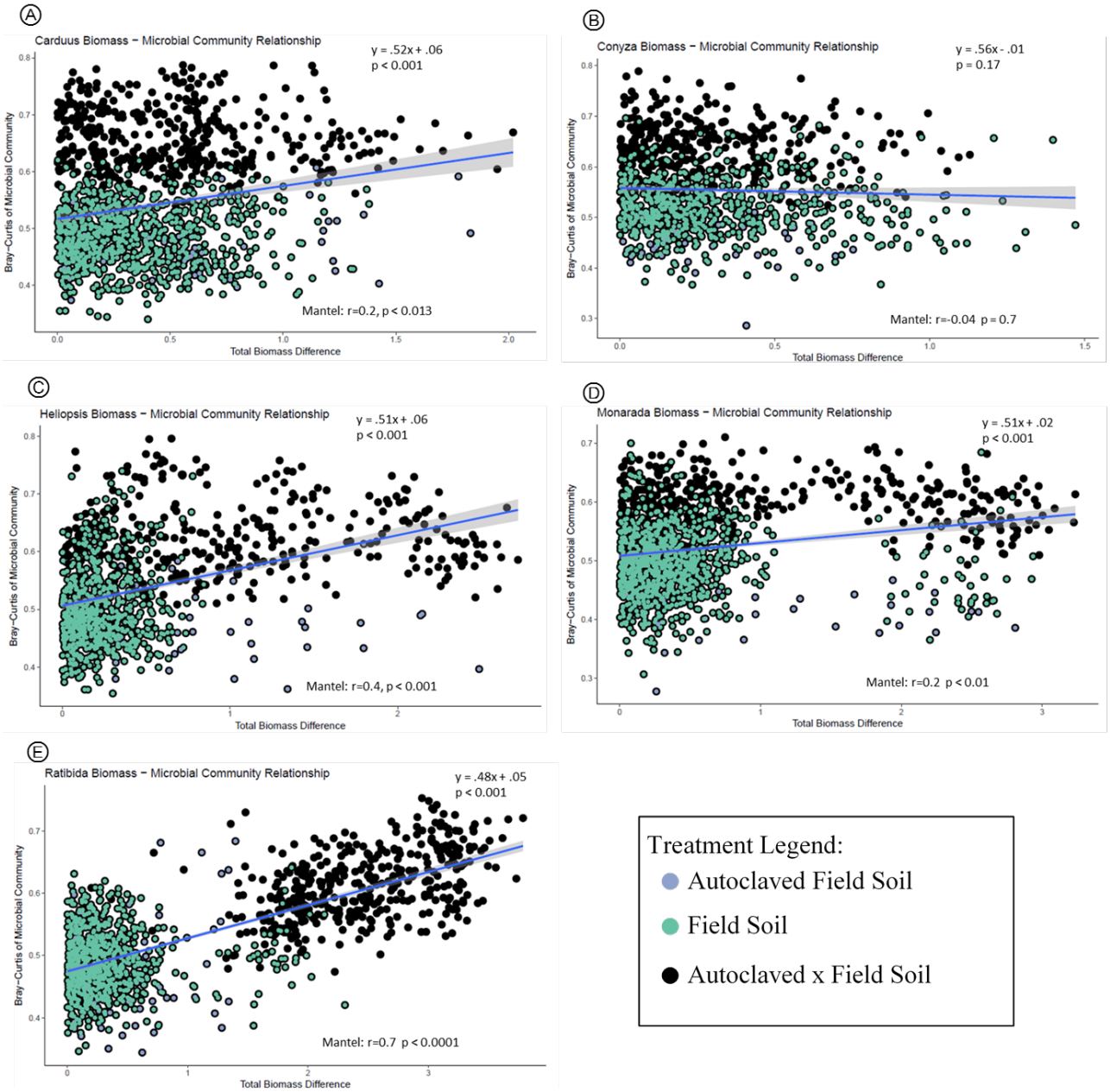


Figure 2-4: Differences in biomass can be explained by variation in microbial community composition for 4 out of 5 plant hosts. Correlations between differences in total biomass and differences in REB communities for (a) *C. nutans* (b) *H. helianthoides* (c) *M. fistulosa* (d) *R. pinnata* (e) *C. canadensis*.

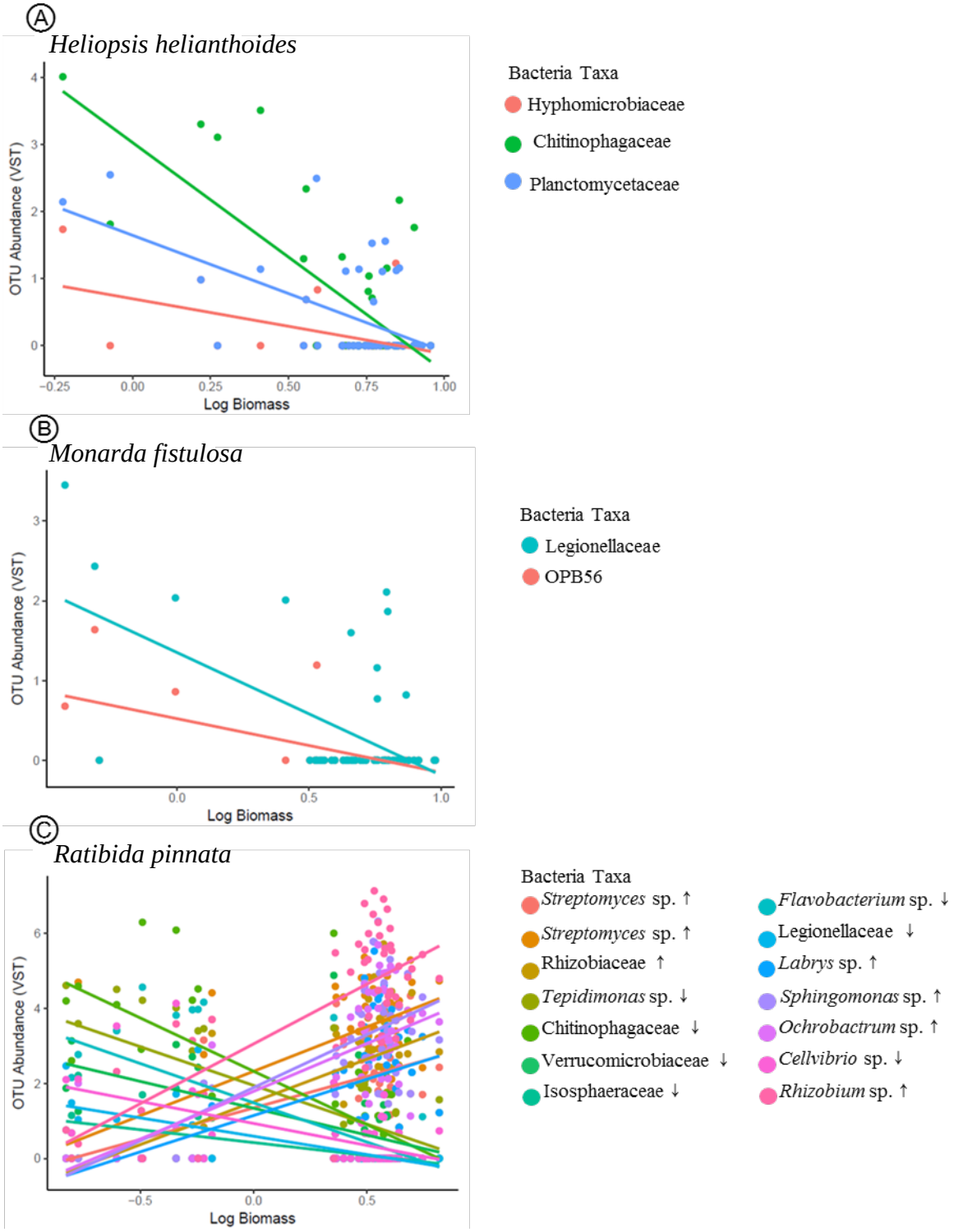


Figure 2-5: **Differences in total biomass is explained by certain bacteria taxa**. Correlations between log total biomass and abundance of OTUs which were significantly correlated for (a) *H. helianthoides* (b) *M. fistulosa* (c) *R. pinnata*.

2.5.1 RESILIENCE OF ROOT ENDOPHYTIC MICROBIAL COMMUNITIES DUE TO HOST CONTROL OF COMMUNITY COMPOSITION

To ensure we characterized the REB, we sequenced the samples collected after sonication and after washing and showed that the community composition was different (Figure 2-S6). Some studies have reported that the REB community is dominated by Proteobacteria [77-79, 81, 119-121] while others are dominated by Actinobacteria [75, 76]. Our results support the theory that plants have a core REB microbiome which is dominated by Proteobacteria. The dominating phyla across all REB communities in this study in decreasing order is Proteobacteria, Firmicutes, Bacteriodetes and Actinobacteria (Figure 2-S7) which have all been reported as dominant members of various REB communities [88]. The main difference in dominating phyla is due to comparisons with *Arabidopsis thaliana* which is a model organism used to understand plant genes and function but do not reflect the ecology of non-agricultural and agricultural plants due the absence of symbiotic relationships with arbuscular mycorrhizal fungi.

We chose to perturb the microbial communities by autoclaving soil, application of antibiotics and host presence in soil prior to collection (soil history). This allowed us to test the strength of deterministic factors that are thought to be responsible for structuring microbial communities. The composition of REB communities was perturbed but the turnover in composition was largely driven by plant hosts (Figure 2-S2, Table 2-S1). In both field and autoclaved field soil, the REB communities clustered based on plant host (Figure 2-1, Figure 2-S4b). Soil history explained only 1.9% of the variation in REB communities (Table 2-S1) suggesting that dispersal may be weak. There were only three bacterial taxa which were differentially abundant and these taxa did not cluster based on soil history (Figure 2-S8). This corroborates theories that microbial taxa may disperse over very short distances, creating non-

random distributions [95]. Other studies have shown that REB communities are very similar regardless of soil source [75, 76, 78]. It is known that fungal communities are also structured by plant host [42, 122, 123] suggesting that plant host structure microbial communities across different kingdoms and provides community resistance to disturbance.

To directly test the resilience nature of the REB communities, we perturbed the communities with antibiotics. We discovered that antibiotics did not affect plant fitness (Table 2-S2) but it did alter the diversity and composition of the bacterial communities (Table 2-S1). The two taxa that were differentially abundant (Figure 2-S9) are in the same phyla, Actinobacteria and family, *Conexibacteraceae*, which has not been intensively studied. *Conexibacteraceae* are known to reduce nitrates, live in high nitrogen environments and are sensitive to streptomycin [124-126]. We have provided evidence that not all strains in the *Conexibacteraceae* family are sensitive to streptomycin. One strain relative abundance increased in the presence of streptomycin while the other decreased. Even in an altered state, this did not weaken the deterministic factors structuring the REB communities. We can conclude that REB communities are resilient to perturbations due to the strong selection force from hosts.

Observational data provide evidence that the REB communities are not stochastically assembled. To directly test this hypothesis, we implemented a null model which created random assemblages of our data set. We used the RC metric because it provides information on the probability that pairs of communities are more similar (or different) than expected by chance [113]. For soils treated with and without antibiotics, the RC values approached 0 providing evidence that the communities are deterministically assembled and more similar than expected by chance (Figure 2-2). This provides direct evidence that endophytic bacterial communities are highly deterministic and the selection is driven by host identity.

2.5.2 THE LINK BETWEEN BELOW AND ABOVE GROUND SPECIES INTERACTION

One way for belowground species to influence plant community composition is by affecting fitness either indirectly or directly. We confirmed that changes in biomass can be attributed to the soil biotic components and not abiotic since biomass in sterile soils were uniform (Figure 2-S4a). Plant biomass was altered by the autoclaving of soil and soil history (Table 2-S2) demonstrating that composition of the soil biotic community could have altered biomass. Previous studies use plant fitness differences in autoclaved soils to approximate whether soils contain beneficial or inhibitory soil biota [66, 93, 127-129]. This could lead to changes in plant diversity through positive or negative feedbacks [100]. Positive feedbacks are when microbial composition increases relative performance of abundant plant species and negative feedbacks reduce relative performance of abundant plant species [100]. Feedbacks regulate coexistence of plant communities through direct feedbacks on conspecifics and indirect feedbacks of competing species [38].

C. nutans and *C. conyza*, biomass was not affected by the reduction of microbial species which indicates that our invasive and weedy plant, respectively, do not have an established relationship with the microbial communities (neither beneficial nor inhibitory) in the prairie system. As for *M. fistulosa*, *R. pinnata*, and *H. helianthoides* which are all native (non-weedy) species, there was a reduction in biomass when grown in autoclaved field soils indicating the potential beneficial relationship between the prairie plants and the microbial community. However, to appropriately test if the reduction of biomass in autoclaved soils could be explained by the divergence of REB communities, we correlated differences in biomass to turnover of REB communities. Not surprisingly, C. *nutans* was the only host which did not have a strong correlation between bacterial community and total biomass (Figure 2-4b). We did not measure

any differences in growth for *C. nutans* when grown in autoclaved soils; therefore, we did not expect there to be a significant correlation. For the three-native species, we identified a correlation between fitness and composition of REB communities. We identified taxa with higher abundance in the field soils compared to autoclaved field soils for *R. pinnata* (Figure 2-S10), *H. helianthoides* (Supplementary Figure 2-S11) and *M. fistulosa* (Figure 2-S12). We determined a core set of taxa which were found in all three natives (Figure 2-S13). Which further suggests that there is a core community of bacteria that can directly impact plant fitness. To directly test this hypothesis, we looked for correlations between individual bacterial taxon abundance and biomass. We demonstrated that there are several bacterial taxa that are correlated with biomass for *H. helianthoides* (Figure 2-5a), *M. fistulosa* (Figure 2-5b) and *R. pinnata* (Figure 2-5c). Two of the taxa that were enriched, *Ochrobactrum* sp. and *Sphingomonas* sp., have been identified as potential growth enhancing bacteria in previous experiments [130, 131]. Additionally, the depletion of certain OTUs belonging to the family Planctomycetaceae, Legionellaceae and Chitinophagaceae were consistent across plant species. This shows that there are bacteria that can be classified as potential growth inhibitors and could be used as a biocontrol for weeds or invasive species. Interestingly, each plant species responded differently to the abundance of bacterial species. Therefore, if individuals of the same plant species are competing for the same resources (bacterial species), then that could lead to negative feedbacks and restrict proliferation of conspecifics. It has already been shown with arbuscular mycorrhizal fungi [132]. To directly test this hypothesis for REB, future studies should focus on characterizing REB communities in the presence or absence of competition with conspecifics and heterospecifics.

2.6 CONCLUSIONS

Our study provides direct evidence that the endophytic root bacterial communities are assembled deterministically mainly driven by host and the strength of this relationship is unaffected by perturbations. Therefore, plant identity is a major determinant of root endophytic microbial communities. There is a strong correlation between plant growth and REB communities proving that the composition of the community is vital for plant growth, but that is not true for all plants. We conclude that there are many complex interactions between aboveground and belowground species that should be accounted for and not lumped into a black box. Our ability to study microbes in different niches will allow us to focus on vital species that will enhance our understanding of how to maintain or restore ecosystems.

2.7 ACKNOWLEDGEMENTS

The work presented in this chapter was done in collaboration with Scott Mangan, Claudia Stein, Jonathan Myers, Christopher Catano and Gautam Dantas. B.A. was supported by National Science Foundation graduate research fellow (award number DGE-1143945). Mike Dryer, Darlene Branson, Khadine Gomez, Max Herzog, Kevin Stiles and the undergraduates of the Mangan Lab helped with collecting data and managing the greenhouse experiment. This work was funded by NIH Director's New Innovator Award, the National Institute of Diabetes and Digestive and Kidney Diseases (DP2DK098089) and National Institute of General Medicine (R01GM099538) awarded to G.D.

2.8 SUPPLEMENTARY FIGURES & TABLES

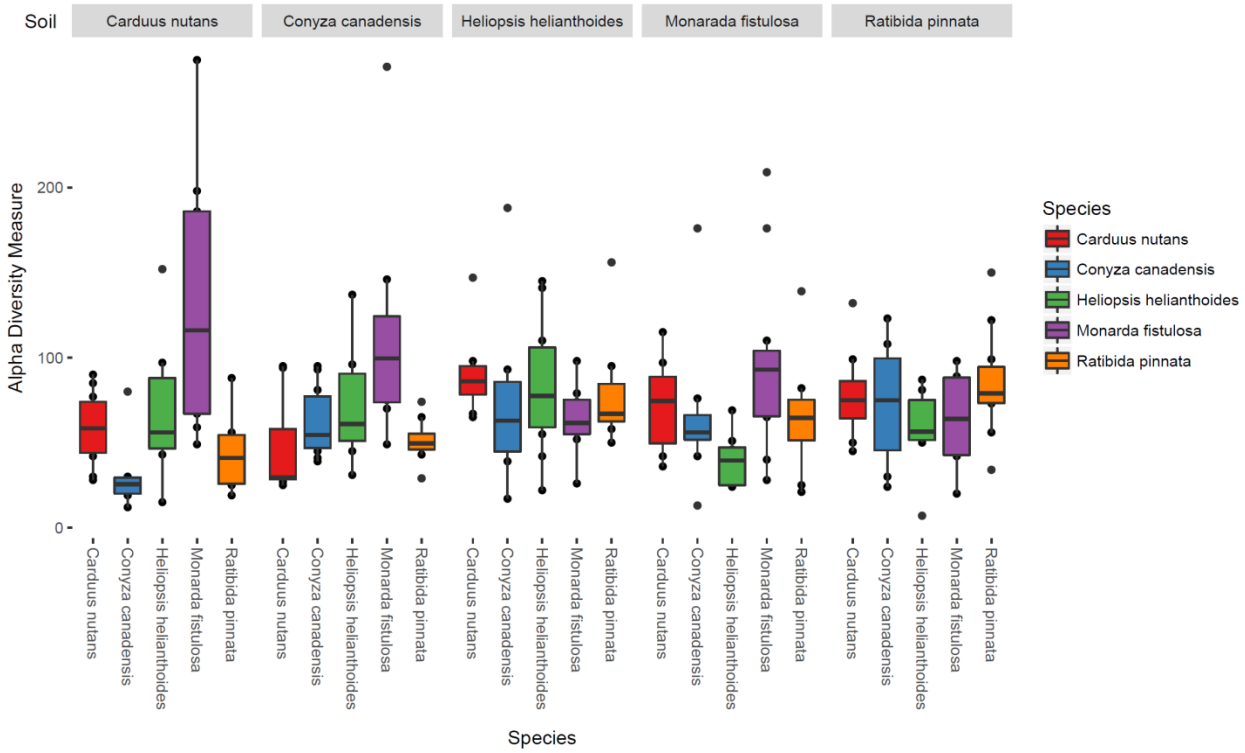


Figure 2-S1: Observed species richness is high between treatments. Observed species richness for each plant host in each soil trained by each plant host.

Table 2-S1: Statistical analysis conducted on Bray-Curtis of REB communities by different treatments. Condition = Autoclaved vs Field Soil. Treatment = Antibiotics vs No Antibiotics. Species = Host identity. Soil = Soil history.

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)	
Condition	1	5.932	5.9321	22.461	0.07836	0.001	***

Treatment	1	0.409	0.4093	1.5499	0.00541	0.026	*
Species	4	5.562	1.3904	5.2645	0.07347	0.001	***
Soil	4	1.467	0.3667	1.3884	0.01938	0.008	**
Residuals	236	62.329	0.2641		0.82338		
Total	246	75.699			1		

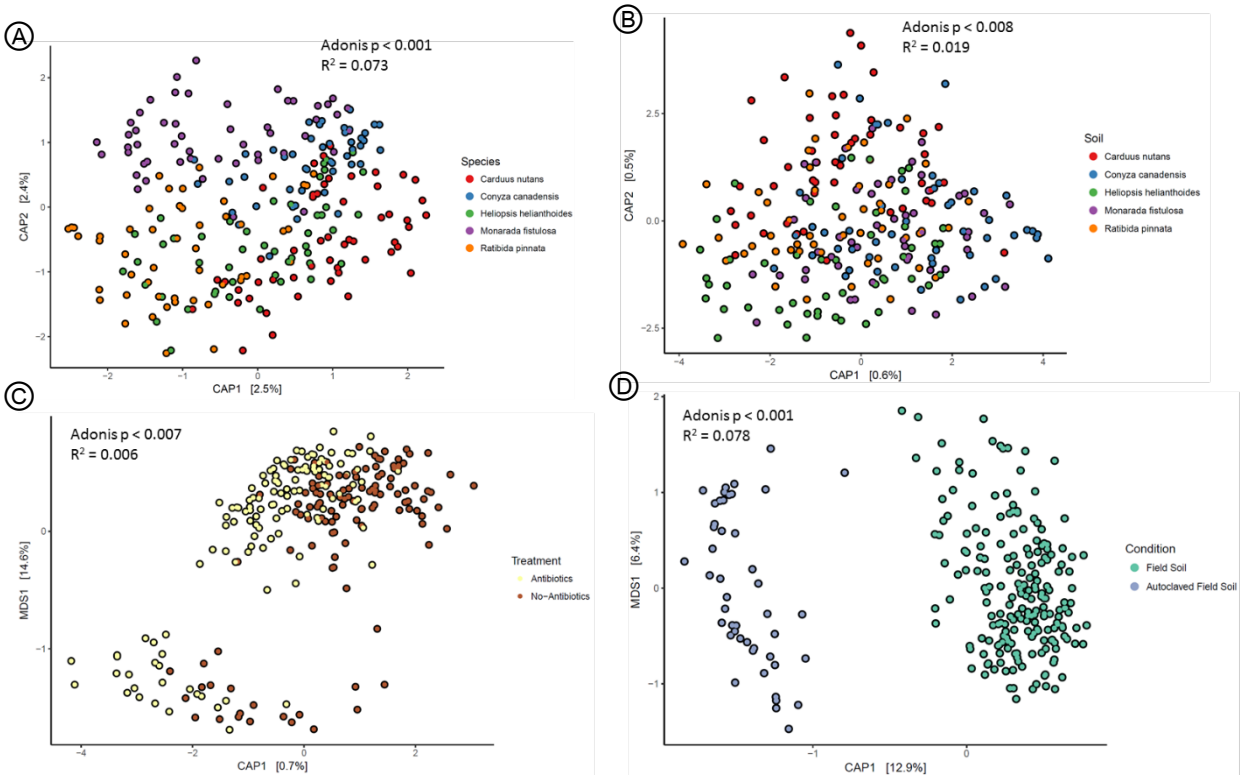


Figure 2-S2: Bacterial communities were altered by host, soil history, antibiotics and soil autoclave treatment. Ordination of Bray-Cutis dissimilarities shows clustering of root endophytic bacterial communities by (a) plant host (b) soil history (c) antibiotic treatment, (d) autoclaving treatment.

Table 2-S2: Statistical analysis of treatment effects on biomass. Full Anova model was used which condition was autoclaved field soil vs. field soil. Treatment was with antibiotics or without antibiotics. Species was the plant identity and soil was soil history or trained soil.

	Df	**Sum Sq**	**Mean Sq**	**F value**	**Pr(>F)**	
Condition	1	83.58	83.58	506.581	<2e-16	***
Treatment	1	0.07	0.07	0.402	0.5265	
Species	4	56.65	14.16	85.841	<2e-16	***
Soil	4	2	0.5	3.033	1.73E-02	*
Condition:Species	4	83.63	20.91	126.717	<2e-16	***
Condition:Species:Soil	36	9.86	0.27	1.659	0.011	*

Plot	5	1.27	0.25	1.542	0.1753
Species: Plot	16	1.25	0.08	0.475	0.9587
Residuals	465	76.72	0.16		

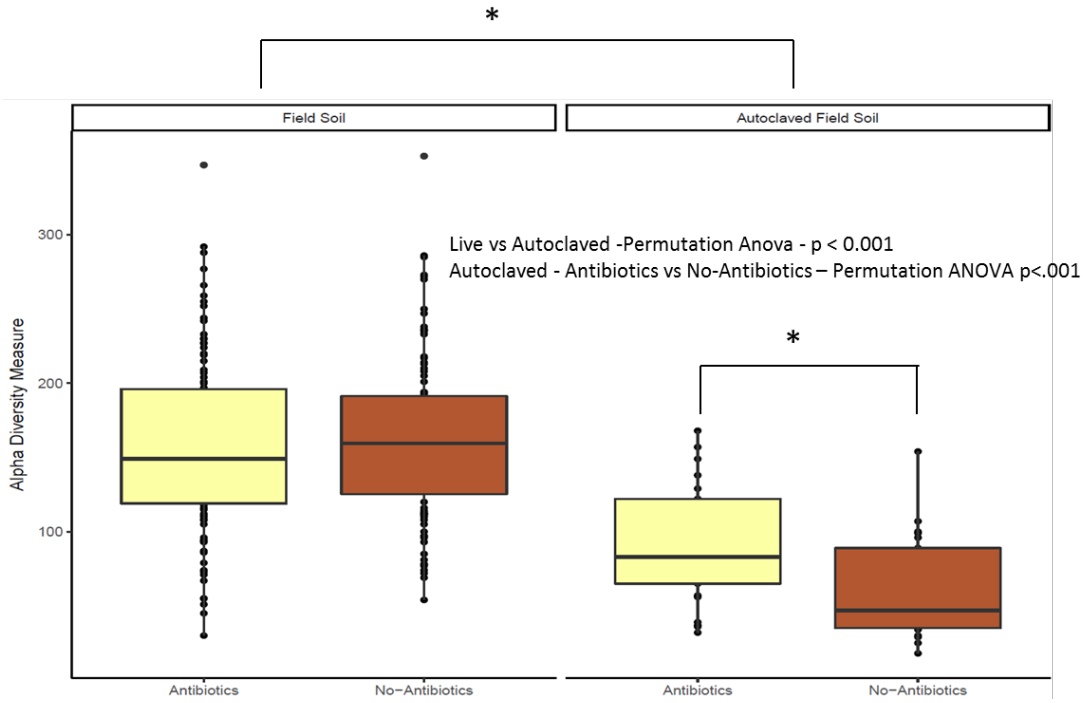


Figure 2-S3: Boxplots of observed bacterial species richness in soil treatments. Observed species richness is lower in autoclaved field soils but not affected by antibiotic

Table 2-S3: Statistical analysis of treatment effects on composition of REB in field soils. Treatment was with antibiotics or without antibiotics. Species was the plant identity and soil was soil history or trained soil.

	Df	**SumsOfSqs**	**MeanSqs**	**F.Model**	**R2**	**Pr(>F)**	
Species	4	2.8847	0.72116	6.3511	0.10987	0.001	***
Soil	4	0.9184	0.22959	2.0219	0.03498	0.001	***
Treatment	1	0.2127	0.21268	1.873	0.0081	0.002	**
Plot	5	1.0144	0.20289	1.7868	0.03864	0.001	***
Species:Plot	16	1.9205	0.12003	1.0571	0.07315	0.164	
Residuals	170	19.3036	0.11355		0.73526		
Total	200	26.2542			1		

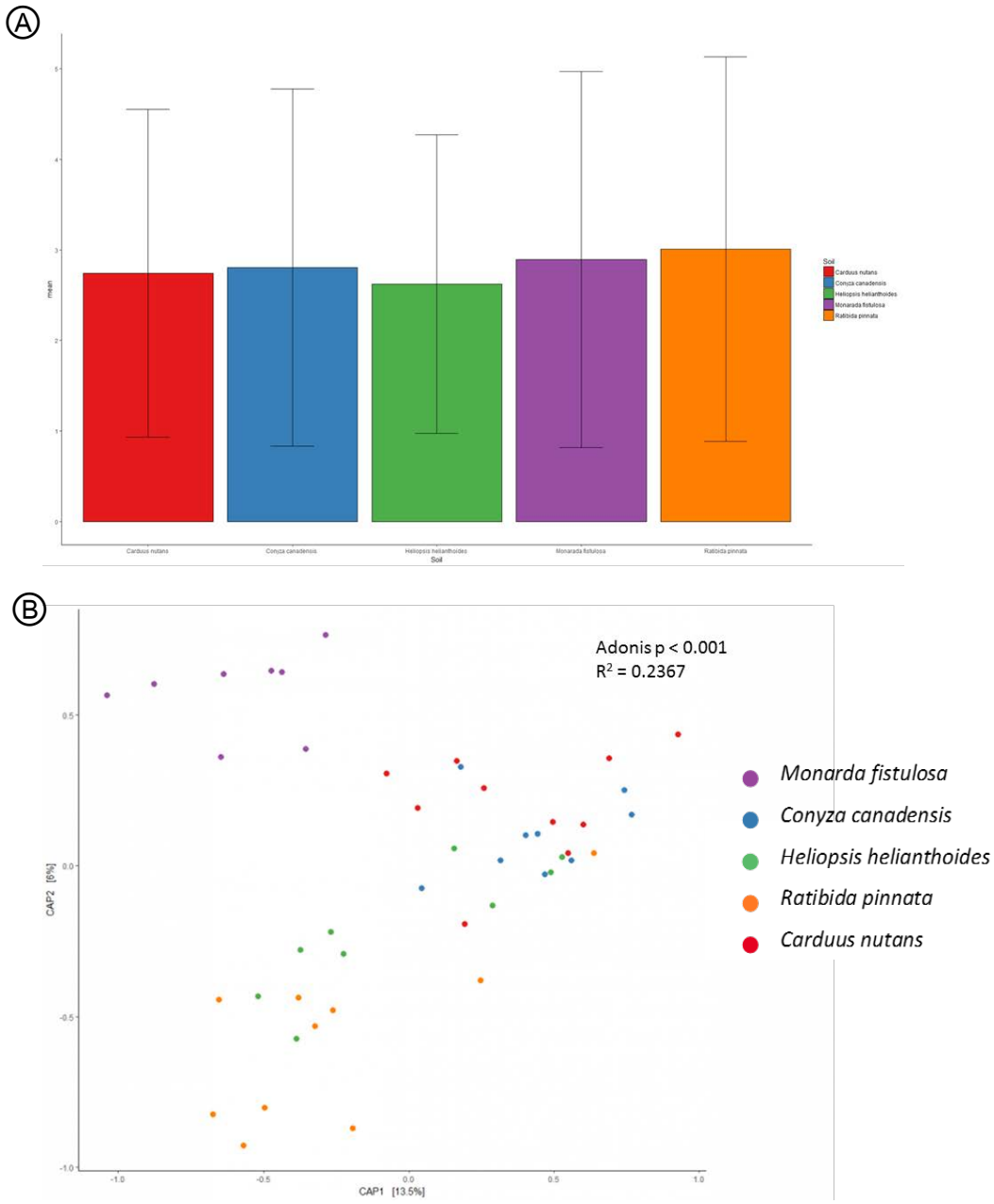


Figure 2-S4: Biomass and composition of microbial communities in autoclaved field soils. (a) Total biomass summed across soil autoclaved treatment. (b) Ordination of Bray-Cutis dissimilarities shows clustering of root endophytic bacterial communities by plant host for plants grown in sterile soils.

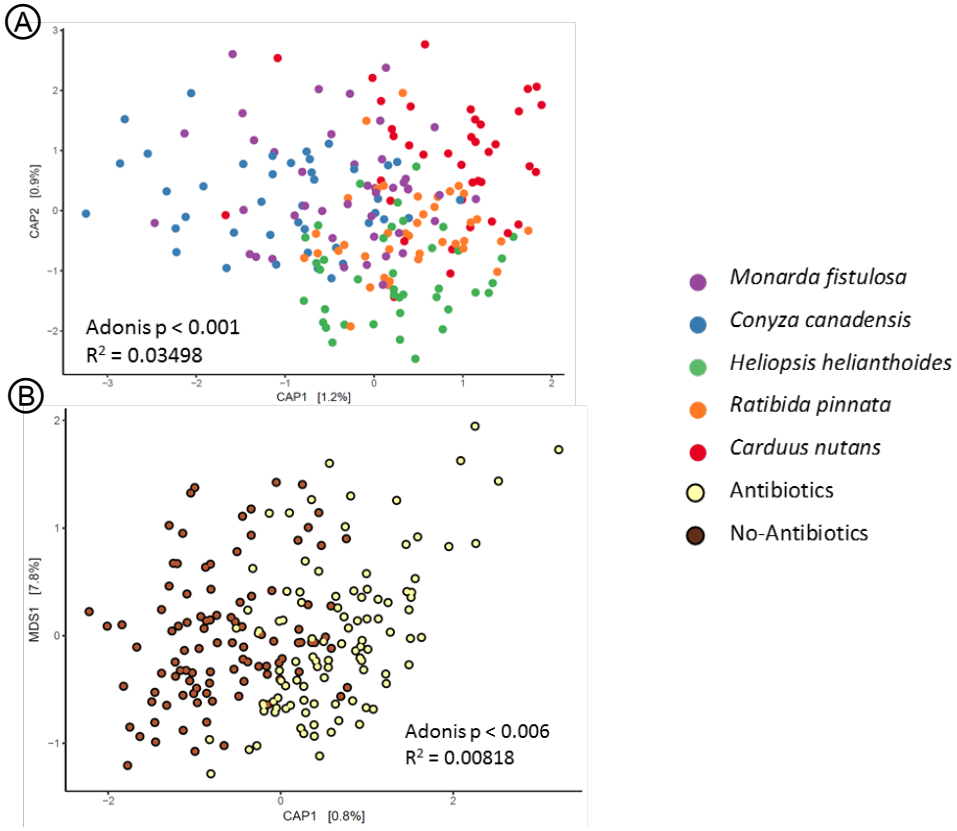


Figure 2-S5: Ordination of REB communities in field soils. Bray-Cutis dissimilarities shows clustering of root endophytic bacterial communities by (a) soil history, (b) antibiotic treatment.

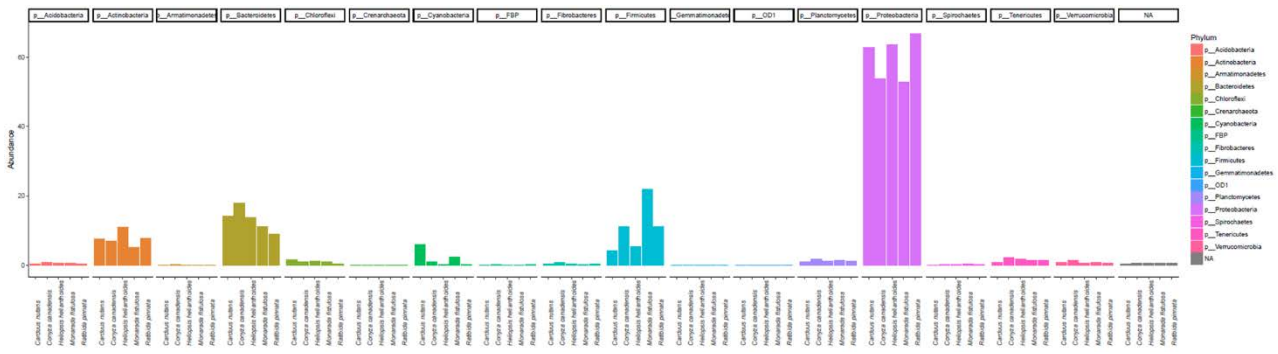


Figure 2-S8: Phyla abundance summed across all REB communities.

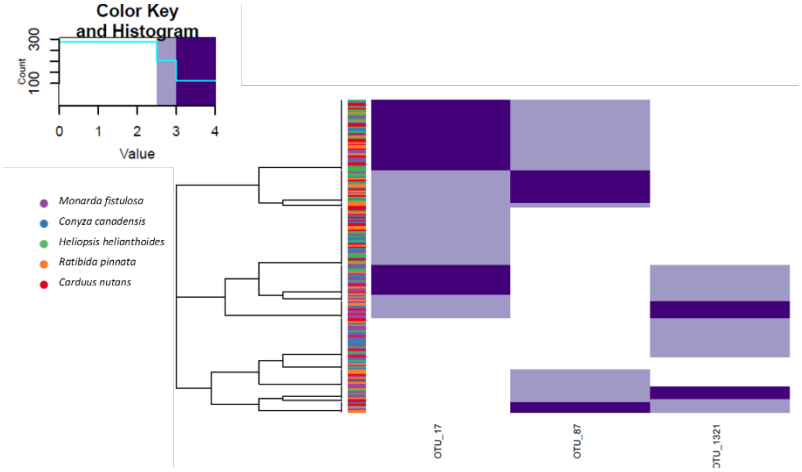
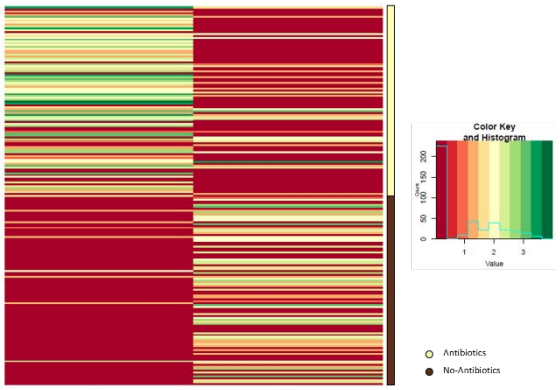


Figure 2-S9: Heatmap and dendrogram of OTUs differentially abundant in soils trained by plant hosts. There is no clustering based on soils.



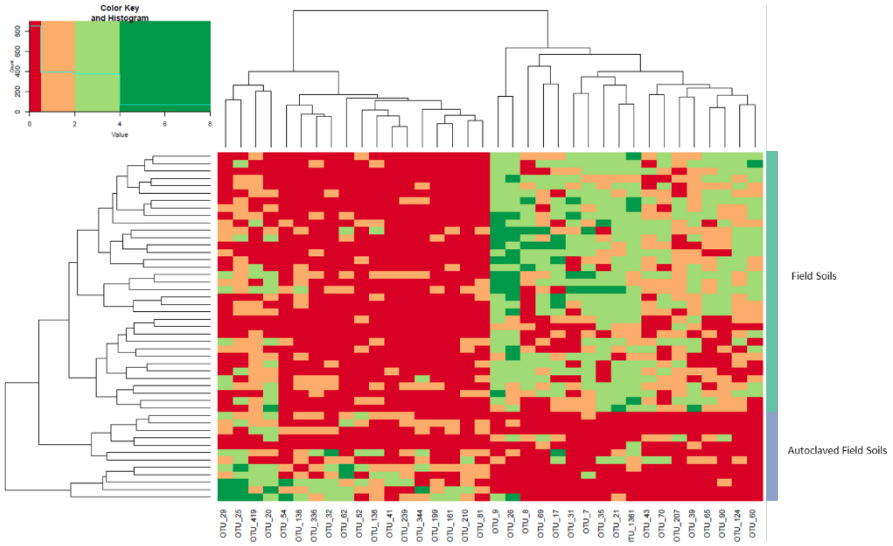


Figure 2-S10: Heatmap of OTUs differentially abundant in *H. helianthoides*' REB communities clustered by field vs autoclaved soil treatment.

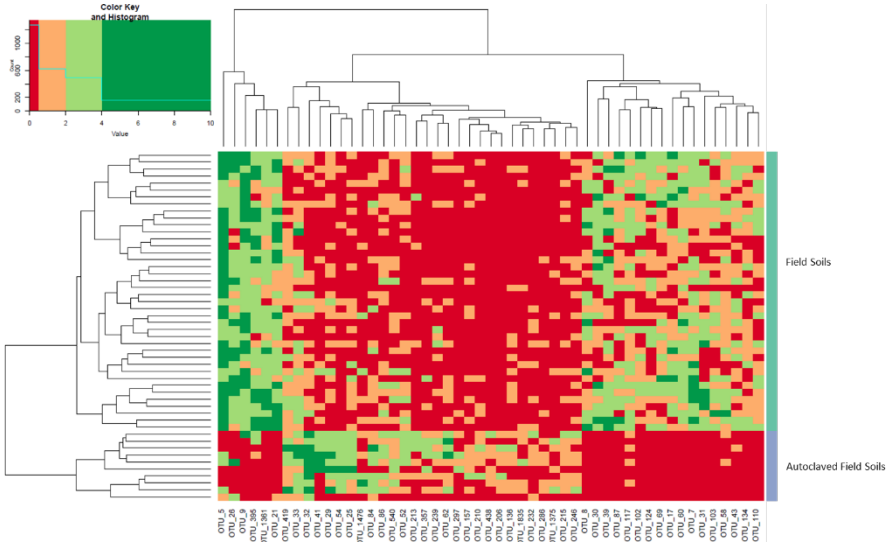


Figure 2-S11: Heatmap of OTUs differentially abundant in *R. pinnata*'s REB communities clustered by field vs autoclaved soil treatment.

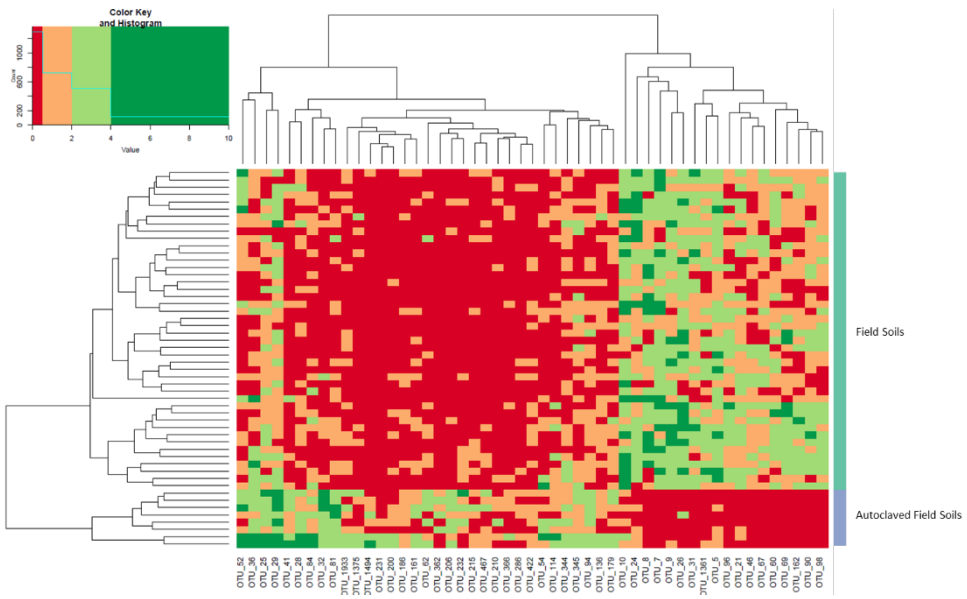


Figure 2-S12: Heatmap of OTUs differentially abundant in *M. fistulosa's* REB communities clustered by field vs autoclaved soil treatment.

Chapter 3 Evaluation of Culture-Based and Culture-Independent Methods to Characterize the Microbial Community of Human Urine

Boahemaa Adu-Oppong[1], Meghan Wallace[2], Carlos Eduardo Posada Perlaza[3], Alejandro Reyes Muñoz[1,3], Carey-Ann D. Burnham[2,4,5], Gautam Dantas[1,2,5,6*]

Affiliations:

[1]Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, Missouri, United States of America
[2]Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO, USA.
[3]Department of Biological Sciences, Universidad de los Andes, Bogotá, Colombia
[4]Department of Pediatrics, Washington University School of Medicine, St. Louis, MO, USA
[5]Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO, USA.
[6]Department of Biomedical Engineering, Washington University, St. Louis, MO, USA.

3.1 ABSTRACT

Urinary Tract Infections (UTIs) are one of the most pervasive urological disorders affecting millions yearly. Current clinical practices are focused primarily of understanding a single, easily-cultured pathogen, that is the most common cause of UTI-- uropathogenic *Escherichia coli*, while ignoring other bacteria (pathogens and commensals) that may not be culturable. We believe valuable, clinically-actionable information could be lost when this uncultured community is ignored. Additionally, improper use of antibiotics, the primary treatment for UTIs and generally targeted against *E. coli*, can lead to substantial selection pressure for the evolution of resistance in uropathogens and commensals. Our study evaluates the extent to which current clinical standards may be not detect key bacterial strains in the setting of UTI (cryptic pathogens

or commensal) which may be important to our understanding of UTI biology and treatment. We establish that the urinary microbiome in asymptomatic women is stable over time and a suspected UTI urine microbiome is compositionally and functionally different with an increase abundance of Proteobacteria and bacteriocin. We demonstrate that for 21% of positive UTI cases, the putative uropathogen identified through conventional diagnostic methods was not the most abundant species in the urine specimen but instead dominated by another known uropathogen. These results suggest the current methodology for classifying UTIs can be amended by the incorporation of next-generation sequencing methods. This will decrease diagnostic time and the risk of evolving antibiotic resistance.

3.2 INTRODUCTION

Urinary Tract Infections (UTIs) are one of the most pervasive urological disorders affecting millions yearly [133, 134]. UTI is a condition in which the urinary tract is colonized by pathogenic bacteria. The bacteria cause inflammation and travel to the bladder and the kidneys [135]. Unlike most bacterial diseases, incidence of UTIs is greater in women than in men. Over the span of a lifetime, women are 50 times more likely than men to contract a UTI [134]. Higher risk of UTIs begins at birth and continues until the age of 60 [136, 137]. Women are at increased risk due to anatomical differences such as a shorter urethra that can be easily colonized by normal vaginal flora [138, 139]. Translocation of external bacteria can happen during sexual intercourse or subsequent to medical interventions such as catheterization [134, 138, 139].

Antimicrobial therapy is the primary treatment for UTIs, but its efficacy is being challenged by increasing antimicrobial resistance (AMR) in UTI pathogens [140]. The growing prevalence of AMR bacterial pathogens has led to more frequent use of broad-spectrum antibiotics, which in

turn indirectly selects for increasingly multi-drug resistant (MDR) pathogens [141-146]. One of the major risks of developing MDR infections is previous exposure to antibiotics [147] and this is primarily because antibiotic usage can drastically change the environment of the urethra causing the structure of the bacterial community to change [148]. Additionally, the gut is frequently the source for organisms ultimately contributing to UTI [136], and there is an overall selective pressure for more resistance in gut-resident bacteria [149-151] due to ingested antibiotics. Therefore, antibiotic therapy may not be a sustainable form of treatment in the near future, as its continued use selects for an increase of MDR uropathogens, steadily decreasing the number of effective treatment options.

Numerous studies of the gut microbiota have established the fundamental role of bacterial community structure in regulating health [152]. It has been proposed that the structure of the gut microbiota is strongly correlated to the incidence of type 2 diabetes, stronger than host genotype. From basic science and translational perspectives, there is great interest in understanding if this strong microbiota-host health dynamic is restricted to the gut or if this correlation to disease occurs in other habitats of the human body as well.

Unfortunately, our understanding of the population structure of microorganisms in the urinary tract is limited [152]. The urinary tract was regarded as a sterile site for decades and only recently acknowledged as a body site that harbors microbes, as evidenced by culture-independent 16S rRNA gene sequencing [153-155]. Additionally, research on uropathogens has been largely focused on uropathogenic *E. coli* (UPEC). Approximately 80% of UTIs in the outpatient setting can be attributed to UPEC, while the remaining 20% can be attributed to other bacteria such as enterococci and staphylococci [140, 156]. However, these statistics are driven by culture dependent methods. Current diagnostic testing approaches include quantifying bacterial density

using culture, microscopy, and/or rapid dip stick or automated urinalysis for biochemical characterization of urine specimens; these methods are tuned for optimal detection of *E. coli*, and likely under-diagnose other uropathogens [135, 157]. Traditional culture methods commonly used for urinary tract pathogens are not suitable for growth of fastidious organisms and/or anaerobic bacteria, and these may be underappreciated as urinary tract pathogens. Unbiased, sequence-based approaches to query urine samples for pathogens would be one way of addressing this question, with the potential to inform improved urine culture methods in the future.

The handful of studies that have performed 16S rRNA gene sequencing on male and female urine samples have reported that an array of bacterial taxa reside in the urinary tract of asymptomatic individuals, such as *Lactobacillus*, *Prevotella*, *Gardnerella*, *Streptococcus* and even *Staphylococcus* species. Some of these bacterial taxa discovered in asymptomatic patient urines have been recognized as potential uropathogens (e.g. *Streptococcus* and *Staphylococcus*) [152]. Therefore, the current antimicrobial therapy for UTI, treatment, based on reviewing the antimicrobial profile of a single uropathogenic bacterium (generally UPEC) may not be appropriate to treat many patients [158]. For example, in a study of 32 suspected UTI samples, 34.4% were determined to be caused by two or three etiological agents [159]. Additionally, as a consequence of considering urine from asymptomatic individuals to be sterile, most UTI studies have focused on solely studying urine samples from infected patients, neglecting to consider the urine microbial composition of asymptomatic patients [160].

We studied the ecology of the urine microbial community in symptomatic and asymptomatic individuals through a combination of culture-dependent and culture-independent next-generation sequencing (NGS)-based methods. In the context of the urine microbiome, we will

characterize the abundance and diversity of known and potentially new uropathogens, their antimicrobial resistance determinants, and their virulence genes. We hypothesize that UTIs should be studied from a bacterial community perspective rather than simply as individual isolates to 1) more accurately identify the etiologic agent(s), and 2) to understand the ecology of the disease by quantifying the community state and dynamics between bacterial species during an infected versus an asymptomatic state. We demonstrate that NGS recapitulated the current standard of care culture based techniques in the state of an infection, highlighting the potential function of NGS in UTI diagnostics.

3.3 MATERIALS AND METHODS

3.3.1 SAMPLE COLLECTION

All suspected UTI patient samples were de-identified and collected from frozen remnant urine specimens and approved by the Human Research Protection Office (approval number 201401115). The samples were submitted to the Barnes-Jewish Hospital/Washington University School of Medicine in Saint Louis, Missouri, United States as part of routine clinical care. In total we selected 162 specimens to use in this study. Of the 152 specimens, 52 specimens have significant growth of one or two uropathogens and classify as positive cultures [161], 71 had no bacterial growth and 20 had less than 10^{-5} CFU/ml of bacterial growth. The remaining 9 specimen have more than 3 bacterial species growing in concentrations above threshold in the standard-of-care clinical routine and classify as contaminated cultures [161].

3.3.2 SAMPLE PRCESSING AND PHENOTYPING

Only the positive specimens were subjected to all culturing protocol. Prior to the culturing protocol, all samples were initially cultured using standard of care methods: 1 uL was plated to each of a BAP and MAC using a 1 uL calibrated loop and incubated at 35C in CO2 for

24 hours. Our culturing protocol consisted of two different methods. The first method, conducted on asymptomatic and positive specimen, consisted of collecting all of the colonies that are grown on MacConkey and sheep's blood agar plates (Hardy Diagnostics) in combination as a "slurry" for DNA extraction using the BiOstic Bacteremia DNA Isolation Kit (Mo-Bio). For the asymptomatic, we did enhanced culturing. The second method, most similar to current clinical methods [161], involve picking four individual colonies per bacterial species from dilutions of the urine sample plated on MacConkey and sheep's blood agar plates and separately extracting gDNA using the BiOstic Bacteremia DNA Isolation Kit (Mo-Bio). Where appropriate, antimicrobial susceptibility testing was performed using Kirby Bauer disk diffusion testing performed and interpreted in accordance with Clinical and Laboratory Standards Institute (CLSI) guidelines [162]. We assayed enteric gram-negative bacteria for susceptibility to nitrofurantoin, cefazolin, cefotetan, ceftazidime, cefepime, ciprofloxacin, trimethoprim-sulfamethoxazole and ceftriaxone. Prior to whole genome sequencing, species identity of isolates was determined with VITEK MALDI-TOF MS v2.0 knowledgebase (bioMerieux) as previously described [163, 164]. For all urine samples, 2ml was used to isolate metagenomic DNA using the BiOstic Bacteremia DNA Isolation Kit (Mo-Bio).

3.3.3 Illumina Library Preparation

Sequencing libraries were prepared from 15 ng – 500 ng of total DNA from each slurry, isolate, and urine sample. DNA was sheered to a target size range of approximately 500-600 bp using the Covaris E220 sonicator with the following settings: peak incident power, 140; duty cycle, 10%, cycles per burst 200; treatment time 75 seconds; temperature 7^{o}C; sample volume 130 µl. Sheared DNA was purified and concentrated using MinElute PCR Purification Kit (Qiagen), eluting in 20 µl pre-warmed nuclease-free H_2O per barcode. Purified sheared DNA

was then end-repaired and Illumina adapters were ligated using the following protocol: A 25 µl reaction volume was prepared containing 20 µl of purified sheered DNA, 2.5 µl T4 DNA ligase buffer with 10mM ATP (10X, New England BioLabs), 1 µl dNTPs (1mM, New England BioLabs), 0.5 µl T4 polymerase (3 U μl^{-1}, New England BioLabs), 0.5 µl T4 PNK (10 U ul^{-1}, New England BioLabs), and 0.5 µl Taq Polymerase (5 U ul^{-1}, New England BioLabs). The reactions were incubated at 25^{o}C for 30 min followed by 20 min at 75^{o}C.

For the barcode mix forward and reverse sequencing adapters were stored in TES buffer (10mM Tris, 1mM EDTA, 50 mM NaCl, pH 8.0) and annealed by heating the 1mM mixture to 95^{o}C followed by slow cooling (0.1 ^{o}C per second) to a final holding temperature of 4^{o}C. A 2.5 µl volume of prepared barcode mix and 0.8 µl of T4 DNA ligase (New England BioLabs) were added to each end-repair reaction and the reaction was incubated at 16^{O}C for 40 min followed by 10 min at 65^{o}C.

Reactions were purified using a MinEluted PCR Purification Kit (Qiagen), eluting in 16 µl pre-warmed elution buffer (Qiagen). The adaptor-ligated, sheered DNA was then size-selected to a target range of 400-900 bp on a 1.5% agarose gel in 0.5X Tris-Borate-EDTA (TBE), stained with GelGreen dye (Biotium) and enriched using the following protocol: A 25 µl reaction volume was prepared containing 2 µl of purified DNA, 12.5 µl 2x Phusion HF Master Mix (New England BioLabs), 1 µl of 10 MM Illumina PCR Primer Mix (5'- AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T -3' and 5'- CAA GCA GAA GAC GGC ATA CGA GAT CGG TCT CGG CAT TCC TGC TGA ACC GCT CTT CCG ATC T -3') and 9.5 µl of nuclease-free H_2O. The PCR cycle temperatures were as follows: 98 ^{o}C for 30s, then 18 cycles of [98 ^{o}C for 10 s, 65 ^{o}C for 30 s, 72 ^{o}C for 30s], then 72 ^{o}C for 5 min.

Amplified DNA was then size-selected to a target range of 500bp on a 1.5% agarose gel in 0.5X TBE, stained with GelGreen dye (Biotium) and purified using a MinEluted PCR Purification Kit (Qiagen), eluting in 15 µl of elution buffer (Qiagen). The purified DNA was measured using the Qubit fluorometer HS assay kit (Life Technologies) and 10nM of each sample were pooled for sequencing. Subsequently samples were submitted for paired-end 150bp sequencing using the Illumina NextSeq-High platform at CGS (Center for Genome Sciences & Systems Biology at Washington University in St. Louis).

Prior to all downstream analysis, Illumina paired-end shotgun metagenomics sequence reads were binned by barcode (exact match of first 7bp), quality filtered using Trimmomatic v0.3.0[165] (*java -Xms1024m -Xmx1024m -jar trimmomatic-0.33.jar PE -phred 33 ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10 LEADING:10 TRAILING:10 SLIDINGWINDOW:4:20 MINLEN:60*) and human DNA was removed from the slurry and urine samples using DeconSeq [166] using build 38 of the human genome using default parameters. The other 54 sequencing libraries were prepared using the Illumina Nextera XT [167] method and submitted for paired-end 150pb sequencing using the Illumina NextSeq-High platform at CGS.

3.3.4 De novo Assembly

De novo assembly of reads for each isolate genomic DNA was completed using VelvetOptimiser (http://bioinformatics.net.au/software.velvetoptimiser.shtml) (*VelvetOptimiser.pl -s 45 -e 91 -t 1 –optFuncKmer 'n50'*). Optimal assembly was determined by n50.

3.3.5 Reference Based Assembly

The best reference sequence was chosen for each isolate by mapping 10,000 reads chosen randomly from that isolate against all reference genomes (from NCBI Genome downloaded April 19, 2016) of the same species as predicted by MALDI-TOF and reconfirmed with MetaPhlAn 2.0 [168] (*metaphlan2.py <forward_paired_reads>,<reverse_paired_reads>,<unpaired_reads> --mpa_pkl mpa_v20_m200.pkl --bowtie2db mpa_v20_m200 --bowtie2out <output_bowtie2_file> --nproc 5 --input_type fastq > <output_file>*). If there was not a consensus between the prediction from MALDI-TOF and MetaPhlAn, that isolate was removed from the analysis. Reads were mapped using SNAP 1.0beta1.8 [169] (*snap paired <index> <forward reads> <reverse reads> -t 1*). The genome against which the highest percentage of reads mapped was used as the reference sequence for that assembly. All reads were mapped to the reference sequence (*bowtie2 -x <reference index> -1 <forward_paired_reads> -2 <reverse_paired_reads> -U <unpaired_reads> -q –phred33 –very-sensitive-local -I 200 -X 1000 -S <sam_file_output> -2> <bowtie2_log_file>*). Variants from the reference were called using samtools (*samtools view -buS <sam_file_output> |samtools sort -m 4000000000 -o <bam_output>*) (*samtools index <bam_output>*) (*samtools mpileup -Ud -f <reference_fasta_file> <bam_output_sorted> > <mpileup_output>*) (*bcftools call –variants-only -O b -c -o <mpileup_bcf_output> <mpileup_output>*) (*bcftools view -O v -o <mpileup_vcf_output> <mpileup_bcf_output>*). The variant call format file was then filtered to remove SNPs with coverage greater than twice the average coverage expected per base (*vcftools –vcf <mpileup_vcf> --max-meanDP 2 --recode --out <filtered_mpileup>*).

3.3.6 COMBINATION OF DE NOVO AND REFERENCE BASED ASSEMBLY

After both assemblies were completed, then another *de novo* assembly was performed using the data from both assemblies. The contigs from the *de novo* assembly and reference mapping were put in an additional velvet assembly step as long reads with the original reads files. We determined hash value based on the optimized velvet assembly hash value (*velveth <output_directory> <hash_value> -fastq -short <unpaired_reads> -fastq -shortPaired <forward_paired_reads> <reverse_paired_reads> -fasta long <contigs_from_ref_bases_assembly> <contigs_from_denovo_assembly>*) (*velvetg <output_directory> -ins_length 400 -clean yes -conserveLong yes -scaffolding yes -long_mult_cutoff 0*). Finally, all contigs were ordered using ABACAS[170] to the reference genome (*abacas.1.3.1.pl -r <reference_genome> -q <contigs> -p 'nucmer' m -b -o <output_contig_file>*). Pilon[171] was used to improve the draft assembly by filling gaps and identifying local misassemblies (*java -jar pilon-1.13.jar –genome <contig_file> --frags <bam_file> --output <pilon_contig_output>*).

3.3.7 Open Reading Frame Prediction and Annotation

Open reading frame prediction for each genome was performed separately using Prokka [172] . Each open reading frame was compared to five databases ResFams, Pfam and TIGRFAMs using ResFams [173] (*annotate_functional_selections.py -proteins <protein_fasta_file> --resfams -o <output_directory>*), the Comprehensive Antibiotic Resistance Database [174], and an in house curated virulence HMM database. All annotations were combined and the annotation with the highest bit score and lowest e-value were assigned to the open reading frame.

3.3.8 Metagenome Assembly

Iterative De Bruijn Graph De Novo Assembler for Short Reads Sequencing data with Highly Uneven Sequencing Depth (IDBA-UD) [175] was used to assemble metagenomes using quality filtered reads with all human reads removed (methods described above).

3.4 RESULTS

3.4.1 DETERMINING STABILITY OF ASYMPTOMATIC URINE MICROBIOME

The control group consisted of clean-catch urine samples collected at three different time points from 10 asymptomatic adult women volunteers who had no evidence of an UTI during the time of collection and had not taken antibiotics in the 14 days prior to collection. Within our control group, we evaluated intraindividual variation (within subject) and interindividual variation (between patients) of the cultured slurries and directly sequenced urine microbiome compositions to determine if the urine microbiome is stable over time. We studied the microbiome via shotgun sequencing and determined taxonomy by MetaPhlAn2 [176]. We noted

higher similarity in intra-variation for both cultured slurries and direct sequenced urine microbiome indicating that the microbiome is individual-specific; therefore, we can average the urine microbiome composition to compare against the samples from suspected UTI patients (Cultured Slurries: P =1.2 x 10^{-13}, Direct Sequenced Urine: P=1.6 x 10^{-6}, Figure 3-1a).

3.4.2 Asymptomatic urine microbiome is compositionally different from suspected UTI microbiome

Banked remnant urine samples from suspected UTI patients (n=152, Table 3-S1) were

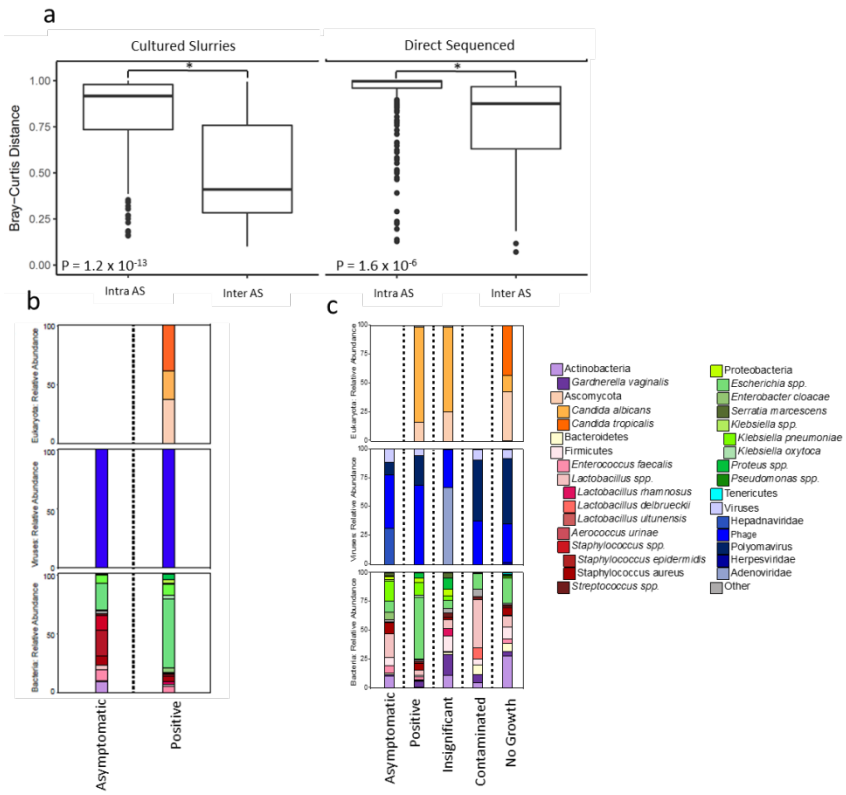


Figure 3-1. Diversity of Urine Microbiome over time and between clinical classifications. a, Box plot quantifying difference in microbial composition over time for the same patient (intra AS) and across patients (inter AS) for the cultured slurries and direct sequenced urine microbiome. b, Bar chart depicting the relative abundance of eukaryotes, bacteria and viruses present in the cultured urine microbiome of asymptomatic (n=10) and positive (n=47). c, Bar chart depicting the relative abundance of eukaryotes, bacteria and viruses present in the uncultured urine microbiome of asymptomatic (n=10), positive (n=48), insignificant (n=17), contaminated (n=7) and no growth (n=61). All P values were calculated using the permutation ANOVA.

classified into one of four categories based on current standard-of-care clinical procedures: (1) *positive*, if the specimen had significant growth ($\geq$100,000 CFU/mL) of one or two uropathogens (n=52), (2) *contaminated*, more than 3 bacterial species growing in concentrations above threshold in standard-of-care clinical culture (10^5) (n=9), (3) *insignificant* (<100,000 CFU/mL during standard-of-care culture) (n=20), and (4) *no growth*, specimen had no visible signs of bacterial or fungal growth during culturing (n=71) (Table 3-S2).

The microbial communities of asymptomatic cultured slurries were enriched with Firmicutes while clinically classified positive specimens were enriched with Proteobacteria (Figure 3-1b). This relationship was reinforced within the directly sequenced urine specimens; however, there was a plethora of viruses and Actinobacteria which were discovered in all suspected UTI specimen (Figure 3-1c). Stratification by gender and race was not significant (Gender: $P = 1.04$, Race: $P = 0.36$, Table 3-S3); however, since nearly all of our asymptomatic volunteers were Caucasian females we compared those samples against Caucasian females from the other cohorts. Examination of the principal component axes of variation in cultured slurries and direct sequenced urine microbiomes showed that asymptomatic specimen segregated from suspected UTI specimen regardless of whether we compared only the Caucasian females or the entire cohort (Caucasian Female Only Cultured Slurries: $P < 0.001$, Figure 3-2a; Caucasian Female Only Direct Sequenced Urine: $P < 0.001$, Figure 3-2b; Entire Cohort Cultured Slurries: $P < 0.001$, Figure 3-2c; Entire Cohort Direct Sequenced Urine: $P < 0.001$, Figure 3-2d). To determine the taxa which were differentially abundant between the categories, we used ANCOM [177]. *Escherichia coli* dominate cultured slurries positive specimens and *Staphylococcus epidermidis, S. haemolyticus, S. hominis, S. lugdunensis*, *Streptococcus anginosus*, *Corynebacterium auriucosum*, *C. sp HFH0082*, *Lactobacillus crispatus*, and *L. jensenii* in the

cultured slurry asymptomatic specimens (Figure 3-S1, FDR-adjusted $P < 0.05$). For the insignificant specimens, *E. coli* remained differentially abundant in positive specimen. *Lactobacillus crispatus* and *L. iners* was abundant in asymptomatic and contaminated specimens (Figure 3-S2, FDR-adjusted P < 0.05). *Propionibacterium acnes* dominated no growth specimens while *Ruminococcus torques* were exclusively found in contaminated and insignificant specimens (Figure 3-S2, FDR-adjusted $P < 0.05$).

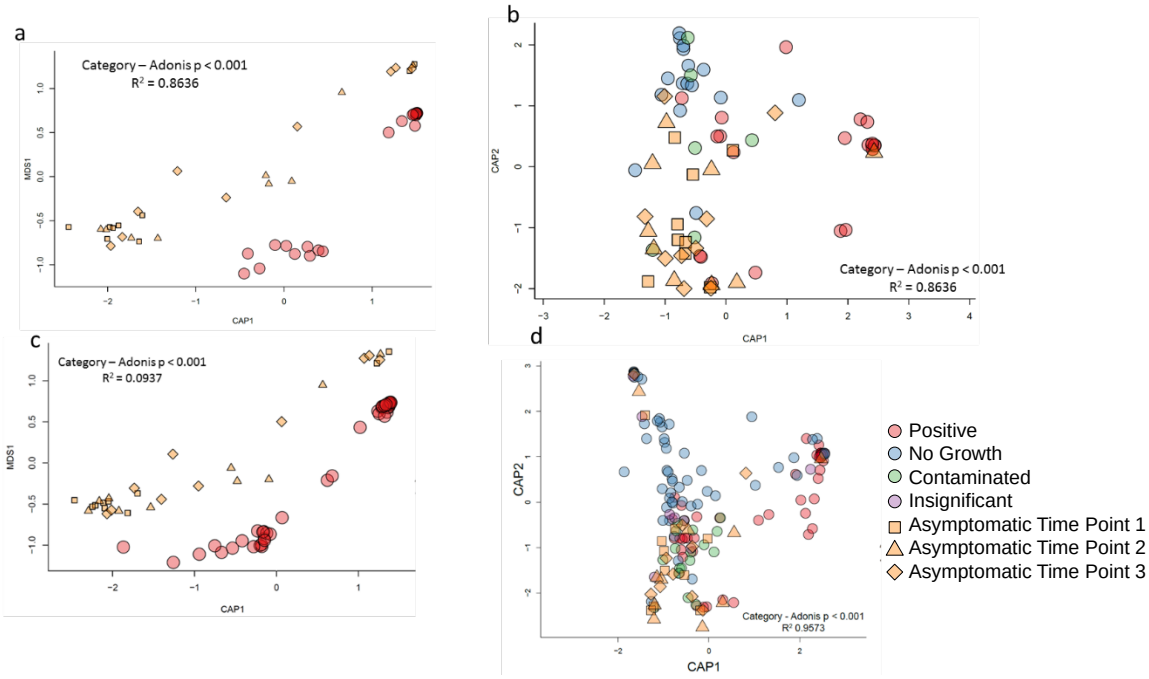


Figure 3-2. Urine microbiome alternative states between clinical classifications. Canonical analysis of principal coordinates: **a,** the cultured microbiome between only asymptomatic samples from Caucasian females (n=10) and positive urine samples from Caucasian females (n=23). **b**, the cultured microbiome between positive (n=10) and asymptomatic samples (n=47). **c**, the uncultured microbiome between only asymptomatic(n=10), positive (n=22), insignificant (n=6), and no growth (n=16) samples from Caucasian females. **c**, the uncultured microbiome between asymptomatic (n=10), positive (n=48), insignificant (n=17), contaminated (n=7), and no growth (n=51) samples. All *P* values were calculated using the ADONIS and pairwise ADONIS was corrected with false discovery rate (FDR).

Stratification by age when assigned into age groups (A – 19-49, n=88; B – 50 – 69, n= 37; C – 70+, n = 28) was significant (Figure 3-3, *P*= 0.034). *Gardnerella vaginalis* was differentially abundant in group A and B relative to group C. Group A and B had a lower overall proportion of Proteobacteria when compared to Group C (Figure 3-3a, *P* =0.03, *P* = 0.004). When stratified by clinical classifications and age group, within the no growth samples group A

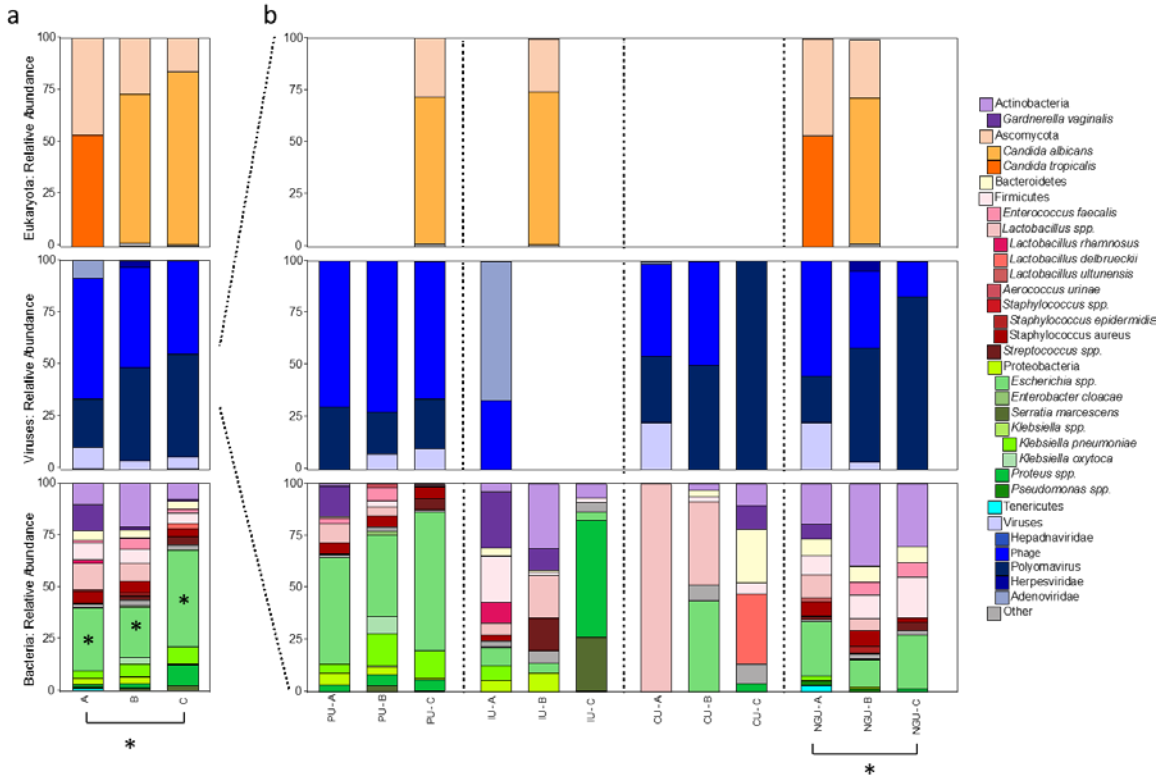


Figure 3-3. Urine microbiome of putative UTI patients clustered by age. Bar plots depicting the relative abundance of eukaryotes, bacteria and viruses present in the uncultured urine microbiome **a**, stratified by age group, A – ages 19-49 (n = 88), B – ages 50 – 69 (n = 37), and C – age 70+ (n = 28). **b,** stratified by age group and clinical classification Positive (P) A (n = 18), B (n = 14), C (n = 17), Insignificant A (n = 10), B (n = 4), C (n = 3), Contaminated A (n= 2), B (n = 2), C (n = 2), No Growth A (n = 28), B (n= 17), C (n = 6). All P values were calculated using the ADONIS and pairwise ADONIS was corrected with FDR.

was compositionally different from group B (Figure 3-3b, *P* = 0.048 A - n=28, B - n=17, C -

We were able to identify metabolic pathways which could be enriched or depleted in asymptomatic individuals using HUMAnN [178]. Bacteriocins were enriched while protein SopB were depleted when comparing the cultured slurries (Table 3-1).

Table 3-1. Cultured slurry metabolic pathways significantly enriched in asymptomatic and positive specimen. Means and p-values of metabolic pathways were calculated using HUMAnN.

	Asymptomatic Mean	Positive Mean	p-value	q-value
Bacteriocin-type signal sequence	96.17	0	8.699e-05	0.002523
Protein SopB	0	42.23	2.803e-05	0.00117

3.4.3 URINE MICROBIOME IS REPRESENTATIVE OF CULTURED SLURRY MICROBIOME FOR SPECIMEN CLASSIFIED AS POSITIVE

Whole-metagenome sequencing of both the cultured slurries and direct sequenced urine allows us the ability to investigate the similarity in compositions across methods. The alpha diversity of the asymptomatic cultured slurry was significantly higher when compared to positive cultured slurry (Figure 3-4a, $P = 2.11 \times 10^{-19}$). However, the alpha diversity was the same in the asymptomatic and positive direct sequenced urine (Figure 3-4b). In order to compare composition across methods, the reads in the cultured slurries were assembled using IDBA-UD [175]. Then reads from the direct sequenced urines were mapped to contigs from the cultured slurries using bowtie2 [179]. The percentage of reads aligned was higher in positive specimens when compared to asymptomatic specimens (Fig 3-4c, $P = 2.2. \times 10^{-16}$) which supports the current methodology for identifying potential uropathogens. The high similarity between direct sequenced urine and cultured slurries is noteworthy since it supports the notion that the number of days it takes to classify a UTI can be decreased by directly sequencing the urine.

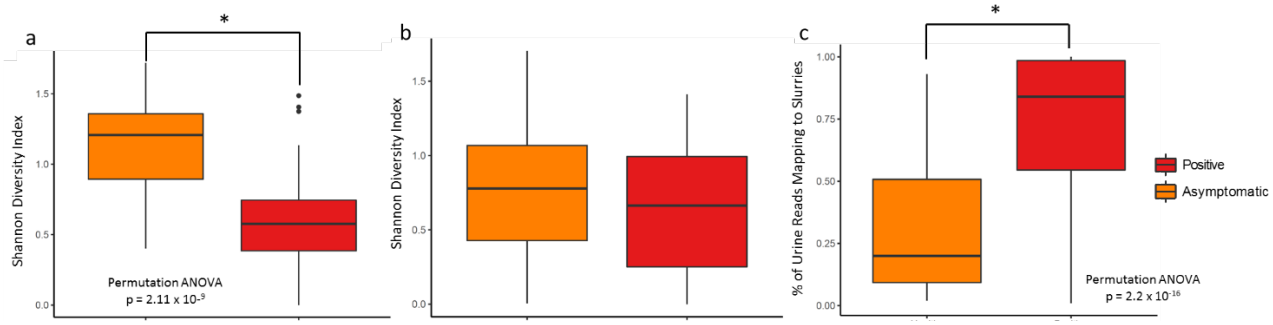


Figure 3-4. Direct sequenced urines are representative of cultured slurry for positive but not asymptomatic urine specimen. a. Boxplots of Shannon diversity index for cultured slurry. b. Boxplots of Shannon diversity index for direct sequenced urine. c. Boxplots of percentage of urine reads which mapped to contigs assembled from cultured slurry reads.

3.4.4 POPULATION STRUCTURE OF PUTATIVE UROPATHOGENIC *ESCHERICHIA COLI*

To investigate clonality and similarity of putative uropathogenic *E. coli*, individual colonies were picked from agar plates and subjected to MALDI-TOF MS for organism identification, Kirby-Bauer Disk Diffusion for antimicrobial resistance detection, and sequenced. Sequencing reads were assembled using both de novo and reference based assembly. Single-nucleotide polymorphisms (SNPs) were identified using kSNP [180] with our cohort of isolates alongside previously sequenced *E. coli* from various clades and pathotypes. A core SNP alignment parsimony tree was used to infer clonality. Multilocus sequence typing (MLST) was identified by mapping contigs to a PubMLST typing schemes (https://github.com/tseemann/mlst). Isolates classified into clades A (6.67%), B1 (10%), B2 (62.5%), D (16.6%), and F (4.2%) (Figure 3-5). Isolates classified as ST-648 from 2 patients clustered into a newly diverging clade in F (Figure 3-5). This indicates that putative uropathogens are evolving in other clades.

Isolates from the same patient clustered together and were the same MLST type, one metric to suggest that that the strains recovered from a single specimen are identical. However,

similar MLST types in the same clade did not have similar antimicrobial resistance profiles proving that resistance is not tied to phylogeny (Figure 3-5).

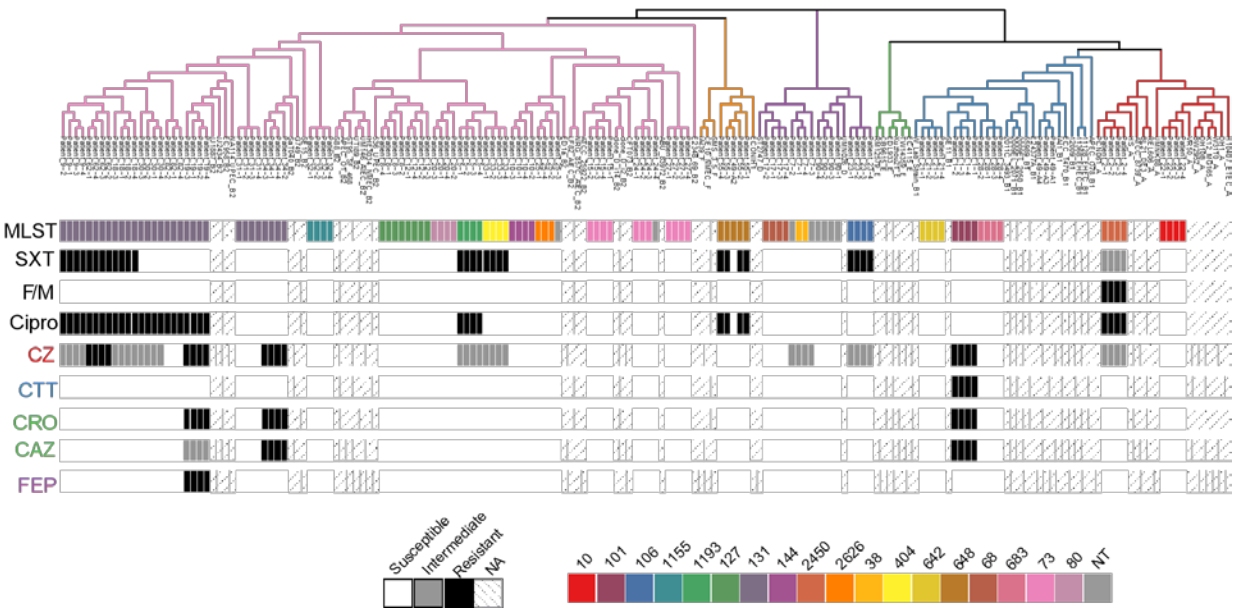


Figure 3-5. Uropathogenic *E. coli* are found in almost every clade and resistance is not mapped to phylogeny. Parsimony core SNP tree of all *E. coli* isolates (n = 120). MLST is depicted in the first row and antibiotic resistance profile for drugs commonly used to treat UTIs are depicted in the next 8 rows.

3.5 DISCUSSION

The confirmation of an active microbial community in the bladder have been an interest to many [153, 155, 181, 182] but understanding the relationship between the microbiome and urological disease has yet to be explored [152] using deep shotgun sequencing . This study focuses on comparing the microbiome of asymptomatic and suspected UTI patients of four different specimen interpretive classifications to gain knowledge on different methods which could be used to diagnose and treat UTIs.

Many studies have relied on marker gene surveys of the urine microbiome which provides a limited scope into function of the microbiome [182]. By utilizing isolate metagenomic and whole-metagenomic shotgun sequencing, we are not only able to study composition but gain

insight into the function. The results presented here prove that whole metagenomic sequencing could be a potential method in diagnosing and treating a UTI rather than using culturing methods which are biased and could take up to a week [51]. Our ability to identify putative pathogens and metabolic pathways that are enriched in specimen that are clinically designated as inconclusive allows us the ability to provide a foundation of new methods to accurately classify patients who have a UTI and provide antibiotics that will not disturb the commensal flora or increase antibiotic resistance.

We observed that the microbiome of asymptomatic individuals is stable over time. Therefore, the urine microbiome is not in flux and a diseased state that is compositionally different is in a state of dysbiosis. This is clear when we compared cultured slurries and direct sequenced urines from asymptomatic and suspected UTI specimens. The composition of the asymptomatic specimen were similar to many other studies who did targeted marker gene sequencing [181] and this is mainly due to the low diversity of the urine microbiome. The composition of suspected UTI specimen which were classified as positive, no growth, and insignificant were compositionally different from asymptomatic specimen but not contaminated. This suggests that the specimen clinically classified as contaminated may be the commensal flora of the urine microbiome rather than a contamination of the skin flora. Future studies where urine is directly extracted from the bladder rather than clean-catch will be necessary to determine if the contamination is from the skin or part of the commensal urine microbiome.

Our ability to recapitulate the cultured slurry microbiome of suspected UTI specimen but not asymptomatic proves to the capability in direct sequencing of the urine to measure disease. Since the urine microbiome has low diversity, samples do not have to be sequenced to the same depths as fecal microbiome studies to fully capture all individuals in the community. Therefore,

culturing only increases time to treatment. All isolates which were extracted from culture were identified in the direct sequenced specimen. The fact that in specimen where the putative pathogen was not the abundant species in the community but rather a different putative pathogen suggests that UTIs can be a multi-pathogen infection rather than a single pathogen. Therefore, culturing only limits our detection to a single pathogen and treatment is based on that single pathogen which could lead to recurrent UTIs and increase in antibiotic resistance.

3.6 CONCLUSIONS

We hypothesize that the urine microbiome of suspected UTI patients classified into the 4 categories was compositionally different when compared to asymptomatic specimens. The results validate this hypothesis, providing evidence that the urine microbiome does not decrease in diversity but compositionally which could lead to a diseased state. The altered state does not mean a dominance of a single pathogen but could be a multi-pathogenic infection. With whole metagenomic sequencing becoming easier to analyze, we predict with the inclusion of such technologies within the clinical classification and treatment of a UTI will aid in precise treatment strategies.

3.7 ACKNOWLEDGEMENTS

The work presented in this chapter was done in collaboration with Meghan Wallace, Carlos Eduardo Posada Perlaza, Alejandro Reyes Muñoz, Carey-Ann Burnham and Gautam Dantas. B.A. was supported by National Science Foundation graduate research fellow (award number DGE-1143945). This work was funded by Barnes Jewish Hospital Foundation-Washington University Institute of Clinical and Translational Sciences, NIH Director's New Innovator Award, the National Institute of Diabetes and Digestive and Kidney Diseases (DP2DK098089) and National Institute of General Medicine (R01GM099538) awarded to G.D.

We would also like to thank WIHSC who consented patients and collected the urine and case report forms from the asymptomatic women.

3.8 SUPPLEMENTARY FIGURES & TABLES

Table 3-S1. Demographic and specimen characteristics

Clinical variable	Entire cohort (n=162)	Asymptomatic (n = 10)	Positive (n=52)	Contaminated (n=9)	No Growth (n= 71)	Insignificant (n = 20)
Age (yr), mean (SD)	50.8 (20.1)	31.4 (8.77)	52.3 (21.2)	48.9 (20.8)	50.3(19.5)	48.1 (21)
Gender, no. (%)						
Females	103 (63.6)	10 (100)	40 (76)	8 (88)	30 (42.2)	15 (75)
Males	59 (36.4)	0	12 (23)	1 (11)	41 (57.8)	5 (25)
Race/ethnicity, no. (%)						
Caucasian	99 (61.1)	9 (90)	30 (57)	4 (44)	45 (63.3)	11 (55)
Black	57 (35.2)	0	21 (40)	4 (44)	23 (32.4)	9 (45)
Asian	2 (1.2)	1 (10)			1 (1.4)	0
Not Specified	4 (2.5)	0	1 (1)	1 (11)	2 (2.8)	0
Patient Type, no. (%)						
Inpatient	48 (29.6)		17 (32)	4 (44)	23 (32.4)	4 (20)
Ouptatient	103 (63.6)		34 (65)	5 (55)	48 (67.6)	16 (80)
Not Specified	1 (0.6)		1 (1)	0 (0)	0	0
Patient Location , no. (%)						
Emergency Department	17 (10.5)		9 (17)	0 (0)	8 (11.3)	0
Medicine	52 (32.1)		13 (25)	4 (44)	29 (40.8)	6 (30)
Oncology	15 (9.3)		3 (5)	1 (11)	10 (14.1)	1 (5)
Gynecology	19 (11.7)		5 (9)	2 (22)	6 (8.5)	6 (30)
Surgery	5 (3.1)		3 (5)	0 (0)	2 (2.8)	0
Other	43 (34)		19 (36)	2 (22)	16 (22.5)	7 (35)
Type of Urine Specimen, no (%)						
Catheter	21 (12.9)		5 (9)	0 (0)	16 (22.5)	0
Illeal Loop	1 (0.6)		1 (1)	0 (0)	0	0
Midstream	86 (53.1)		21 (40)	8 (88)	42 (59.2)	15 (75)

Urine					
Not Specified	40 (24.7)	25 (48)	1 (11)	11 (15.5)	5 (25)

Table 3-S2. Number of samples meeting the inclusion criteria during each step of the analysis.

	Obtained	**Sequenced**	**Minimum Number of Reads Acquired**	**Successfully Analyzed with Pipeline**
Isolates	224	220	219	212
Slurries				
Asymptomatic	30	30	30	30
Positive	52	48	48	47
No-Growth	6	6	6	6
Urines				
Asymptomatic	30	30	30	30
Insignificant	20	17	17	17
Contaminated	9	7	7	7
Positive	52	49	49	48
No-Growth	71	69	61	61

Table 3-S3. Statistical analysis of conducted on Bray-Curtis of direct sequenced urine specimen. Category = Positive, Asymptomatic, Insignificant, Contaminated and No Growth.

	Df	**SumsOfSqs**	**MeanSqs**	**F.Model**	**R2**	**Pr(>F)**	
Category	4	6.034	1.50838	4.4433	0.09236	0.001	***
Age	60	27.405	0.45675	1.3455	0.41952	0.001	***
Gender	1	0.482	0.48229	1.4207	0.00738	0.104	
Race	2	0.719	0.35946	1.0589	0.01101	0.36	

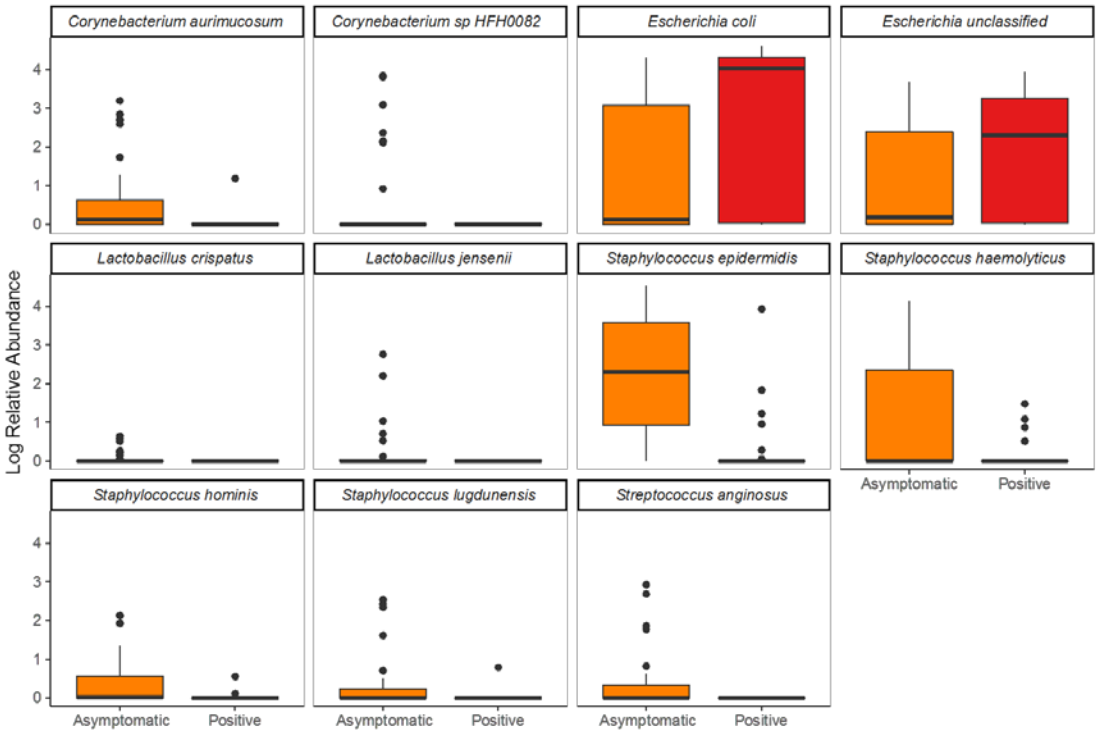


Figure 3-S2. Differentially abundant bacteria between asymptomatic and positive cultured slurries. Boxplots depicting log relative abundance of bacterial species between positive (n = 47) and asymptomatic (n = 30) cultured slurries. Differential abundance was calculated using ANCOM and FDR was used to correct p-values.

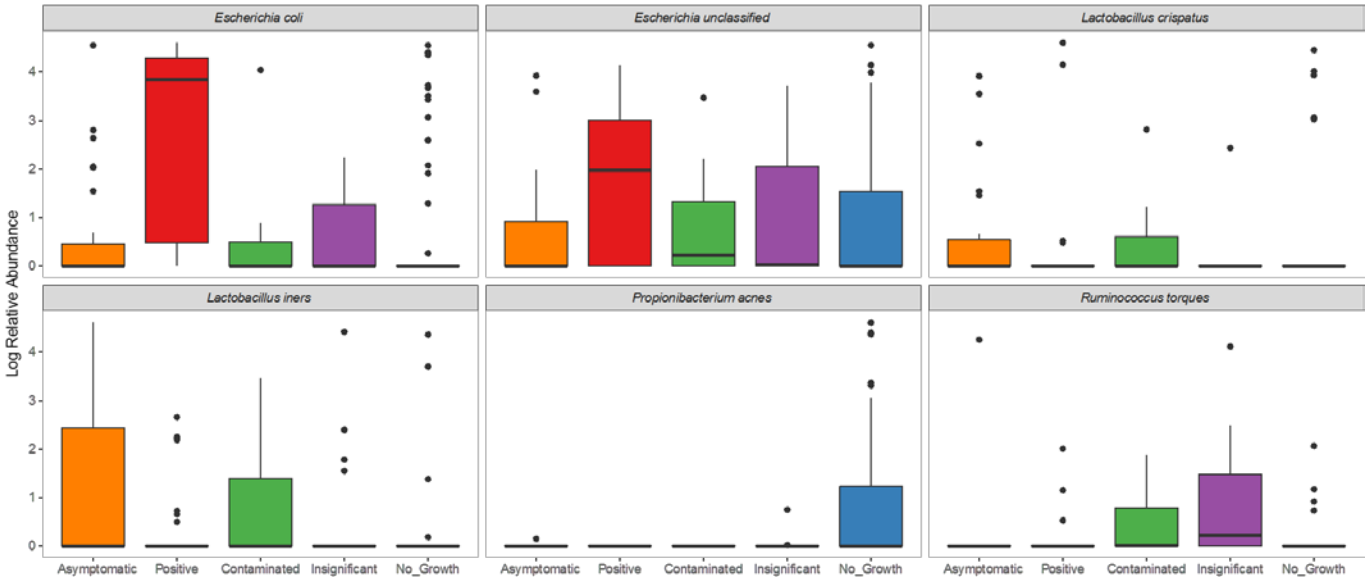


Figure 3-S1. Differentially abundant bacteria between asymptomatic, positive, contaminated, insignificant and no growth direct sequenced specimen. Boxplots depicting log relative abundance of bacterial species between positive (n=47), asymptomatic (n = 30), contaminated (n = 7), insignificant (n = 17), no growth (n = 61). Differential abundance was calculated using ANCOM and FDR was used to correct p-values.

Chapter 4 Conclusions and Future Directions

My work has shown that the microbiome can directly affect the fitness of the host. Endophytic root bacteria are correlated with total biomass of plant hosts which are sensitive to microbial composition disturbance. However, this disturbance does not affect the strength of the assembly processes on the overall composition. We provide evidence that the assemblage of root endophytic bacteria is structured by deterministic factors and the identity of the host has the strongest effect. This study provides a foundation to study root endophytic communities in prairie plants. However, we have yet to know if the bacterial species identified truly can alter plant fitness. Challenging plants with bacterial species found in the roots will elucidate direct function of bacterial species alongside whole genome sequencing.

Our results show that the urine microbiome is in an altered state when the individual has a suspected UTI. The microbiome is enriched with pathogens and often more than one. This poses a potential problem as conventional clinical diagnostics typically view UTI as a monomicrobial infection and focus on the isolation of a single pathogen. Furthermore, conventional urine culture methods are primarily tuned for the isolation and recovery of *E. coli*. The results shown in my dissertation provides further evidence that direct sequencing not only recapitulates the cultured composition but provides further necessary details for future diagnostic testing. Therefore, there isn't a need to spend time culturing when we can sequence the urine directly. Before we can make such claim, we need to have a solid classification system that can replace the current clinical classifications. To do so, we need to have access to all patient outcomes and treatments to better compare the treatment recommendations concluded using solely sequencing data compared to the current method.

Studying two different habitats have led to similar conclusions: microbiomes can affect their hosts. Microbes have been shown to expand the genetic architecture of their hosts by providing a service the hosts are not genetically capable of doing. However, these microbes are not randomly placed in certain niches and then co-opted based on a function. Instead, the microbes are directly selected by the host based on their function. Due to evolution of sequencing technology, elucidating function of microbiomes and single bacterial species is feasible. Once we can tie both function and assembly theory, we will be able to understand answer the questions central to ecology: how did they get there, what are they doing and why are they there.

References

1. Groombridge, B. and M. Jenkins, *World Atlas of Biodiversity. Prepared by the UNEP World Conservation Monitoring Centre.* Unversity of California Press, Berkeley, 2002.
2. Horner-Devine, M.C., K.M. Carney, and B.J. Bohannan, *An ecological perspective on bacterial biodiversity.* Proceedings of the Royal Society of London B: Biological Sciences, 2004. **271**(1535): p. 113-122.
3. Darcy, J.L., et al., *Global distribution of Polaromonas phylotypes-evidence for a highly successful dispersal capacity.* PLoS One, 2011. **6**(8): p. e23742.
4. Finlay, B.J., *Global dispersal of free-living microbial eukaryote species.* Science, 2002. **296**(5570): p. 1061-1063.
5. Finlay, B.J. and K.J. Clarke, *Ubiquitous dispersal of microbial species.* Nature, 1999. **400**(6747): p. 828.
6. Wilkinson, D.M., et al., *Modelling the effect of size on the aerial dispersal of microorganisms.* Journal of Biogeography, 2012. **39**(1): p. 89-97.
7. Lennon, J.T. and S.E. Jones, *Microbial seed banks: the ecological and evolutionary implications of dormancy.* Nature reviews. Microbiology, 2011. **9**(2): p. 119.
8. Heidelberg, J.F., et al., *Genome sequence of the dissimilatory metal ion-reducing bacterium Shewanella oneidensis.* Nature biotechnology, 2002. **20**(11): p. 1118.
9. Fredrickson, J.K., et al., *Towards environmental systems biology of Shewanella.* Nature reviews. Microbiology, 2008. **6**(8): p. 592.
10. Nemergut, D.R., et al., *Patterns and processes of microbial community assembly.* Microbiology and Molecular Biology Reviews, 2013. **77**(3): p. 342-356.
11. Hanson, C.A., et al., *Beyond biogeographic patterns: processes shaping the microbial landscape.* Nat Rev Micro, 2012. **10**(7): p. 497-506.
12. Mihaljevic, J.R., *Linking metacommunity theory and symbiont evolutionary ecology.* Trends in ecology & evolution, 2012. **27**(6): p. 323-329.
13. Horner-Devine, M.C., et al., *A taxa-area relationship for bacteria.* Nature, 2004. **432**(7018): p. 750.
14. Amend, A.S., et al., *Macroecological patterns of marine bacteria on a global scale.* Journal of Biogeography, 2013. **40**(4): p. 800-811.
15. Whitaker, R.J., D.W. Grogan, and J.W. Taylor, *Geographic Barriers Isolate Endemic Populations of Hyperthermophilic Archaea.* Science, 2003. **301**(5635): p. 976-978.
16. O'Malley, M.A., *'Everything is everywhere: but the environment selects': ubiquitous distribution and ecological determinism in microbial biogeography.* Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 2008. **39**(3): p. 314-325.
17. Vellend, M. and A. Agrawal, *Conceptual synthesis in community ecology.* The Quarterly review of biology, 2010. **85**(2): p. 183-206.
18. Costello, E.K., et al., *The Application of Ecological Theory Toward an Understanding of the Human Microbiome.* Science, 2012. **336**(6086): p. 1255-1262.
19. Prosser, J.I., et al., *The role of ecological theory in microbial ecology.* Nature Reviews Microbiology, 2007. **5**(5): p. 384-392.
20. Prosser, J.I., *Dispersing misconceptions and identifying opportunities for the use of'omics' in soil microbial ecology.* Nature reviews. Microbiology, 2015. **13**(7): p. 439.

21. Fierer, N. and J.T. Lennon, *The generation and maintenance of diversity in microbial communities.* American Journal of Botany, 2011. **98**(3): p. 439-448.
22. Webb, C.O., *Exploring the phylogenetic structure of ecological communities: an example for rain forest trees.* The American Naturalist, 2000. **156**(2): p. 145-155.
23. Graham, C.H. and P.V. Fine, *Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time.* Ecology letters, 2008. **11**(12): p. 1265-1277.
24. Chase, J.M., et al., *Using null models to disentangle variation in community dissimilarity from variation in α-diversity.* Ecosphere, 2011. **2**(2): p. 1-11.
25. Chase, J.M. and J.A. Myers, *Disentangling the importance of ecological niches from stochastic processes across scales.* Philosophical transactions of the Royal Society B: Biological sciences, 2011. **366**(1576): p. 2351-2363.
26. Gotelli, N.J. and G.L. Entsminger, *Swap and fill algorithms in null model analysis: rethinking the knight's tour.* Oecologia, 2001. **129**(2): p. 281-291.
27. Gotelli, N.J. and B.J. McGill, *Null versus neutral models: what's the difference?* Ecography, 2006. **29**(5): p. 793-800.
28. Woodcock, S., et al., *Neutral assembly of bacterial communities.* FEMS microbiology ecology, 2007. **62**(2): p. 171-180.
29. Sloan, W.T., et al., *Quantifying the roles of immigration and chance in shaping prokaryote community structure.* Environmental microbiology, 2006. **8**(4): p. 732-740.
30. Burns, A.R., et al., *Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development.* The ISME journal, 2016. **10**(3): p. 655.
31. Stegen, J.C., et al., *Stochastic and deterministic assembly processes in subsurface microbial communities.* The ISME journal, 2012. **6**(9): p. 1653.
32. Fierer, N. and R.B. Jackson, *The diversity and biogeography of soil bacterial communities.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(3): p. 626-631.
33. Baker, B.J. and J.F. Banfield, *Microbial communities in acid mine drainage.* FEMS Microbiology Ecology, 2003. **44**(2): p. 139-152.
34. Fierer, N., et al., *Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients.* The ISME Journal, 2012. **6**(5): p. 1007-1017.
35. Bardgett, R.D., et al., *Seasonal changes in soil microbial communities along a fertility gradient of temperate grasslands.* Soil Biology and Biochemistry, 1999. **31**(7): p. 1021-1030.
36. Andrew, D.R., et al., *Abiotic Factors Shape Microbial Diversity in Sonoran Desert Soils.* Applied and Environmental Microbiology, 2012. **78**(21): p. 7527-7537.
37. Stegen, J.C., et al., *Quantifying community assembly processes and identifying features that impose them.* ISME J, 2013. **7**(11): p. 2069-2079.
38. Bever, J.D., et al., *Rooting theories of plant community ecology in microbial interactions.* Trends in Ecology & Evolution, 2010. **25**(8): p. 468-478.
39. Maherali, H. and J.N. Klironomos, *Influence of phylogeny on fungal community assembly and ecosystem functioning.* science, 2007. **316**(5832): p. 1746-1748.
40. Wardle, D.A., *The influence of biotic interactions on soil biodiversity.* Ecology letters, 2006. **9**(7): p. 870-886.

41. Van Der Heijden, M.G.A., R.D. Bardgett, and N.M. Van Straalen, *The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems.* Ecology Letters, 2008. **11**(3): p. 296-310.
42. Krüger, C., et al., *Plant communities rather than soil properties structure arbuscular mycorrhizal fungal communities along primary succession on a mine spoil.* Frontiers in Microbiology, 2017. **8**.
43. Horn, S., et al., *Linking the community structure of arbuscular mycorrhizal fungi and plants: a story of interdependence?* The ISME Journal, 2017. **11**(6): p. 1400-1411.
44. Martínez, I., et al., *The gut microbiota of rural Papua New Guineans: composition, diversity patterns, and ecological processes.* Cell reports, 2015. **11**(4): p. 527-538.
45. Ley, R.E., D.A. Peterson, and J.I. Gordon, *Ecological and Evolutionary Forces Shaping Microbial Diversity in the Human Intestine.* Cell, 2006. **124**(4): p. 837-848.
46. Hubbell, S.P., *The unified neutral theory of biodiversity and biogeography (MPB-32)*. Vol. 32. 2001: Princeton University Press.
47. Davies, J., *Inactivation of antibiotics and the dissemination of resistance genes.* Science, 1994. **264**(5157): p. 375-383.
48. Xu, J., *Invited review: microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances.* Molecular ecology, 2006. **15**(7): p. 1713-1731.
49. Vartoukian, S.R., R.M. Palmer, and W.G. Wade, *Strategies for culture of 'unculturable' bacteria.* FEMS Microbiology Letters, 2010. **309**(1): p. 1-7.
50. Lieberman, T.D., et al., *Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures.* Nature Genetics, 2014. **46**(1): p. 82-87.
51. Didelot, X., et al., *Transforming clinical microbiology with bacterial genome sequencing.* Nature reviews. Genetics, 2012. **13**(9): p. 601-12.
52. Shapiro, B.J., et al., *Population genomics of early events in the ecological differentiation of bacteria.* science, 2012. **336**(6077): p. 48-51.
53. Wilson, D.J., *Insights from genomics into bacterial pathogen populations.* PLoS Pathog, 2012. **8**(9): p. e1002874.
54. Mättö, J., et al., *Composition and temporal stability of gastrointestinal microbiota in irritable bowel syndrome—a longitudinal study in IBS and control subjects.* FEMS Immunology & Medical Microbiology, 2005. **43**(2): p. 213-222.
55. Sbordone, L. and C. Bortolaia, *Oral microbial biofilms and plaque-related diseases: microbial communities and their role in the shift from oral health to disease.* Clinical Oral Investigations, 2003. **7**(4): p. 181-188.
56. Zhao, J., et al., *Decade-long bacterial community dynamics in cystic fibrosis airways.* Proceedings of the National Academy of Sciences, 2012. **109**(15): p. 5809-5814.
57. Yassour, M., et al., *Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability.* Science translational medicine, 2016. **8**(343): p. 343ra81-343ra81.
58. Woyke, T., et al., *Symbiosis insights through metagenomic analysis of a microbial consortium.* Lawrence Berkeley National Laboratory, 2006.
59. Hacquard, S., et al., *Microbiota and host nutrition across plant and animal kingdoms.* Cell Host & Microbe, 2015. **17**(5): p. 603-616.
60. Bao, Y., et al., *Questions and challenges associated with studying the microbiome of the urinary tract.* Annals of Translational Medicine, 2017.

61. Turner, T.R., E.K. James, and P.S. Poole, *The plant microbiome.* Genome Biology, 2013. **14**(6): p. 209.
62. Berendsen, R.L., C.M. Pieterse, and P.A. Bakker, *The rhizosphere microbiome and plant health.* Trends in plant science, 2012. **17**(8): p. 478-486.
63. Ferrenberg, S., et al., *Changes in assembly processes in soil bacterial communities following a wildfire disturbance.* The ISME journal, 2013. **7**(6): p. 1102.
64. Allison, S.D. and J.B. Martiny, *Resistance, resilience, and redundancy in microbial communities.* Proceedings of the National Academy of Sciences, 2008. **105**(Supplement 1): p. 11512-11519.
65. Tilman, D. and S. Pacala, *The maintenance of species richness in plant communities.* Species diversity in ecological communities, 1993: p. 13-25.
66. Klironomos, J.N., *Feedback with soil biota contributes to plant rarity and invasiveness in communities.* Nature, 2002. **417**(6884): p. 67-70.
67. Reynolds, H.L., et al., *GRASSROOTS ECOLOGY: PLANT–MICROBE–SOIL INTERACTIONS AS DRIVERS OF PLANT COMMUNITY STRUCTURE AND DYNAMICS.* Ecology, 2003. **84**(9): p. 2281-2291.
68. Schnitzer, S.A., et al., *Soil microbes drive the classic plant diversity–productivity pattern.* Ecology, 2011. **92**(2): p. 296-303.
69. Prober, S.M., et al., *Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide.* Ecology Letters, 2015. **18**(1): p. 85-95.
70. Clay, K. and J. Holah, *Fungal endophyte symbiosis and plant diversity in successional fields.* Science, 1999. **285**(5434): p. 1742-1744.
71. Van Der Heijden, M.G., et al., *Symbiotic bacteria as a determinant of plant community structure and plant productivity in dune grassland.* FEMS Microbiology Ecology, 2006. **56**(2): p. 178-187.
72. Bever, J.D., *Soil community feedback and the coexistence of competitors: conceptual frameworks and empirical tests.* New Phytologist, 2003. **157**(3): p. 465-473.
73. Hartnett, D.C. and G.W. Wilson, *The role of mycorrhizas in plant community structure and dynamics: lessons from grasslands*, in *Diversity and Integration in Mycorrhizas*. 2002, Springer. p. 319-331.
74. HilleRisLambers, J., et al., *Rethinking community assembly through the lens of coexistence theory.* Annual Review of Ecology, Evolution, and Systematics, 2012. **43**: p. 227-248.
75. Bulgarelli, D., et al., *Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota.* Nature, 2012. **488**(7409): p. 91-95.
76. Lundberg, D.S., et al., *Defining the core Arabidopsis thaliana root microbiome.* Nature, 2012. **488**(7409): p. 86-90.
77. Edwards, J., et al., *Structure, variation, and assembly of the root-associated microbiomes of rice.* Proceedings of the National Academy of Sciences, 2015. **112**(8): p. E911-E920.
78. Gottel, N.R., et al., *Distinct microbial communities within the endosphere and rhizosphere of Populus deltoides roots across contrasting soil types.* Applied and environmental microbiology, 2011. **77**(17): p. 5934-5944.
79. Shakya, M., et al., *A multifactor analysis of fungal and bacterial community structure in the root microbiome of mature Populus deltoides trees.* PLoS One, 2013. **8**(10): p. e76382.

80. Reinhold-Hurek, B., et al., *Roots shaping their microbiome: global hotspots for microbial activity.* Annual review of phytopathology, 2015. **53**: p. 403-424.
81. Coleman-Derr, D., et al., *Plant compartment and biogeography affect microbiome composition in cultivated and native Agave species.* New Phytologist, 2016. **209**(2): p. 798-811.
82. Peiffer, J.A., et al., *Diversity and heritability of the maize rhizosphere microbiome under field conditions.* Proceedings of the National Academy of Sciences, 2013. **110**(16): p. 6548-6553.
83. Hallmann, J., et al., *Bacterial endophytes in agricultural crops.* Canadian Journal of Microbiology, 1997. **43**(10): p. 895-914.
84. Conrath, U., et al., *Priming: getting ready for battle.* Molecular Plant-Microbe Interactions, 2006. **19**(10): p. 1062-1071.
85. Hardoim, P.R., L.S. van Overbeek, and J.D.v. Elsas, *Properties of bacterial endophytes and their proposed role in plant growth.* Trends in Microbiology, 2008. **16**(10): p. 463-471.
86. Gaiero, J.R., et al., *Inside the root microbiome: Bacterial root endophytes and plant growth promotion.* American Journal of Botany, 2013. **100**(9): p. 1738-1750.
87. Pillay, V.K. and J. Nowak, *Inoculum density, temperature, and genotype effects on in vitro growth promotion and epiphytic and endophytic colonization of tomato (Lycopersicon esculentum L.) seedlings inoculated with a pseudomonad bacterium.* Canadian Journal of Microbiology, 1997. **43**(4): p. 354-361.
88. Bulgarelli, D., et al., *Structure and functions of the bacterial microbiota of plants.* Annual review of plant biology, 2013. **64**: p. 807-838.
89. Houseman, G.R., et al., *Perturbations alter community convergence, divergence, and formation of multiple community states.* Ecology, 2008. **89**(8): p. 2172-2180.
90. Chaparro, J.M., D.V. Badri, and J.M. Vivanco, *Rhizosphere microbiome assemblage is affected by plant development.* ISME J, 2014. **8**(4): p. 790-803.
91. Schweitzer, J.A., et al., *Plant-soil microorganism interactions: heritable relationship between plant genotype and associated soil microorganisms.* Ecology, 2008. **89**(3): p. 773-81.
92. Turner, T., et al., *Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants*, in *ISME J*. 2013.
93. Callaway, R.M., et al., *Soil biota and exotic plant invasion.* Nature, 2004. **427**(6976): p. 731-733.
94. Ehrenfeld, J.G., B. Ravit, and K. Elgersma, *Feedback in the plant-soil system.* Annu. Rev. Environ. Resour., 2005. **30**: p. 75-115.
95. Martiny, J.B.H., et al., *Microbial biogeography: putting microorganisms on the map.* Nat Rev Micro, 2006. **4**(2): p. 102-112.
96. Bever, J.D., *Feeback between plants and their soil communities in an old field community.* Ecology, 1994. **75**(7): p. 1965-1977.
97. Mangan, S.A., et al., *Negative plant-soil feedback predicts tree-species relative abundance in a tropical forest.* Nature, 2010. **466**(7307): p. 752-755.
98. Pendergast, T.H.t., D.J. Burke, and W.P. Carson, *Belowground biotic complexity drives aboveground dynamics: a test of the soil community feedback model.* New Phytol, 2013. **197**(4): p. 1300-10.

99. Pernilla Brinkman, E., et al., *Plant–soil feedback: experimental approaches, statistical analyses and ecological interpretations.* Journal of Ecology, 2010. **98**(5): p. 1063-1073.
100. Bever, J.D., T.G. Platt, and E.R. Morton, *Microbial population and community dynamics on plant roots and their feedbacks on plant communities.* Annual review of microbiology, 2012. **66**: p. 265-283.
101. Handelsman, J., *Metagenomics: application of genomics to uncultured microorganisms.* Microbiol Mol Biol Rev, 2004. **68**(4): p. 669-85.
102. Torsvik, V., R. Sørheim, and J. Goksøyr, *Total bacterial diversity in soil and sediment communities—a review.* Journal of industrial microbiology & biotechnology, 1996. **17**(3): p. 170-178.
103. Riesenfeld, C.S., P.D. Schloss, and J. Handelsman, *Metagenomics: genomic analysis of microbial communities.* Annu. Rev. Genet., 2004. **38**: p. 525-552.
104. Rappé, M.S. and S.J. Giovannoni, *The uncultured microbial majority.* Annual Reviews in Microbiology, 2003. **57**(1): p. 369-394.
105. Samson, F.B. and F.L. Knopf, *Prairie conservation: preserving North America's most endangered ecosystem*. 1996: Island Press.
106. Caporaso, J.G., et al., *Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms.* ISME J, 2012. **6**(8): p. 1621-1624.
107. Caporaso, J.G., et al., *QIIME allows analysis of high-throughput community sequencing data.* Nature methods, 2010. **7**(5): p. 335-336.
108. Edgar, R.C., *UPARSE: highly accurate OTU sequences from microbial amplicon reads.* Nature methods, 2013. **10**(10): p. 996-998.
109. McMurdie, P.J. and S. Holmes, *Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible.* PLoS Comput Biol, 2014. **10**(4): p. e1003531.
110. Oksanen, J., et al., *Package 'vegan'*. Community ecology package, version, 2013. **2**(9).
111. Hervé, M., *RVAideMemoire: Diverse basic statistical and graphical functions. R package version 0.9-45-2.* Computer software.(http://CRAN. R-project. org/packa ge¼RVAideMemoire), 2015.
112. McMurdie, P.J. and S. Holmes, *phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data.* PloS one, 2013. **8**(4): p. e61217.
113. Chase, J.M., et al., *Using null models to disentangle variation in community dissimilarity from variation in α -diversity.* Ecosphere, 2011. **2**(2): p. 1-11.
114. Alberti, J., et al., *Herbivore exclusion promotes a more stochastic plant community assembly in a natural grassland.* Ecology, 2017. **98**(4): p. 961-970.
115. Catano, C.P., T.L. Dickson, and J.A. Myers, *Dispersal and neutral sampling mediate contingent effects of disturbance on plant beta -diversity: a meta- ana* Ecology Letters, 2017. **20**(3): p. 347-356.
116. Tucker, C.M., et al., *Differentiating between niche and neutral assembly in metacommunities using null models of β -diversity.* Oikos, 2015.
117. Wickham, H., *ggplot2: elegant graphics for data analysis*. 2016: Springer.
118. Lebeis, S.L., *Greater than the sum of their parts: characterizing plant microbiomes at the community-level.* Current Opinion in Plant Biology, 2015. **24**: p. 82-86.
119. Aleklett, K., et al., *Wild plant species growing closely connected in a subalpine meadow host distinct root-associated bacterial communities.* PeerJ, 2015. **3**: p. e804.

120. Sessitsch, A., et al., *Functional characteristics of an endophyte community colonizing rice roots as revealed by metagenomic analysis.* Mol Plant Microbe Interact, 2012. **25**(1): p. 28-36.
121. Manter, D.K., et al., *Pyrosequencing reveals a highly diverse and cultivar-specific bacterial endophyte community in potato roots.* Microb Ecol, 2010. **60**(1): p. 157-66.
122. Christian, N., et al., *Plant Host and Geographic Location Drive Endophyte Community Composition in the Face of Perturbation.* Microbial ecology, 2016. **72**(3): p. 621-632.
123. Broeckling, C.D., et al., *Root exudates regulate soil fungal community composition and diversty.* Appl Environ Microbiol, 2008. **74**.
124. Deng, J., et al., *Shifts of tundra bacterial and archaeal communities along a permafrost thaw gradient in Alaska.* Molecular Ecology, 2015. **24**(1): p. 222-234.
125. Monciardini, P., et al., *Conexibacter woesei gen. nov., sp. nov., a novel representative of a deep evolutionary line of descent within the class Actinobacteria.* International journal of systematic and evolutionary microbiology, 2003. **53**(2): p. 569-576.
126. Albuquerque, L. and M.S. da Costa, *The Families Conexibacteraceae, Patulibacteraceae and Solirubrobacteraceae*, in *The Prokaryotes: Actinobacteria*, E. Rosenberg, et al., Editors. 2014, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 185-200.
127. Middleton, K.J., M.J. Bell, and J.P. Thompson, *Effects of soil sterilization, inoculation with vesicular-arbuscular mycorrhizal fungi and cropping history on peanut (Arachis hypogaea L.) growth in an oxisol from subtropical Australia.* Plant and Soil, 1989. **117**(1): p. 41-48.
128. Reinhart, K.O., et al., *Plant–soil biota interactions and spatial distribution of black cherry in its native and invasive ranges.* Ecology Letters, 2003. **6**(12): p. 1046-1050.
129. Knevel, I.C., et al., *Release from native root herbivores and biotic resistance by soil pathogens in a new habitat both affect the alien Ammophila arenaria in South Africa.* Oecologia, 2004. **141**(3): p. 502-510.
130. Kim, H., et al., *High population of Sphingomonas species on plant surface.* Journal of applied microbiology, 1998. **85**(4): p. 731-736.
131. Faisal, M. and S. Hasnain, *Growth stimulatory effect of Ochrobactrum intermedium and Bacillus cereus on Vigna radiata plants.* Lett Appl Microbiol, 2006. **43**(4): p. 461-6.
132. van der Heijden, M.G.A., et al., *Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity.* Nature, 1998. **396**(6706): p. 69-72.
133. Tambyah, P.A. and D.G. Maki, *Catheter-associated urinary tract infection is rarely symptomatic: a prospective study of 1497 catheterized patients.* Archives of Internal Medicine, 2000. **160**(5): p. 678-682.
134. Litza, J.A. and J.R. Brill, *Urinary Tract Infections.* Primary Care, 2010. **37**(3): p. 491-507.
135. Franz, M. and W.H. Hörl, *Common errors in diagnosis and management of urinary tract infection. I: Pathophysiology and diagnostic techniques.* Nephrology Dialysis Transplantation, 1999. **14**(11): p. 2746-2753.
136. Foxman, B., *Epidemiology of urinary tract infections: incidence, morbidity, and economic costs.* The American Journal of Medicine, 2002. **113**(1): p. 5S-13S.
137. Johnson, C., *Definitions, classification, and clinical presentation of urinary tract infections.* The Medical Clinics of North America, 1991. **75**(2): p. 241-252.

138. American College of, O. and Gynecologists, *ACOG Practice Bulletin No. 91: Treatment of urinary tract infections in nonpregnant women.* Obstetrics and gynecology 2008. **111**(3): p. 785-94.
139. Sheffield, J.S. and F.G. Cunningham, *Urinary tract infection in women.* Obstetrics & Gynecology, 2005. **106**(5, Part 1): p. 1085-1092.
140. Foxman, B., *The epidemiology of urinary tract infection.* Nature Reviews Urology, 2010. **7**(12): p. 653-660.
141. Dyer, I.E., T.M. Sankary, and J.A. Dawson, *Antibiotic resistance in bacterial urinary tract infections, 1991 to 1997.* Western Journal of Medicine, 1998. **169**(5): p. 265-268.
142. Gupta, K., D. Scholes, and W.E. Stamm, *Increasing prevalence of antimicrobial resistance among uropathogens causing acute uncomplicated cystitis in women.* Jama, 1999. **281**(8): p. 736-738.
143. Kahlmeter, G., *An international survey of the antimicrobial susceptibility of pathogens from uncomplicated urinary tract infections: the ECO· SENS Project.* Journal of Antimicrobial Chemotherapy, 2003. **51**(1): p. 69-76.
144. Manges, A.R., et al., *Widespread distribution of urinary tract infections caused by a multidrug-resistant Escherichia coli clonal group.* New England Journal of Medicine, 2001. **345**(14): p. 1007-1013.
145. Silverman, J.A., et al., *From Physiology to Pharmacy: Developments in the Pathogenesis and Treatment of Recurrent Urinary Tract Infections.* Current Urology Reports, 2013. **14**(5): p. 448-456.
146. Kostakioti, M., S.J. Hultgren, and M. Hadjifrangiskou, *Molecular blueprint of uropathogenic Escherichia coli virulence provides clues toward the development of anti-virulence therapeutics.* Virulence, 2012. **3**(7): p. 592-593.
147. Kurtaran, B., et al., *Antibiotic resistance in community-acquired urinary tract infections: prevalence and risk factors.* Medical science monitor: international medical journal of experimental and clinical research, 2010. **16**(5): p. CR246-51.
148. Tempera, G., et al., *The impact of prulifloxacin on vaginal lactobacillus microflora: an in vivo study.* Journal of Chemotherapy, 2009. **21**(6): p. 646-650.
149. Dantas, G. and M.O. Sommer, *How to Fight Back Against Antibiotic Resistance.* American Scientist 2014. **102**(1): p. 42-51.
150. Sommer, M.O. and G. Dantas, *Antibiotics and the resistant microbiome.* Current opinion in microbiology, 2011. **14**(5): p. 556-563.
151. Sommer, M.O., G.M. Church, and G. Dantas, *The human microbiome harbors a diverse reservoir of antibiotic resistance genes.* Virulence, 2010. **1**(4): p. 299-303.
152. Whiteside, S.A., et al., *The microbiome of the urinary tract [mdash] a role beyond infection.* Nature Reviews Urology, 2015.
153. Wolfe, A.J., et al., *Evidence of uncultivated bacteria in the adult female bladder.* Journal of Clinical Microbiology, 2012. **50**(4): p. 1376-1383.
154. Nelson, D.E., et al., *Characteristic male urine microbiomes associate with asymptomatic sexually transmitted infection.* PLoS One, 2010. **5**(11): p. e14116.
155. Lewis, D.A., et al., *The human urinary microbiome; bacterial DNA in voided urine of asymptomatic adults.* Frontiers in cellular and infection microbiology, 2013. **3**.
156. Flores-Mireles, A.L., et al., *Urinary tract infections: epidemiology, mechanisms of infection and treatment options.* Nature Reviews Microbiology, 2015. **13**(5): p. 269-284.

157. Schmiemann, G., et al., *The diagnosis of urinary tract infection: a systematic review.* Deutsches Ärzteblatt International, 2010. **107**(21): p. 361-367.
158. Siddiqui, H., et al., *Pathogens in Urine from a Female Patient with Overactive Bladder Syndrome Detected by Culture-independent High Throughput Sequencing: A Case Report.* The open microbiology journal, 2014. **8**: p. 148.
159. Domann, E., et al., *Culture-independent identification of pathogenic bacteria and polymicrobial infections in the genitourinary tract of renal transplant recipients.* Journal of clinical microbiology, 2003. **41**(12): p. 5500-5510.
160. Hasman, H., et al., *Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples.* Journal of clinical microbiology, 2014. **52**(1): p. 139-46.
161. McCarter, Y., et al., *Cumitech 2C: laboratory diagnosis of urinary tract infections.* American Society for Microbiology, Washington, DC, 2009.
162. Patel, J., et al., *Performance standards for antimicrobial susceptibility testing; twenty-fourth informational supplement.* CLSI standards for antimicrobial susceptibility testing, 2014. **34**(1): p. 1-226.
163. Manji, R., et al., *Multi-center evaluation of the VITEK® MS system for mass spectrometric identification of non-Enterobacteriaceae Gram-negative bacilli.* European Journal of Clinical Microbiology & Infectious Diseases, 2014. **33**(3): p. 337-346.
164. Richter, S.S., et al., *Identification of Enterobacteriaceae by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry using the VITEK MS system.* European Journal of Clinical Microbiology & Infectious Diseases, 2013. **32**(12): p. 1571-1578.
165. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-2120.
166. Schmieder, R. and R. Edwards, *Fast identification and removal of sequence contamination from genomic and metagenomic datasets.* PLoS One, 2011. **6**(3): p. e17288.
167. Baym, M., et al., *Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes.* PLoS ONE, 2015. **10**(5): p. e0128036.
168. Segata, N., et al., *Metagenomic microbial community profiling using unique clade-specific marker genes.* Nature Methods, 2012. **9**(8): p. 811-814.
169. Zaharia, M., et al., *Faster and more accurate sequence alignment with SNAP.* arXiv preprint arXiv:1111.5572, 2011.
170. Assefa, S., et al., *ABACAS: algorithm-based automatic contiguation of assembled sequences.* Bioinformatics, 2009. **25**(15): p. 1968-1969.
171. Walker, B.J., et al., *Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement.* PLoS ONE, 2014. **9**(11): p. e112963.
172. Seemann, T., *Prokka: rapid prokaryotic genome annotation.* Bioinformatics, 2014. **30**(14): p. 2068-9.
173. Gibson, M.K., K.J. Forsberg, and G. Dantas, *Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology.* The ISME journal, 2014.
174. Jia, B., et al., *CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database.* Nucleic Acids Res, 2017. **45**(D1): p. D566-d573.
175. Peng, Y., et al., *IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth.* Bioinformatics, 2012. **28**(11): p. 1420-1428.

176. Truong, D.T., et al., *MetaPhlAn2 for enhanced metagenomic taxonomic profiling.* Nature methods, 2015. **12**(10): p. 902-903.
177. Mandal, S., et al., *Analysis of composition of microbiomes: a novel method for studying microbial composition.* Microbial ecology in health and disease, 2015. **26**.
178. Abubucker, S., et al., *Metabolic reconstruction for metagenomic data and its application to the human microbiome.* PLoS computational biology, 2012. **8**(6): p. e1002358.
179. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nature methods, 2012. **9**(4): p. 357-359.
180. Gardner, S.N. and B.G. Hall, *When whole-genome alignments just won't work: kSNP v2 software for alignment-free SNP discovery and phylogenetics of hundreds of microbial genomes.* PLoS One, 2013. **8**(12): p. e81760.
181. Hilt, E.E., et al., *Urine is not sterile: use of enhanced urine culture techniques to detect resident bacterial flora in the adult female bladder.* Journal of clinical microbiology, 2014. **52**(3): p. 871-876.
182. Pearce, M.M., et al., *The female urinary microbiome: a comparison of women with and without urgency urinary incontinence.* MBio, 2014. **5**(4): p. e01283-14.