CHARACTERIZATION OF THE BACTERIAL COMMUNITIES OF ROMAINE LETTUCE: INTERACTIONS WITH ENVIRONMENTAL CONDITIONS AND FOOD SAFETY

by

Diana Vanessa Sarria Zuniga

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Botany and Plant Pathology

West Lafayette, Indiana

December 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Robert E. Pruitt, Chair

Department of Botany and Plant Pathology

Dr. Amanda Deering

Department of Food Science

Dr. Stephen B. Goodwin

Department of Botany and Plant Pathology

Dr. M. Catherine Aime

Department of Botany and Plant Pathology

Approved by:

Dr. Christopher Staiger

To my beloved family who has been my faithful support through every step in my life. Thanks mom, dad and grandma for every pray, and your understanding that my dreams would take me far away from home.

To my soul mate, Angel, literally you are my angel. Thanks for your love, friendship, company, for every hour you helped me out with my research, for cheering me up and taking food on the long working days.

To every person that made my life easier during these 5 years at Purdue with their help, friendship, advices, laughs and wine.

Dedicada a mi amada familia quienes han sido mi fiel apoyo a través de cada paso en mi vida, Gracias mami, papi y abuelita por cada oración y por su comprensión de que mis sueños me llevarían lejos de casa.

A mi alma gemela, Ángel, literalmente tú eres mi ángel. Gracias por tu amor, amistad, compañía, por cada hora que me ayudaste en el laboratorio e invernadero, por animarme y llevarme comida en los días largos de trabajo.

A cada persona que hizo mi vida más fácil durante estos 5 años en Purdue, con su ayuda, amistad, consejos, risas y vino.

ACKNOWLEDGMENTS

I would like to acknowledge that many people who made this thesis possible.

First, all my gratitude to my advisor Dr. Robert Pruitt, for his continuous guidance, support and trust that have brought me to the end of my PhD program. I am also grateful with my committee members, Drs. Cathie Aime, Amanda Deering and Steven Goodwin for all their help and time provided during my graduate career. I would like also to extend my appreciation to all the members of the Department of Botany and Plant Pathology, past and present, for any academic, funding and personal help received. I would like to also acknowledge every lab and person who helped me out in some way with reagents, advices, access to equipment, etc. Special thanks to Aime lab members, Deering lab members, Lee lab members, Cai lab members, Yoon lab members, and to the staff of greenhouse and growth chamber facilities. I also want to acknowledge the collaborators on my research projects: Dr. Euiwon Bae and Dr. J. Paul Robinson lab members for the fruitful discussions and help with the BARDOT project, Michael Gosney for assisting me on the stomatal conductance measurements, Dr. Manoj Ghaste and Dr. Joshua Widhalm for their collaborative work on the cuticular waxes characterization, Dr. Kendra Erk lab members for the access and training on the use of the goniometer. Finally, I want to acknowledge the institutions that funded my PhD program and research project, Ministerio de Ciencia of Colombia for my PhD fellowship, and Center of Food Safety Engineering – United States Department of Agriculture for the research funding.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The leafy green industry in the United States has positioned the country as the second world leader in lettuce production. Romaine lettuce has been associated with several outbreaks of *E. coli* O157:H7 during the last decade, producing economic losses, as well as negative impacts on human health and consumer confidence. This pathogen has been demonstrated to actively colonize plants and persist for weeks; therefore, dealing with this issue will require an understanding of the interactions happening between plant host, human pathogen, environment, and the resident microbial communities. This research aimed to provide insights for control strategies at the level of prevention in the field, as well as of detection. Based on this, our research goals were: 1. Describe how environmental factors affect the leaf properties and microbial ecology of romaine lettuce plants, as well as the fate of *E. coli* O157:H7 on their leaves; 2. Evaluate the application of a new light scattering technology (BARDOT) developed at Purdue, as an alternative tool to characterize culturable bacterial communities from plants through the recognition of scatter patterns produced by bacterial colonies.

Lettuce plants were grown under three relative humidity (RH) levels: A. 83% (SD= 7.0); B. 62% (SD=9.0); C. 43% (SD= 7.4); significant changes in leaf properties such as responses of stomatal resistance to water loss were observed. RH was the main factor explaining the variation of resident bacterial communities, changes of leaf properties and the fate of *E. coli* O157:H7. Humid condition A produced the lowest bacterial diversity, which was mainly explained by the decreased transpiration rates of these plants, while at the same time this condition allowed the highest *E. coli* O157:H7 growth. Under RH condition C, differences in leaf properties and their distributions across the lettuce leaves were found to be correlated with the composition and localization of the resident microbial communities. *E. coli* O157:H7 growth on leaves was also negatively correlated with the inoculum dose, and it was enhanced on leaf areas with increased stomatal density and size, and decreased leaf wettability. We found that resident bacterial communities are disturbed after the introduction of the human pathogen, *Microbacterium*, and one unclassified genus from the *Rhizobiaceae* family were found as biomarkers of communities where *E. coli* O157:H7 reached higher and lower population counts, respectively.

For the BARDOT technology, three libraries containing 8,418 images of scatter patterns from the nine most abundant bacterial genera from conventional and organic romaine lettuce were created. These libraries covered around 70-76% of the total isolated populations. The training parameters achieved classifiers at genus level with positive predictive values (PPVs) between 90.6-99.8%. The validation with blind samples resulted in sensitivity and average classification accuracy values above 90% for both pure and mixed cultures. The sensitivity and classification accuracy per genus when new lettuce samples were tested, showed values between 51.9-79.1% and 42.9%-100%, correspondingly. Some bacterial genera were identified as challenging for the BARDOT and improvements for the technology have been suggested. BARDOT technology represents a rapid and easy-to-use alternative to conventional microbiological and molecular methods for identification of culturable bacteria.

CHAPTER 1. BACTERIAL COMMUNITIES AND HUMAN PATHOGENS ON PLANTS (ROMAINE LETTUCE): INTRODUCTION AND BACKGROUND

1.1 Human Pathogens on Plants: Plant Pathology's Role in Food Safety

Plants are naturally associated with microorganisms such as bacteria, fungi, and yeasts. These plant-associated microbes live on the host phyllosphere (plant aerial surfaces, i.e. leaf surfaces), on the rhizosphere (the soil-root interface), and in the endosphere (within plant tissues) in a harmless relation with the plant host. Microbial inhabitants of the phyllosphere and rhizosphere are denominated epiphytes, while those ones living in the internal tissues of leaves, stems or roots are called endophytes (1–3).

While the majority of microbiota on plants is likely to be harmless, some species are responsible for food spoilage, plant diseases and a few cause serious disease in humans. In the United States, foodborne diseases cause an estimated 48 million illnesses each year, and 9.4 million are caused by a known pathogen (4, 5). The outbreaks (defined as the occurrence of two or more cases of similar illness resulting from the ingestion of a common food) of foodborne illness linked to plant-based products such as sprouts, herbs, vegetable crops, fruits, grains and beans have represented around 30-40% of the total reported outbreaks in the last 10 years in United States (6). Specifically, outbreaks associated with leafy vegetables increased from 13% during the period 1998-2008 (7, 8), to 26.3% during the period 2004-2012 (9).

The increased number of outbreaks linked with leafy greens could be explained because of the growing consumption of fresh produce due to current concerns about a healthy lifestyle, larger scale production and distribution, as well as enhanced surveillance by public health officials (9, 10). This great emergence of outbreaks associated with fresh produce is not only a problem of public health but also triggers social impacts such as the disturbance of consumer confidence, which affects the economy of the industry (11).

The leafy green industry in the United States has positioned the country as the second world leader in lettuce production, with a total annual production around 8,000 million pounds. The value of U.S. lettuce production in 2015 totaled nearly $1.9 billion, making lettuce the leading vegetable crop in terms of value (12). However, leafy greens such as Romaine lettuce, are highly susceptible

to contamination by human pathogens since they grow low to the ground and therefore have more probabilities of being in contact with contaminated sources as soil, water, manure, animal feces, among others (13–15). Additionally, leafy greens are mostly consumed raw, without a cooking step which might eradicate the bacteria present.

One important foodborne human pathogen which causes around 73,000 illnesses, 2,200 hospitalizations, and 60 deaths annually in United States is *Escherichia coli* O157:H7 (16). This bacterium was first described as an enteric human pathogen associated with hemorrhagic colitis in 1982 (17). In the beginning this pathogen was associated with contaminated food from bovine sources such as beef, milk, etc. (18). However, several cases of *E. coli* O157:H7 human infections linked to consumption of leafy greens have been reported since 1990 (19, 20), and lettuce has been the most implicated type of fresh produce (21, 22).

These numerous outbreaks have demonstrated the fact that plants can be vectors for human pathogens (23). Therefore, several studies have focused on the steps of the life cycle of human bacterial pathogens such as *E. coli* O157:H7, which are life-threatening by using important crop plants as vectors (24–27).

In general, animals and humans are the reservoirs of enteric pathogens, therefore these bacteria reach the agricultural environment through their feces. Feces may contain between 10^2 and 10^5 CFU/g of *E. coli* (28). Several routes of crop contamination have been described, such as soil, water, compost, manure, animals, insects, seeds, etc. (13, 15, 29). *E. coli* O157:H7 could survive in soil, at low temperatures (25ºC) and high soil moisture (100%) for more than 80 days (30). Additionally, the surface water can be contaminated by direct deposition of fecal matter or by dispersal following rain or flood events. Finally, this water may be used for irrigation, or preparation of pesticides and fertilizers (31).

Consequently, once human bacterial pathogens have reached a route of dispersal which leads them to plants, they can attach, establish and colonize seeds, substomatal chambers, stomata, trichomes, cuticle cracks, and intercellular spaces (32), in the same way that plant pathogens do.

Recent studies have demonstrated that some bacterial human pathogens such as *Salmonella* spp. and *Escherichia coli* O157:H7 not only colonize plant surfaces, they also can internalize within the plant tissues (25, 33, 34). For instance, the internalization of these pathogens was

described in lettuce and spinach plants after the cultivation of contaminated seeds (35), and in lettuce plants grown in contaminated soils and hydroponic systems (36). Therefore, the suitable approach to this food safety issue is not just the superficial sanitation of the crops during harvest and postharvest. It is necessary to understand the microbial ecology of how these human bacterial pathogens can adapt to establish themselves as plant-associated bacteria, overcoming the harsh environmental factors and becoming part of the native microbial communities already present in the fresh produce. Application of plant pathology fundamentals such as the plant disease triangle and the plant disease cycle to this food safety issue, will help to develop efficient management and control strategies.

1.2 Factors Influencing Plant Colonization by Human Bacterial Pathogens

The understanding of what factors of the plant disease triangle and plant disease cycle make plants more susceptible to human bacterial pathogen colonization and persistence could lead to better methods of control. Factors such as human pathogen genotype, plant genotype, physiological state of the plant (leaf age, tissue damage, nutrient availability), abiotic conditions (fluxes in temperature, moisture, solar radiation, osmotic stress, desiccation, precipitation, wind), plant immune responses against human bacterial pathogens, and interactions of the human pathogen with native bacteria are thought to modulate the enteric pathogen's process of colonization (27).

Human pathogens require special adaptations such as efficient use of carbon sources of plants (sucrose, fructose and glucose), DNA repair systems, pigmentation, and production of extracellular polysaccharides, to survive the harsh environment of plant surfaces. The intestinal environment offers a high nutrient content, anaerobic conditions and shield from solar rays, while plant surfaces are poor in nutrients, mostly aerobic, exposed to UV radiation and osmotic stress (14). This adaptation process was observed through the transcriptional profile of *E. coli* O157:H7 associated with lettuce leaves, where genes involved in stress and starvation responses were upregulated, while genes mediating cellular energy, metabolism and transport were downregulated (37). For that reason, the genotype of the enteric bacteria may determine divergent levels of plant colonization between strains of the same species. For instance, *E. coli* K-12 reached higher levels

of colonization in the lettuce rhizoplane with more extensive transcriptional changes (~7.6% of the total genes) than *E. coli* O157:H7 (~1.4% of the total genes) (38).

Plant colonization by human pathogens is similar to colonization by plant pathogens. First, these bacteria are attracted to the crop plants by root exudates, free water accumulation, microclimates, shelter from UV rays, nutrient-rich secretions of glandular trichomes, and nutrients from photosynthesizing guard cells (stomata) (13, 39, 40). Then, bacterial factors such as flagella, fimbriae, curli, cellulose production, and capsule are indispensable for the attachment process including motility, sense of external wetness, adhesion to surfaces, biofilm formation and stress tolerance (13, 32, 41). For instance, *E. coli* O157:H7 utilizes the same molecular mechanism via EspA filaments to colonize the mammalian intestine and to adhere to the plant phyllosphere (42). Furthermore, strains of *E. coli* O157:H7 mutated in the flagellin gene or in the ATPase required for translocation through the type 3 secretion system (T3SS), were significantly reduced in their ability to adhere to spinach and lettuce leaves (41). Similar results were observed with mutations in proteins involved in biofilm formation such as *ycfR* in *E. coli* O157:H7 (37, 43) and *sirA* in *Salmonella* species (44).

The internalization in plant tissues might occur by two routes: natural openings such as stomata, hydathodes, lenticels, and lateral roots or uptake through the root system along with water. Additionally, damage of tissues such as trichome collapse, leaf tip burn lesions, soft rot pathogens and insects are contributing factors (33, 34). After that, the survival as endophytes depends on the interactions with the plant immune response by overcoming the recognition of pathogen-associated molecular patterns (PAMPs), the induction of stomata closure, and the activation of PR genes (16, 45). Also, the bacterial Type 3 Secretion System (TTSS) plays a role in suppressing the plant defense (46), but also in being recognized by the plant immune system. Differences in this plant-human pathogen interaction have been found for the two most popular bacterial pathogens *E. coli* O157:H7 and *S. enterica* serovar Typhimurium. For instance, *E. coli* O157:H7 has been shown to produce a prolonged stomatal closure in Arabidopsis and lettuce leaves even under high relative humidity and light stimulus, while *S. enterica* induced a transient closure (45, 47). Additionally, *E. coli* O157:H7 induced a stronger activation of plant immunity when Arabidopsis leaves were infected, causing higher expression levels of the gene PR1 compared to *S. enterica* (45). Previous studies have reported suppression of plant defense response

caused by the TTSS of *Salmonella*, since TTSS mutants reached lower bacterial populations and induced stronger symptoms in plants (46, 48–50).

In this human pathogen-plant interaction, the genotypic and phenotypic characteristics of the plant host can modulate the colonization potential. Quilliam *et al.* (51) demonstrated that the metabolic activity of *E. coli* O157:H7 varied significantly among different lettuce cultivars and it was totally dependent on the host presence. This cultivar effect on *E. coli* O157:H7 was observed in leaves and rhizosphere, and the metabolic activity of the human pathogen in the rhizosphere was undetectable after the removal of the host plant (51). These findings suggest that plant compounds, root architecture, root exudates and microbial communities may play a role in the differential effects that these cultivars had on *E. coli* activity. Moreover, spinach varieties with differing leaf topologies such as rough-surface, semisolid-leaf and cordate-leaf exhibited differences in the *E. coli* O157:H7 cell numbers after leaf drop inoculation. The rough-surface variety resulted in higher bacterial numbers which may be associated with the more prominent ridges and valleys providing protected niches for survival and replication (52). However, not only leaf topology can influence the bacterial growth, also the characteristics that affect whole-plant architecture such as plant size and head size, and the physiochemical properties such as levels of soluble carbohydrate, calcium, leaf wax, water content, and phenolic compounds were found as significant factors in differentiating bacterial populations in the lettuce phyllosphere (53). Moreover, the physiological state of the plant in terms of leaf age influences *E. coli* O157:H7 populations; younger leaves were found to be more susceptible to contamination with human pathogens (54).

Finally, the plant microbiota interactions play a role in the colonization or inhibition of human enteric pathogens in leafy greens. When human pathogens land in the plant environment, they have to compete with the resident bacterial communities for resources and adaptability in this ecological niche (55). For instance, Cooley's study (56) showed that the presence of the epiphyte *Wausteria paucula* enhanced the survival of *E. coli* O157:H7; on the other hand, a possible diffusible factor produced by *Enterobacter asburiae* decreased the human pathogen's growth on in vitro lettuce plants (56). Some studies have shown how the competition among soil microbiota can adversely affect the survival of *Salmonella* Newport and *E. coli* O157:H7 on *Arabidopsis thaliana* (57). Others found that the species richness of natural microbiota on lettuce cultivars was

negatively correlated with the *Salmonella* population (39). Significant progress has been made to understand how these human pathogens interact with native members of the leaf microbial community (58–60), however many questions about how external factors such as environment, plant genotype and human pathogen genotype can modulate these interactions still remain.

Research studies in this area are highly encouraged because there is increasing evidence that a healthy population of epiphytic and endophytic species protects host plants from invasion by phytopathogens (61–63); therefore, it is feasible that through the characterization and understanding of the roles of these plant native bacterial communities on Romaine lettuce, we will be better able to predict and protect against plant invasion by human pathogens too.

1.3 Bacterial Communities Associated with Romaine Lettuce Leaves

Leaf surfaces can be colonized by up to 10^7 microorganisms per cm^2 (64). Lettuce phyllosphere has reported approximately 10^5 CFU of aerobic bacteria per gram of leaf tissue (65). Next-generation sequencing studies of bacterial communities associated with lettuce leaves have worked with entire leaves, assuming that the communities are distributed spatially in a homogeneous manner (66–68). However, the microbial colonization of the plant phyllosphere is not distributed uniformly across the entire leaf surface. This distribution is affected by the presence and density of leaf structures such as stomata, trichomes and veins (64). Also, the survival of the phyllosphere microbiota is limited by the availability of carbon and nitrogen sources, which are not uniform around the leaf (69), and by the amount of usable water which is favored by high relative humidity (57, 70, 71).

At a bigger scale, geographic location, environmental factors and plant genotypes have been proposed as the main drivers of bacterial community composition on leaves (53, 60, 67). The 'core' community of lettuce leaves has been determined to be composed of *Pseudomonas, Bacillus, Massilia, Arthrobacter* and *Pantoea*, and the geographical location was an important determinant of community composition (67). All these mentioned studies (67, 68, 72–74) have described the bacterial communities through pyrosequencing (454). They have covered the topic of variability in bacterial community composition, due to factors such as organic and conventionally grown lettuce (68, 74), time, space and environment (67), season, irrigation method and presence of *E. coli* O157:H7 (72), using field or laboratory-grown lettuce (73). Dees *et al.* (2015) was the first

study to determine bacterial communities associated with leafy greens across plant development and time, assuring a deeper insight into the phyllosphere composition by using Illumina-based 16S rRNA gene sequencing (66). However, the experimental design of these studies is diverse, and their results cannot be compared. For instance, Jackson *et al.*(68) and Leff *et al.* (74) used lettuce heads from grocery stores without biological and technical replicates in the first case, and with only biological replicates in the second case. On the other hand, Rastogi *et al.* (67) described the geographical and seasonal variation of bacterial communities from romaine lettuce grown in Salinas Valley and Yuma Desert using two samples per field with two technical replicates each.

Studies such as Hunter *et al*. 2010, addressed the influence of the plant genotype and leaf properties (topography, head and leaf size) on the bacterial population diversity in lettuce (53). However, this study did not use second-generation sequencing (clone library sequencing was used), the leaf samples were pooled from different growth stages and positions in the plant architecture, and it did not explore the topographical differences of leaf features at a small scale in each leaf, such as the distribution of veins, stomata, hydathodes, trichomes, etc.

While it is well established that there is significant variability of bacterial community composition between plant species (inter-species) and between plants (intra-species), it is unclear how the biotic and abiotic factors interact to influence the microbiota composition and its resistance to change against colonization of exogenous microorganisms such as human pathogens on plants. Additionally, all these studies have taken the lettuce leaf as sample unit, providing only a general snapshot of the differences in bacterial communities without a small-scale analysis of the "micro communities" present across an entire leaf.

To understand the interactions between environment (abiotic factors), plant, microbial communities and human pathogen with their implications on food safety, new studies have started to apply the "omics" sciences (37, 75, 76). Transcriptional profiling was used to identify genes that are differentially expressed during attachment, interaction, and survival of *E. coli* O157:H7 on lettuce leaf surfaces (37). Also, proteomic analysis has offered new insights in to the protein content of fresh-cut lettuce during the storage process (76). Finally, comparative metagenomics have been used to study the alteration of the soil microbiota in lettuce as a response to different fertilizers (75).

As conclusion, several studies have revealed important aspects of the microbial ecology in lettuce and the possible role of native bacterial communities as modulators of human pathogen plant colonization. However, newer molecular techniques are currently available which could improve our understanding in this topic, to be able to address research priorities such as what elements in the phyllosphere or rhizosphere encourage or discourage human pathogens on plants, what epiphytic phylloplane microbes can trigger plant resistance, and what plant cultivars are more resistant to human pathogen colonization, among others (23).

1.4 New Technologies to Study Culturable Bacterial Communities Associated with Plants

The characterization of bacterial communities can be performed through culture-dependent or culture-independent methods such as Next-Generation Sequencing (NGS) (68, 74, 77). Although NGS offers a better approximation of the real microbial diversity, it does not allow to isolate and preserve the identified bacteria for future evaluations, for instance potential phenotypes for biocontrol. Additionally, in the field of food safety, both big and small fresh produce growers aim to assure the quality of their products; hence, the availability of more simple, rapid, accessible and inexpensive technologies to identify culturable bacterial populations is desirable.

The identification of culturable bacterial populations usually involves multiple steps after their isolation such as the establishment of pure culture, the performance of biochemical tests or polymerase chain reaction (PCR) and Sanger sequencing. These methods add more time to the identification process and in the last case demand access to resources of molecular biology and external sequencing services. A laser light-scattering technology called BARDOT (Bacterial Rapid Detection using Optical Scatter Technology) was created as a direct, real-time and label-free identification method (78). Compared with other direct microbial identification methods such as matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS) (79), BARDOT is applied directly on the bacterial colonies growing on the culture media and preserves the viability of the microorganisms.

The BARDOT system uses a 635-nm laser, that interrogates the whole volume of the colony, generating a scatter pattern influenced by refractive indices, extracellular material, density and individual shape of the bacterial cells (80). These patterns have been used as a fingerprint to be

compared against a pre-trained scatter-signature image database which can detect specific targets such as foodborne bacterial pathogens (78, 81–83), differentiate serovars (84), serogroups (85), and virulence gene-associated mutant colonies (86). However, a broader application such as the characterization of entire culturable bacterial communities from environmental samples has not been tested yet using the BARDOT technology.

CHAPTER 2. EFFECT OF EXPERIMENTAL METHODS ON STUDIES OF BACTERIAL COMMUNITIES AND HUMAN PATHOGENS ON PLANTS

2.1 Abstract

Experimental and biological factors affected the results obtained of growth of *Escherichia coli* O157:H7 on the phyllosphere of romaine lettuce. First, the population dynamics of this pathogen on harvested leaf samples which were cut, inoculated and incubated in a humid petri dish chamber were significantly different than the growth observed on inoculated leaves of living plants, even though they also were under humid growth chamber conditions. Secondly, not only the environmental conditions at the time of inoculation determine the fate of *Escherichia coli* O157:H7 on leaves, also the conditions used for plant growth are crucial. Lettuce age produced significant differences in the growth of the pathogen; we recommend the use of lettuce plants of same age and close to harvest time if the effect of age is not the research question. Sample processing through blending of lettuce leaves was demonstrated to decrease the number of viable cells of *Escherichia coli* O157:H7 that can be recovered for plate counts. Additionally, this blending method is not suitable for preparation of DNA for next-generation sequencing of the bacterial 16S rRNA since the contamination with plant DNA is extremely high. Co-amplification of chloroplast and mitochondrial DNA is a big challenge for studies of plant-associated bacteria; the use of plant samples with higher bacterial population numbers was demonstrated to improve the number of bacterial reads obtained. Overall, our results suggest that all research studies in this area must optimize their experimental methods prior to performing the research study to avoid results being driven by the effect of the methodology and not by the real performance of the pathogen on plants.

2.2 Importance

Research about the phytobiome (organisms that influence or are influenced by the plant or plant environment) could lead to a better understanding of how human-pathogenic bacteria can become plant-associated microorganisms. In the end, this knowledge will help to design better strategies of control for our current food safety issues. However, these types of studies face challenges in their methodological design and approach, such as difficulty of working with

threatening human pathogens under real field conditions, simulation of environmental conditions that allow inoculation and growth of the pathogen on plants, reproducibility and comparison of results among published research due to multiple sources of variation such as genotype of the human pathogen, genotype of the plant host, age of the plant, method and dose of inoculation, sample processing methodology, etc. Additionally, there is a big limitation in studies of bacterial communities from plants caused by the high similarity between the DNA sequences of chloroplasts, plant mitochondria and bacterial 16S rRNA. For those reasons, this study evaluated how the experimental design and methods can affect the results of evaluating *Escherichia coli* O157:H7 growth on romaine lettuce leaves and the output of MiSeq sequencing of bacterial communities, demonstrating the importance of a previous optimization of materials and methods in this type of research.

2.3 Materials and Methods

2.3.1 Bacterial Strain and Culture Conditions

The strain *Escherichia coli* O157:H7 B6-914, which does not produce Shiga-like toxins Stx1 and Stx2, was used for laboratory safety advantages (87). To track the inoculated pathogen in lettuce plants, this strain was transformed with the plasmid pGFP (cDNA vector, Clontech, Mountain View, CA, USA) to express green fluorescent protein (GFP) and an ampicillin resistance gene for selection (88). The presence of GFP in *E. coli* O157:H7 had no effect on its behavior when compared with the parent strain (89) and the lack of toxin genes had no influence on the bacterial growth (90).

Inoculum was prepared by streaking frozen GFP-labeled *E. coli* O157:H7 onto Luria Bertani (LB) agar supplemented with 100 µg/ml of ampicillin (LBA-Amp) and incubating at 37˚C for 24 h. An individual colony was transferred to one milliliter of LB broth with 100 µg/ml ampicillin (LB-Amp) and incubated at the lower temperature of 30˚C to acclimatize the bacterial strain for subsequent plant inoculations. Incubation of this liquid culture was carried out for 7 h in constant agitation (250 rpm) to the early stationary phase of growth, when optical density at 600 nm (OD_{600}) was around 1.5. For accurate quantification of *E. coli* O157:H7 inoculum, this liquid culture was serially 10-fold diluted in phosphate buffer (PB) 0.1M pH 7.0 and plated onto LBA-Amp. Plate count was performed after 16 h of incubation at 37˚C, while the liquid culture was

stored at 4˚C. Based on this first quantification of viable culturable cells, our liquid cultures ranged between 1.0 – 9.9 x 10^8 CFU/ml, then dilutions were freshly prepared in PB 0.1M pH 7.0 to yield 10^3 and 10^6 CFU in 50 µl. These inoculum doses were plated onto LBA-Amp right before being used to inoculate lettuce plants, so the exact number of CFU placed onto lettuce leaves was known.

2.3.2 Romaine lettuce source

Two cultivars of romaine lettuce were used, Green Tower (GT) and True Heart (TH). Seeds were purchased from Urban Farmer (Westfield, IN, US) and Synergene Seed & Technology, Inc. (Salinas, CA, US), respectively. The seeds were sown in trays using a soil mix prepared in a 2:1 ratio by volume of PRO-MIX FLX growing medium (Premier Tech Horticulture, PA, USA) and field soil (sandy loam collected from the top 10 cm of a conventional agricultural field in Indiana). Three weeks after germination, seedlings were transplanted into pots (15.5 cm x 17.8 cm) containing the same soil mix with 2 grams of fertilizer (Osmocote® Smart-Release®). Plants were grown and tested between 6-8 weeks after sowing the seeds, using always the fully developed first and second emerged leaves. Plants were watered with the frequency required to avoid the soil mix to dry under each environmental condition. Water was added directly on top of the soil avoiding its accumulation in the tray placed under the pots.

2.3.3 Greenhouse Conditions for Cultivation of Romaine Lettuce and Inoculation of *E. coli* O157:H7

Greenhouse facilities allowed light-controlled conditions operated on a daily cycle of 16 h of light and 8 h of darkness, with temperature-controlled conditions set to 22-24˚C. Relative humidity cannot be controlled, and RH measurements ranged between 20% to 50%.

2.3.4 Growth chamber conditions for cultivation of romaine lettuce and inoculation of *E. coli* O157:H7

A growth chamber (Percival AR-75L) was programmed for a 16-h light cycle (130 µmol m^{-2} s^{-1}) and 8-h dark at constant temperature of 24˚C (SD= 1˚C), and constant relative humidity of 83% (SD= 7.0). Environmental conditions were recorded with a HOBO MX2301 Temp/RH data logger.

2.3.5 Measurement of *E. coli* O157:H7 growth on leaves of romaine lettuce grown and inoculated under different environmental conditions

To evaluate which environmental conditions allow the inoculation and optimal growth of *E. coli* O157:H7 on leaves, romaine lettuce plants were cultivated and inoculated under greenhouse, laboratory and growth chamber environments. Four experiments were performed:

Test #1. Plants were grown and inoculated under greenhouse conditions. A total of 8 leaves from two GT and two TH plants were divided into 12 testing areas (Figure 2.1) using petroleum jelly for delimitation, then each area was drop-inoculated with 50 µl of PB 0.1M pH 7.0 containing 10^2 CFU of *E. coli* O157:H7. The bacterial suspension was dispensed with a pipette and sterile tip by randomly dispersing droplets inside each area. Bacterial inoculum was incubated for 24 hours on leaves.

Test #2. Plants were grown under greenhouse conditions; when the time of growth was completed, they were transferred to the laboratory and were inoculated after a 24-h period of acclimatization. The environmental conditions in the laboratory cannot be controlled, but they were recorded as temperature of 23˚C (SD= 1˚C) and RH of 20% - 40%. Three different methods of inoculation were assessed (Figure 2.2): by Plant which was performed on leaves attached to living plants (same as Test #1); by Leaf which was performed on harvested leaves dipped into water; and by Square which used a piece of leaf excised with a razor blade that is incubated inside a petri dish with wet filter paper. All samples inoculated through these three methods used 50 µl of PB 0.1M pH 7.0 containing 10^2 CFU of *E. coli* O157:H7. Bacterial inoculum was incubated for 24 hours on leaves.

Figure 2.1 Twelve testing areas for drop-inoculation of *Escherichia coli* O157:H7 and 16S MiSeq sequencing.

Figure 2.2. Three methods of inoculation of *Escherichia coli* O157:H7 assessed in Test #2.

Test #3. Plants were grown under greenhouse conditions; when the time of growth was completed, they were transferred to growth chamber conditions and were inoculated after a 24-h period of acclimatization. A total of 18 GT plants were drop-inoculated only on testing area A using three leaves per plant to test different incubation times. Inoculation was performed as previously described in Test #1 but using 10^4 CFU of *E. coli* O157:H7; bacterial inoculum was incubated for 16, 24 and 48 hours on leaves.

Test #4. Plants were grown and inoculated under growth chamber conditions. A total of 6 GT and 5 TH plants were drop-inoculated on testing areas A, B, C, and D. Inoculation was performed as previously described in Test #1 but using 10^4 CFU of *E. coli* O157:H7; bacterial inoculum was incubated for 12, 16, 24 and 48 hours on leaves.

2.3.6 Leaf sample processing for plate counts of *E. coli* O157:H7 CFU

Samples from the previous four tests were processed as follows: To avoid disruption of the *E. coli* O157:H7 cells on leaf surfaces, harvested leaves were placed onto a flat surface, testing areas were immediately excised using a razor blade and transferred separately to a sterile 50-ml conical tube. Afterwards, 30 ml of PB 0.1M pH 7.0 were added and samples were sonicated for 7 minutes using a Branson 5800 ultrasonic cleaner water bath (Branson Ultrasonics Corporation, Danbury, CT). Dilutions from wash buffer were prepared and plated onto LBA-Amp. After an incubation at 37˚C for 24 h, the plate counts were recorded to calculate the growth of *E. coli* O157:H7.

2.3.7 Recovery of *E. coli* O157:H7 CFU from inoculated leaves through different sample preparation methods

Three sample preparation methods were evaluated as follows to determine their efficiency in the recovery of *E. coli* O157:H7 cells from inoculated lettuce leaves: sonication A= 2.5 minutes performed through 5 cycles of 30 seconds of sonication and 30 seconds rest, sonication C= 7 continuous minutes, and Blending= 20 seconds. Sonication was performed in a Branson 5800 ultrasonic cleaner water bath (Branson Ultrasonics Corporation, Danbury, CT) and Blending was done using a Ninja professional blender. Samples were tested after 15 minutes, 16 h, 24 h and 48 h of inoculation of leaf pieces with 10^4 - 10^6 CFU of *E. coli* O157:H7. Samples tested after 15 minutes were incubated on an open tray inside a laminar flow hood to evaluate the recovery of the same number of inoculated CFU avoiding the effect of bacterial growth. Samples tested after 16 h, 24 h and 48 h were incubated inside wet petri dishes. Each treatment consisted of 10 biological repetitions.

2.3.8 Measurement of *E. coli* O157:H7 growth on leaves after different incubation times

E. coli O157:H7 growth was determined at different times after inoculation using leaf pieces incubated in wet petri dishes and plants grown and inoculated under growth chamber humid conditions. 10^4 and 10^5 CFU of *E. coli* O157:H7 were inoculated and samples were prepared for plate counts by sonication for 7 minutes. Each incubation time consisted of 10 biological repetitions for the experiment with wet petri dishes, and 24 biological repetitions for the experiment with plants in the growth chamber.

2.3.9 Measurement of *E. coli* O157:H7 growth on first-emerged leaves from plants of different ages

Six fully developed leaves from three GT and three TH plants grown in growth chamber conditions were drop-inoculated on testing areas A, B, C and D when plants were 5, 6 and 7 weeks old. Inoculation was performed under growth chamber conditions, following the procedure previously described in Test #1 but using 10^4 CFU of *E. coli* O157:H7. Samples were processed after 12 and 16 hours of incubation through sonication for 7 minutes. Each leaf age was evaluated with 24 biological repetitions.

2.3.10 Statistical analysis of *E. coli* O157:H7 growth

E. coli O157:H7 populations (CFU per testing area) obtained from plate counts were expressed as *E. coli* O157:H7 growth change (log CFU) by calculating the logarithmic change between the number of CFU inoculated and the number of CFU after the time of inoculation on testing areas. Comparison of bacterial growth among the experimental methods and conditions was conducted by analysis of variance (ANOVA) followed by Tukey's multiple comparison test when data were normally distributed and by Kruskal-Wallis test followed by the Steel-Dwass method when data was not. Statistical tests were performed with JMP software (SAS Institute Inc., Cary, NC).

2.3.11 Evaluation of type of lettuce and sample preparation method for DNA extraction and 16S MiSeq sequencing

Two romaine lettuce plants, one TH grown under our greenhouse conditions and one organic head bought from grocery stores were used to evaluate two sample preparation methods for DNA extraction and MiSeq sequencing. Blending method was performed by blending twenty-five grams of leaf tissue into 225 ml of PB for 20 seconds. Washing method was performed by shaking twenty-five grams of leaf tissue into 225 ml of PB for 30 minutes at 300 rpm. Two type of samples were collected from the washing method: the washing buffer which was centrifuged for concentration, and the washed leaf tissue which was blended as previously described. For DNA extraction, 500 µl from each sample type were used with the DNeasy PowerSoil DNA Isolation Kit (Qiagen, MD, USA). Library preparation was carried out using the Illumina 16S Metagenomic Sequencing Library protocol with some modifications. A dual indexing protocol with a two-step PCR was performed to amplify the V5 through V6 region of the bacterial 16S rRNA gene and to add dual TruSeq indices. Primer for 16S sequences (underlined portion) with TruSeq adaptor tails are shown: primer 799F (5' ACACTCTTTCCCTACACGACGCTCTTCCGATCT AACMGGATTAGATACCCKG 3') was chosen to exclude chloroplast amplification (91, 92), and primer 1064R (5' GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT CGACARCCATGCASCACCT 3') was chosen as universal primer (93). Amplicons were pooled in molar equivalent concentrations based on qPCR quantification with primers 1.1 (5' AATGATACGGCGACCACCGAGAT 3') and 2.1 (5' CAAGCAGAAGACGGCATACGA 3'). The pool was sequenced and demultiplexed at Purdue Genomics Center using Illumina MiSeq paired-end 250-bp sequencing. Data processing and analysis were performed using packages in R (v 3.6.1), DADA2 (v 1.14.1) and Phyloseq (v1.22.3) (94–96). Reads with unclassified Phylum or identified as Chloroplast or Mitochondria were filtered out. A more detailed protocol can be found in the methods section of chapter 3.

2.4 Results and Discussion

2.4.1 Effect of environmental conditions on growth of *E. coli* O157:H7 on romaine lettuce leaves

Four tests were performed to evaluate what environmental conditions not only during *E. coli* O157:H7 inoculation, but also during seed germination and plant growth of romaine lettuce

allow the successful growth of this bacterial pathogen on leaf surfaces. No growth of *E. coli* O157:H7 was achieved in any of the 120 repetitions when lettuce plants were grown and inoculated under our environmental conditions in a greenhouse (Test #1). First to corroborate if this phenomenon was caused by the climate generated in our greenhouse (mix of temperature, relative humidity, solar radiation, ventilation fans, etc), we carried out new inoculations (Test #2) using the same low inoculum dose, same incubation time and again plants grown in a greenhouse but they were transferred to be inoculated and incubated under laboratory conditions. In Test #2 besides leaves attached to living plants, also harvested leaves dipped into water and leaf pieces inside wet petri dishes were inoculated. As observed in Figure 2.3, there was not a significant difference in *E. coli* O157:H7 growth on leaves attached to living plants when inoculated under greenhouse conditions (Gh/Gh-Plant) or laboratory conditions (Gh/Lab-Plant); however, some outliers finally showing growth were spotted in the Gh/Lab-Plant group. Although the inoculations on harvested leaves (Gh/Lab-Leaf) also had negative mean values, this condition was significantly different showing less death of *E. coli* O157:H7 than the previous ones. Finally, an optimal *E. coli* O157:H7 growth was observed when inoculation and incubation are performed on excised leaf pieces inside wet petri dishes (Gh/Lab-Square).

Tests #1 and #2 demonstrated that both our greenhouse and laboratory conditions were not conducive for optimal growth of this human pathogen after drop-inoculation on romaine lettuce leaves. On the other hand, a very humid environment such as the moist chambers built with petri dishes containing wet filter paper was favorable for the bacterial proliferation. Previous studies similarly have reported that very high relative humidity conditions (100%) can increase *E. coli* O157:H7 numbers up to 7 log CFU/g on intact leaf tissue of *Arabidopsis thaliana* (57). Additionally, this Gh/Lab-Square condition might offer a greater source of nutrients due to possible leakage from the borders of the excised leaf pieces, which has been demonstrated to enhance *Salmonella enterica* attachment, motility, biofilm formation and growth on salad leaves when traces of salad juices are present (97). Our results also underscore how chopped leafy greens might be more susceptible to higher levels of *E. coli* O157:H7 colonization, which can be aggravated by high moisture.

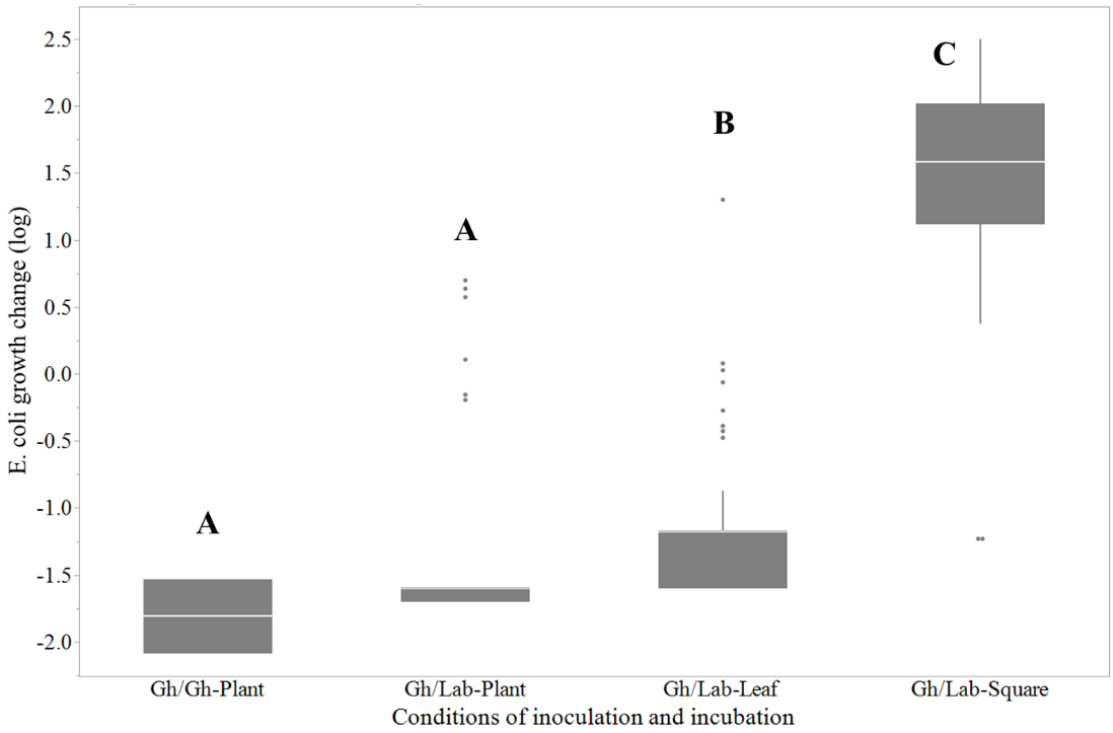


Figure 2.3 Tests #1 and #2: Growth of *Escherichia coli* O157:H7 on greenhouse-grown lettuce plants when inoculated under greenhouse conditions using leaves attached to living plants (Gh/Gh-Plant), or under laboratory conditions using leaves attached to living plants (Gh/Lab-Plant), harvested leaves (Gh/Lab-Leaf), and excised leaf pieces inside wet petri dishes (Gh/Lab-Square). Different letters indicate a significant difference ($P < 0.05$).

However, as our interest was to investigate the *E. coli* O157:H7 growth on living romaine lettuce plants, we performed Test #3 to evaluate if these plants grown under dry conditions in a greenhouse can allow higher growth of the pathogen if they are inoculated with 10^4 CFU of *E. coli* O157:H7 under high relative humidity achieved in a growth chamber (RH; 83% [SD= 7.0]). After 24 hours of incubation, as observed in Figure 2.4, lettuce plants grown in greenhouse did not favor *E. coli* O157:H7 growth on their leaves, no matter the increase of relative humidity when inoculated in a growth chamber. To determine if the lack of growth of the human pathogen was caused by our method of drop-inoculation or by plant-associated factors, Test #4 was performed in which romaine lettuce plants were germinated, grown and inoculated under high RH in growth chamber conditions. Results in Figure 2.4 show that *E. coli* O157:H7 growth was achieved with drop-inoculation of plants which were germinated and grown in humid conditions. Therefore, the drop-inoculation method was kept as our inoculation method. Although Moyne *et al.* (98) found that plants inoculated by drop showed a more rapid decrease of *E. coli*

O157:H7 CFU than plants inoculated with a spray bottle, this study only tested this phenomenon under a low RH of 30%. On the other hand, we were able to achieve bacterial growth on leaf surfaces by drop-inoculation using a higher RH and, most important, this inoculation method offers consistency in the sample's contamination and accurate quantification of how many CFU are placed on each testing area.

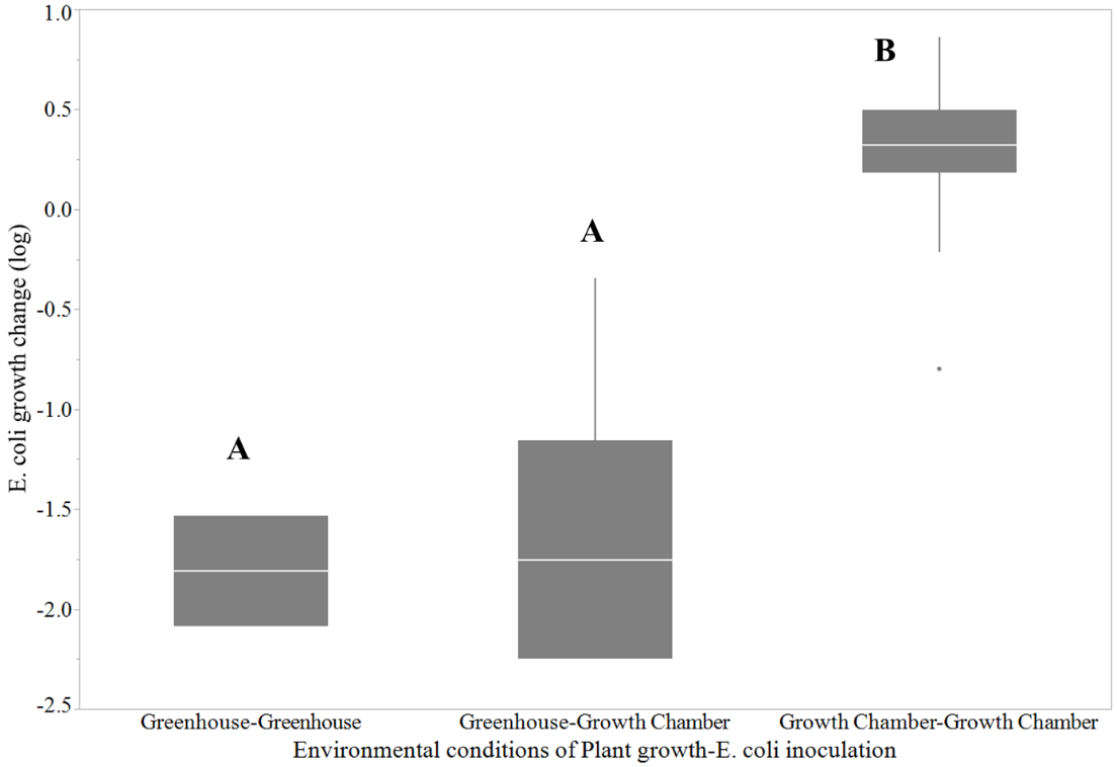


Figure 2.4 Tests #3 and #4: Comparison of *Escherichia coli* O157:H7 growth after 24 hours of inoculation with 10^4 CFU. Leaves of romaine lettuce plants grown in greenhouse conditions were inoculated under the same greenhouse conditions (Greenhouse-Greenhouse) and inoculated under growth chamber conditions (Greenhouse-Growth Chamber). Also, leaves of romaine lettuce plants grown in growth chamber conditions were inoculated under the same growth chamber conditions (Growth Chamber-Growth Chamber). Different letters indicate significant difference ($P < 0.05$).

As the environmental conditions used for plant germination and growth seem to limit the successful growth of *E. coli* O157:H7 on leaf surfaces, more plants grown in greenhouse and growth chamber conditions were inoculated under same the growth chamber humid conditions to be tested at different times of incubation. Results in Table 2.2 show that no matter the incubation

time and the high RH during inoculation and incubation, only plants grown in humid conditions favored *E. coli* O157:H7 survival.

Table 2.1 Growth of *Escherichia coli* O157:H7 on romaine lettuce plants grown under greenhouse and growth chamber conditions and inoculated under these two same conditions

Plant growth condition	*E. coli* O157:H7 inoculation condition	Incubation hours after inoculation	*E. coli* O157:H7 growth change (log CFU)[a]
Greenhouse	Growth Chamber	16	-1.16 ± 0.62 ****
Growth Chamber			0.25 ± 0.59 ****
Greenhouse	Greenhouse	24	-1.81 ± 0.28 B
	Growth Chamber		-1.67 ± 0.57 B
Growth Chamber			0.30 ± 0.36 A
Greenhouse	Growth Chamber	48	-1.49 ± 0.71 ****
Growth Chamber			0.28 ± 0.36 ****

[a] Values are means ± standard deviations of log(CFU after incubation/CFU inoculated). Values followed by a different letter are significantly different ($P < 0.05$). **** Paired values are significantly different ($P < 0.0001$).

These findings suggest that the environmental conditions in which romaine lettuce plants are germinated and grown play an important role on defining plant-associated factors which can contribute to *E. coli* O157:H7 growth on leaf surfaces. For that reason, Chapter 3 of this research thesis studied leaf properties such as hydrophobicity, stomatal density, and composition of bacterial communities of lettuce plants grown under different levels of RH; and Chapter 4 aims to correlate these findings with differential *E. coli* O157:H7 fate on those plants.

2.4.2 Effect of sample preparation method on the recovery of *E. coli* O157:H7 CFU from inoculated romaine lettuce leaves

As plate counts of *E. coli* O157:H7 CFU were used to evaluate the influence of environmental and plant-associated factors on the growth of this pathogen on romaine lettuce leaves, optimization of the method used to process these samples is imperative. In that way, the effects observed on bacterial counts will not be caused by the sample preparation method and can be attributed solely to the experimental treatments applied. Based on that premise, we would expect that the values of *E. coli* O157:H7 growth change (log CFU) were not significantly different among the sample preparation method used. However, Figure 2.5 shows how inoculated leaves

which were processed by the blending method exhibited significantly lower recovery of *E. coli* O157:H7 CFU no matter if they are tested 15 minutes, 16, 24 or 48 hours after inoculation. Also, results obtained by blending are subjected to a higher variance compared with results obtained by sonication (Table 2.3). On the other hand, sonication methods always showed consistent bacterial recovery even if samples were sonicated for 2.5 minutes or 7 minutes.

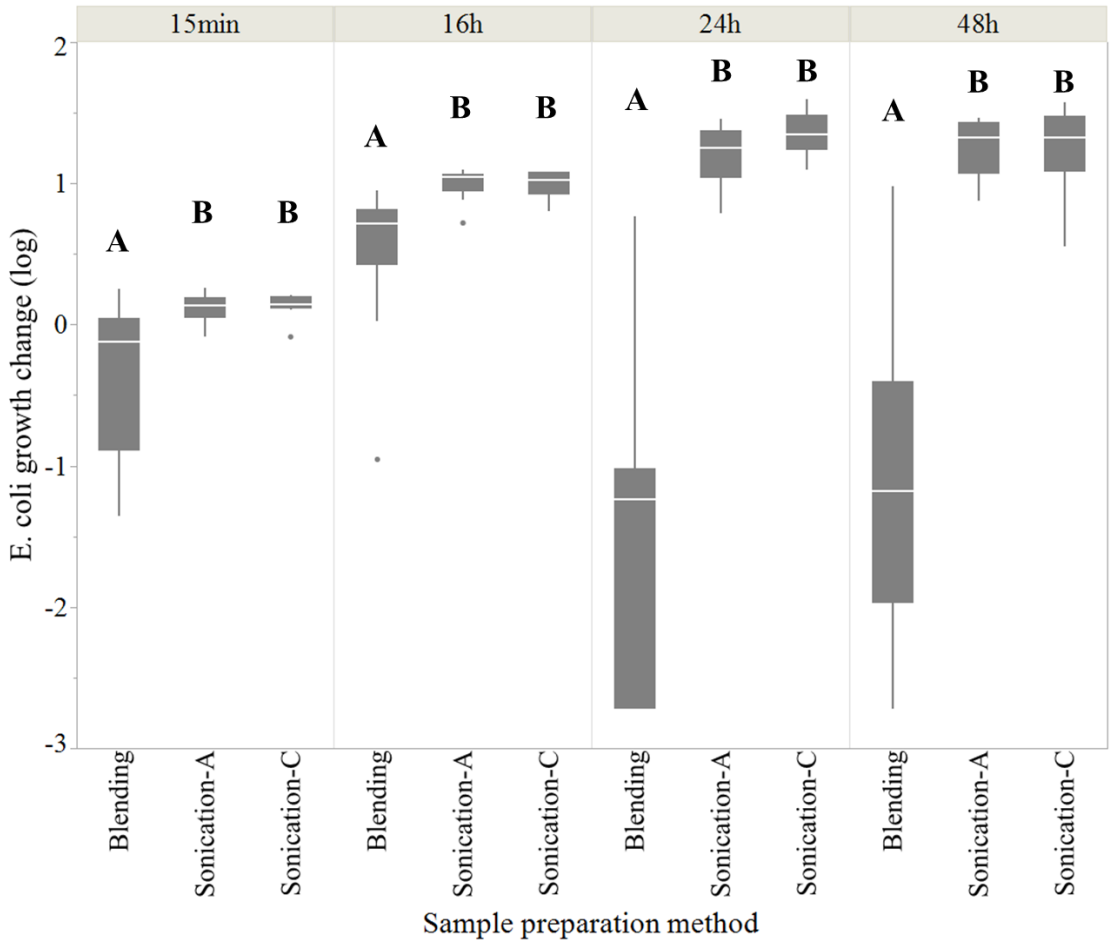


Figure 2.5 Recovery of *Escherichia coli* O157:H7 CFU through different sample preparation methods: Sonication-A (2.5 minutes in cycles of 30 seconds), Sonication-C (7 continuous minutes) and Blending (20 seconds). Inoculated leaves were tested after 15 minutes when incubated in a laminar flow hood, and after 16, 24 and 48 hours when incubated in wet petri dishes. Different letters indicate significant difference ($P < 0.05$).

Table 2.2 Values of *Escherichia coli* O157:H7 growth change (log CFU) from experiment of recovery of *E. coli* O157:H7 CFU through different sample preparation methods.

Sample preparation	*E. coli* O157:H7 growth change (log CFU)[a]			
method	15 minutes	16 hours	24 hours	48 hours
Blending	-0.33 ± 0.54 A	0.52 ± 0.58 A	-1.41 ± 1.09 A	-1.11 ± 1.22 A
Sonication-A	0.12 ± 0.10 B	0.99 ± 0.12 B	1.20 ± 0.23 B	1.25 ± 0.21 B
Sonication-C	0.13 ± 0.09 B	0.99 ± 0.09 B	1.35 ± 0.17 B	1.23 ± 0.37 B

[a] Values are means ± standard deviations of log(CFU after incubation/CFU inoculated). Values in same column followed by a different letter are significantly different ($P < 0.05$).

Additionally, we found that incubation of inoculated leaves inside a laminar flow hood is not useful if growth of *E. coli* O157:H7 on leaf surface is expected. This incubation method was only used here to evaluate the recovery of the same number of inoculated CFU avoiding the effect of bacterial growth, and it needs to be tested after no more than 15 minutes. When inoculated leaves under laminar flow hood conditions were tested after 30 minutes and one hour, only 21.7 ± 9.5 % and 18.0 ± 27.7 % of *E. coli* O157:H7 cells were alive, respectively. These results concur with Kim *et al.* (99) who recovered significantly fewer bacteria from vegetable pieces inoculated with four different foodborne pathogens including *E. coli* O157:H7 when placed inside a laminar flow hood (RH: 40%) than when placed inside a plant growth chamber (RH: 90%) (99). Kim *et al* suggested that these reductions of recovered CFU may be a result of binding of bacterial cells to plant tissue, death of bacterial cells due to exposure to plant antimicrobial compounds or drying after inoculation. According to our results, the most probable cause is the rapid desiccation of the inoculum under the ventilation of a laminar flow hood. This hypothesis is encouraged because our sample preparation method through sonication showed consistent recovery of *E. coli* O157:H7 (attached or not attached to leaf tissue) even 48 hours after inoculation, but only when bacteria were inoculated under favorable environmental conditions of RH such as wet petri dishes (Figure 2.5).

2.4.3 Population dynamics of *E. coli* O157:H7 on romaine lettuce leaves

It is necessary to determine the population dynamics of our strain of *E. coli* O157:H7 on romaine lettuce leaves before selecting a specific incubation time to evaluate future environmental and plant-associated factors. The ideal time to perform future analysis would be during the

exponential phase of growth since any factor impacting the colonization and establishment of *E. coli* O157:H7 on lettuce leaves may be more evident. For that reason, the growth of our strain was tested for several hours after inoculation on lettuce plants grown and infected under humid conditions in a growth chamber. As a comparative reference of *E. coli* O157:H7 population dynamics, leaf pieces inside wet petri dishes also were inoculated and tested. Although we were able to observe growth of the human pathogen on leaf surfaces across the different times of test after inoculation as follows (mean ± SD): 12h: 0.28 ± 0.27 log CFU, 16h: 0.25 ± 0.59 log CFU, 24h: 0.30 ± 0.36 log CFU, 48h: 0.28 ± 0.36 log CFU, this growth was not exponential nor significantly different among times. On the contrary, our same strain exhibited the typical exponential bacterial growth after being inoculated on leaf pieces in wet petri dishes, also the growth change at 16 hours after inoculation (0.99 ± 0.10 log CFU) was significantly different from the ones observed at 24 and 48 hours after inoculation (1.27 ± 0.21 and 1.24 ± 0.29 log CFU, respectively).

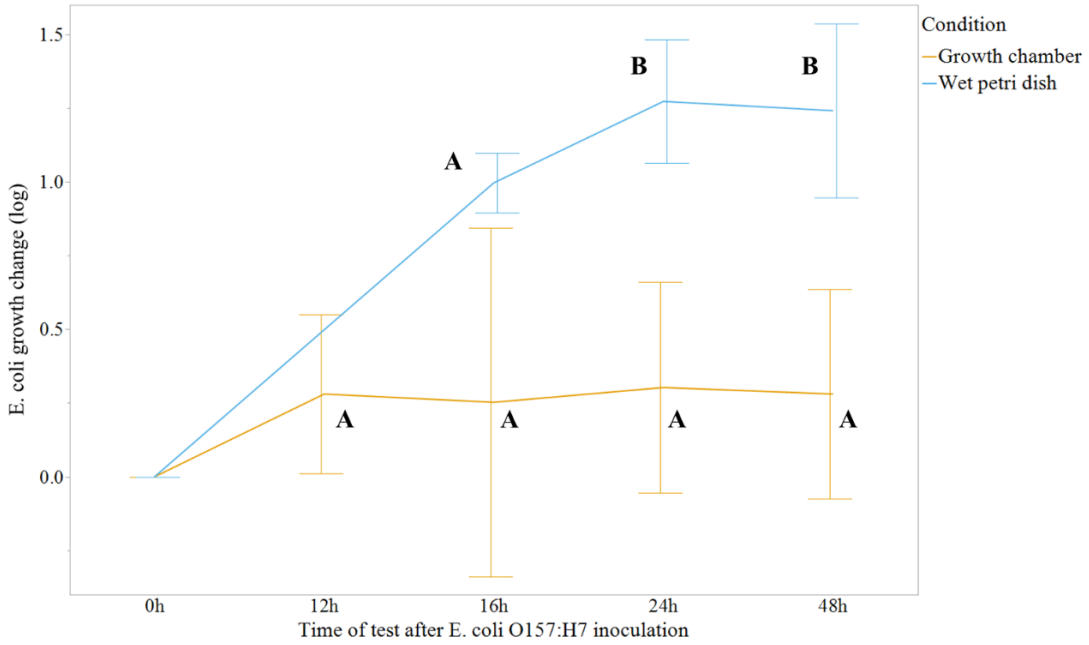


Figure 2.6 *Escherichia coli* O157:H7 population dynamics on leaf surface of romaine lettuce plants grown and inoculated under growth chamber humid conditions (orange line) and on excised leaf pieces from the same plants which were inoculated in moist chambers with wet petri dishes (blue line). Inoculations were performed with 10^4 CFU for the growth chamber condition and 10^5 CFU for the wet petri dishes condition. Lines show the mean of growth change (log CFU) and error bars represent the SD. A different letter within same Condition line means a significant difference ($P < 0.05$) among the times of test after inoculation.

These results suggest, first, that 16 hours is an ideal time for evaluating future factors affecting the growth of *E. coli* O157:H7 on lettuce leaves, because under optimal conditions (wet petri dishes) the bacterial population is undergoing the late logarithmic phase at this time before reaching a plateau between 24 and 48 hours after inoculation. Although under growth chamber conditions this phase is not clear, the growth of *E. coli* O157:H7 at 16 hours showed the higher variation, exhibiting higher growth change in some cases. Secondly, the lettuce phyllosphere of intact leaves is not the most favorable environment for human pathogens such as *E. coli* O157:H7, even though the infection occurred in a humid environment with 10^4-10^5 CFU. It is known that the leaf surface is a harsh environment since it is poor in nutrients, mostly aerobic, exposed to UV radiation and osmotic stress, very different from the human intestinal environment (14, 27). This difference of *E. coli* O157:H7 growth between cut leaves and intact leaves was also observed in a study with harvested leaves from greenhouse-grown romaine lettuce plants; four hours after inoculation, the pathogen's population size increased 4.54-fold on cut leaf pieces and 1.95-fold on intact leaves (100). The fate of *E. coli* O157:H7 in lettuce phyllosphere of field-grown plants has been reported as little or no growth changes with gradual decline after inoculation (101–103). One study evaluated loose-leaf lettuce grown and inoculated in a growth chamber with 70% RH; it was found that nearly half of the pathogen's population died after 2 days of inoculation with high-dose inoculum (6.9 to 8.0 log CFU/ml), while a very small growth increase of 0.08 log CFU was observed after 2 days of inoculation with low-dose inoculum (2.7 to 4.2 log CFU/ml) (104). Therefore, our optimized plant growth and inoculation conditions in a growth chamber allowed to increase the population size of the pathogen, in this way additional factors causing negative impacts on *E. coli* O157:H7 growth can be studied.

2.4.4 Effect of plant age on growth of *E. coli* O157:H7 on fully developed romaine lettuce leaves

To characterize how RH conditions during plant growth can affect leaf properties, fully expanded mature leaves (first and second emerged leaves) were chosen because they offer a larger leaf area for *E. coli* O157:H7 inoculations, and also because stomata are produced early in leaf expansion so their density is at its maximum when leaves are young, and then decreases as the leaf area expands to a relatively stable density (105–107). Additionally, old lettuce leaves have showed 10 times more native bacteria than middle (fifth and sixth emerged leaves) and young (inner rosette)

leaves while not having a significant difference in preharvest population dynamics of *E. coli* O157:H7 when inoculated (54). As leaf age was chosen based on the above reasons, plant age was evaluated to determine if it can affect the growth of *E. coli* O157:H7 on the leaf surface. After 16 hours of incubation, the growth of *E. coli* O157:H7 was significantly different among lettuce plants of diverse age, even though mature leaves (first and second emerged leaves) were used in all cases. On the other hand, although 6-weeks-old plants exhibited higher bacterial growth, this was not significantly different from 7-weeks-old plants when samples were evaluated after 12 hours (Table 2.3). These results suggest that plant age is affecting the speed at which *E. coli* O157:H7 reaches the logarithmic phase (population doubles in number). For that reason, the effect of plant age was not significant at 12 hours when population number still has not doubled (change of log CFU was lower than 0.30), and it becomes evident at 16 hours when the pathogen is in exponential growth on 6-weeks-old and 7-weeks-old plants. Similar results were found when *E. coli* 700728 was inoculated on field-grown romaine lettuce in the Salinas Valley, CA, where bacterial cell numbers were significantly lower on 4-week-old compared to 6-week-old plants (108). Therefore, these results suggest that lettuce plants closer to harvest time are the best models to study additional factors influencing *E. coli* O157:H7 fate and potential risk to the consumer.

Table 2.3 *Escherichia coli* O157:H7 growth after inoculations in a growth chamber using mature leaves (first and second emerged leaves) of plants of diverse age (5, 6, and 7 weeks old) grown in humid conditions.

Lettuce plant age (weeks)	*E. coli* O157:H7 growth change (log CFU)[a]	
	12 hours	16 hours
5	ND[b]	0.002 ± 0.509 A
6	0.370 ± 0.239 A	0.666 ± 0.260 B
7	0.245 ± 0.201 A	0.306 ± 0.269 C

[a] Values are means ± standard deviations of log(CFU after incubation/CFU inoculated). Values in same column followed by a different letter are significantly different ($P < 0.05$). [b] Not determined.

2.4.5 Effect of type of lettuce and sample processing for DNA extraction on yield of bacterial reads from 16S MiSeq sequencing

To evaluate which types of lettuce and sample preparation method yield more bacterial reads than plant reads, blending and washing methods were applied to greenhouse-grown and grocery store-bought romaine lettuce heads. Three types of samples from each type of lettuce were collected as follows: Non-Washed (lettuce was blended), Washed (lettuce was washed as described in methods and blended), and Wash-Buffer (buffer where lettuce was washed). As observed in Table 2.4, blended samples only produced reads from plant DNA; even when the lettuce was not washed bacterial reads accounted for less than 0.1%. The blending method produced a leaf homogenate where all the DNA from bacteria, plant mitochondria and chloroplasts are released, and the number of chloroplast DNA copies can be as high as 10,000 copies per cell (92). Although the sonication method provided a higher percent of bacterial reads, the type of lettuce also has an important effect on how much bacterial data can be obtained. The organic lettuce sample processed by sonication yielded 56.8% of bacterial reads, while the greenhouse-grown lettuce sample processed by the same method only yielded 1.4%. This might be related with the size of bacterial populations present in these two types of lettuce; as described in Chapter 5, culturable bacterial population counts of organically grown romaine lettuce from grocery stores were significantly higher than the conventionally grown lettuce from the same stores ($p<0.01$). Moreover, lettuce plants grown under our greenhouse conditions exhibited lower bacterial population counts than the conventionally grown lettuce from stores. Therefore, it is very important to select optimal sample preparation methods and try to increase the bacterial load in the plant samples which will be sequenced.

2.5 Conclusion

For studies where growth of *E. coli* O157:H7 on lettuce leaves is intended, not only the environmental conditions at the time of inoculation are important, but the environment where the plants are grown is also crucial. Also, results of *E. coli* O157:H7 growth from studies performing the inoculations of the pathogen using harvested leaves or cut leaf pieces, are not comparable with the results obtained when inoculations are performed on living plants.

Biological and methodological factors can affect the results of growth of this pathogen on romaine lettuce such as plant age and sample preparation method for plate counts. Overall,

sonication was demonstrated to be the best sample preparation method for recovery of inoculated *E. coli* O157:H7 cells from lettuce, as well as for DNA extraction and 16S rRNA amplicon next-generation sequencing since it avoids high levels of contamination with plant DNA. The use of plant samples with high bacterial population numbers helps to deal with the co-amplification of chloroplasts and mitochondrial DNA.

Table 2.4 Number of reads obtained through MiSeq 2x250 16S amplicon sequencing before and after filtering Chloroplast and Mitochondria contamination

Sample	Number of reads		Percent of plant contaminant reads	Percent of bacterial reads
	Before filter	After filter		
Non-Washed_Greenhouse	77,451	8	99.99	0.01
Non-Washed_Organic	56,994	32	99.94	0.06
Washed_Greenhouse	11,031	0	100	0
Washed_Organic	72,154	0	100	0
Wash-Buffer_Greenhouse	302,564	4,280	98.59	1.41
Wash-Buffer_Organic	20,250	11,501	43.20	56.80

CHAPTER 3. ENVIRONMENTAL CONDITIONS OF RELATIVE HUMIDITY DETERMINE LEAF FACTORS, COMPOSITION AND SPATIAL DISTRIBUTION OF BACTERIAL COMMUNITIES ACROSS ROMAINE LETTUCE LEAVES

3.1 Abstract

Lettuce is one of the leading vegetable crops in the U.S in terms of value and its consumption and production has been increasing for the last ten years. For that reason, this study evaluated romaine lettuce plants grown under three different levels of relative humidity (RH): A. 83% (SD= 7.0), B. 62% (SD=9.0), C. 43% (SD= 7.4), with constant light and temperature conditions, and adequate watering. Leaf properties and traits such as stomatal density (SD), stomatal size, stomatal conductance (g_s), leaf wettability, amount of cuticular waxes, among others, were significantly different according to the RH conditions. Responses of stomata resistance to water loss were observed from RH conditions A to B, while plants under condition C expressed interesting responses as a balance between stomata resistance and continuation of photosynthetic activity. Under RH condition C, changes of leaf properties and their distributions across the lettuce leaves were correlated with the composition and localization of the resident microbial communities. In general, RH was the main factor explaining the variation of resident bacterial communities of romaine lettuce leaves, humid condition A produced lower bacterial diversity than dryer conditions B and C, and this was mainly explained by the transpiration rates of plants. These results demonstrated how this crop is affected by relative humidity, as well as the microbial ecology of this plant; future studies about additive effects of other environmental factors can reinforce the development of better strategies for crop management, food safety, and response to climate change.

3.2 Importance

Almost 30% of the foodborne disease outbreaks reported annually in the United States are linked to leafy greens, with at least one multistate recall associated with romaine lettuce (109). As observed with plant pathogens, the presence of a healthy and diverse microbiota in the host can help to reduce the susceptibility of the plant to be infected (61, 63, 110). For those reasons, we evaluated the effect of three different environmental conditions of relative humidity (RH) on romaine lettuce leaf development. This study demonstrates how RH conditions can determine the

density of culturable bacterial populations in the lettuce plant, the hydrophobicity of the leaf surface, the development of stomata and the composition of the bacterial communities. Moreover, previous studies (66–68, 72, 74) in the area of microbial ecology in lettuce leaves, have taken the field, the plant or the entire leaf of lettuce as sampling unit to report the native bacterial communities. Here in this study, we approached the spatial distribution of bacterial communities across the lettuce leaf and their association with leaf structures/properties.

3.3 Materials and Methods

3.3.1 Romaine lettuce source and cultivation in growth chamber

Seed germination, plant growth and experimental studies were performed in an environmental growth chamber (Percival AR-75L) which was programmed for a 16-h light cycle (130 µmol m^{-2} s^{-1}) and 8-h dark cycle at constant temperature of 24˚C (SD= 1˚C). Three different conditions of relative humidity (RH) were evaluated: A. 83% (SD= 7.0), B. 62% (SD=9.0), C. 43% (SD= 7.4). Conditions A and B were achieved through the growth chamber performance, while condition C required the introduction of an external dehumidifier. Environmental conditions were recorded with a HOBO MX2301 Temp/RH data logger (Table 3.1, Figure 3.1).

Table 3.1. Environmental data recorded with a HOBO MX2301 Temp/RH data logger from the growth chamber conditions A, B and C.

Condition	Planting date	Inoculation date	Relative humidity (%)				Temperature (˚C)			
			Max	Min	Median	Mean	Max	Min	Median	Mean
A	11/01/19	12/14/19	96.9	51.0	81.1	83.0	26.4	21.5	24.0	24.2
	11/27/19	01/21/20								
	12/04/19	01/26/20								
	12/22/19	02/05/20								
B	02/26/20	04/09/20	91.5	32.1	62.2	61.6	26.0	20.6	23.8	23.8
	03/02/20	04/17/20								
C	05/26/20	07/08/20	71.7	30.2	41.9	43.1	26.0	18.8	23.7	23.6
	06/08/20	07/21/20								

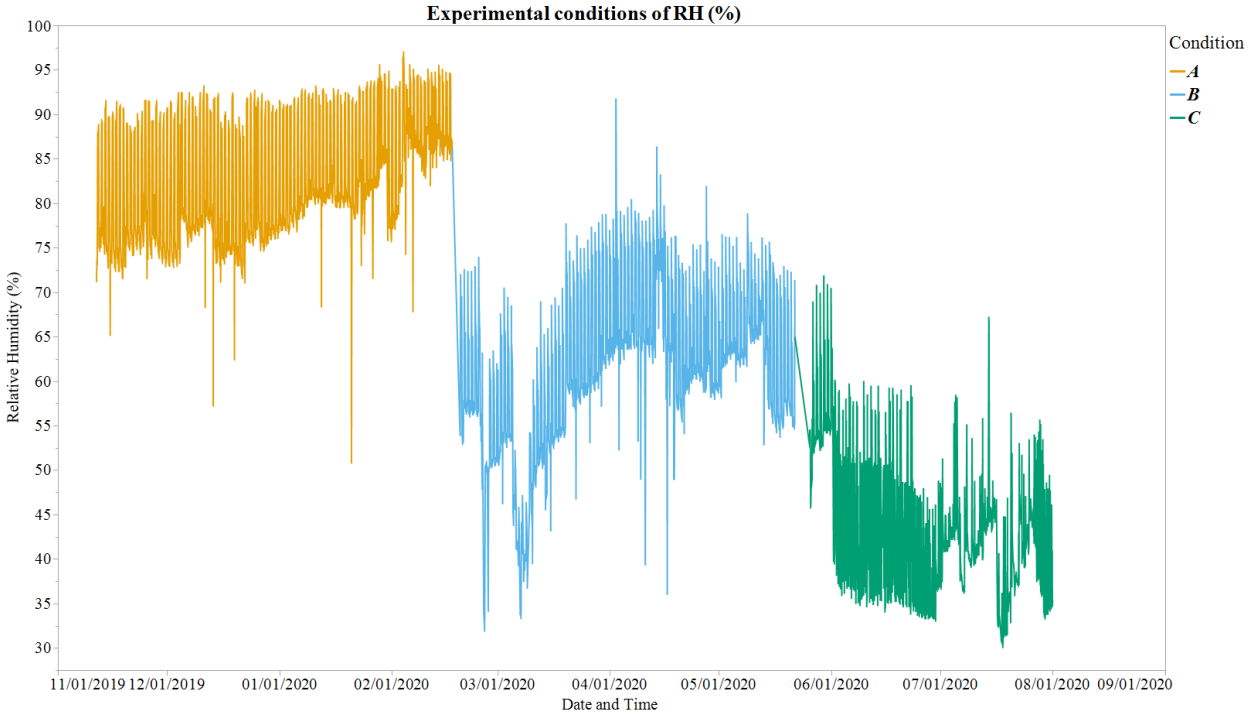


Figure 3.1 Relative humidity conditions in the growth chamber used for seed germination, plant growth and experimental studies.

Two cultivars of romaine lettuce were grown, Green Tower (GT) and True Heart (TH). Seeds were purchased from Urban Farmer (Westfield, IN, US) and Synergene Seed & Technology, Inc. (Salinas, CA, US), respectively. The seeds were sown in trays using a soil mix prepared in a 2:1 ratio by volume of PRO-MIX FLX growing medium (Premier Tech Horticulture, PA, USA) and field soil (sandy loam collected from the top 10 cm of a conventional agricultural field in Indiana). Three weeks after germination, seedlings were transplanted into pots (15.5 cm x 17.8 cm) containing the same soil mix with 0.25 US cup of fertilizer (Osmocote® Smart-Release®). Plants were grown and tested between 6-8 weeks after sowing the seeds, using always the fully developed first and second emerged leaves. Plants were not subjected to water deprivation in any of the conditions, they were watered with the frequency required to avoid the soil mix becoming dry under each RH. Water was added directly on top of the soil avoiding its accumulation in the tray placed under the pots. For every environmental condition, two batches of growth (seeds of both cultivars were sown at two different dates) were used to carry out in all the experimental evaluations.

3.3.2 Counts of culturable bacteria from entire lettuce plants and from testing areas across leaves

Romaine lettuce plants between 7-8 weeks old were used to enumerate the culturable bacterial population of the entire plant. For each lettuce head, all the leaf tissue was collected avoiding the central vein. The total weight of leaf tissue of each plant was thoroughly mixed and divided in half to be processed through sonication or blending. One half was submerged and sonicated for 14 minutes into 300 milliliters of phosphate buffer (PB) 0.1M pH 7.0 using a Branson 5800 ultrasonic cleaner water bath (Branson Ultrasonics Corporation, Danbury, CT), while the other half was blended for 20 seconds with the same buffer (PB) using a 10-fold dilution using a Ninja professional blender. Then, additional 10-fold serial dilutions were performed and spread onto Plate Count Agar (PCA) plates, which were incubated at 30˚C for 72 hours (111). After the time of incubation, colonies were counted, and the total bacterial populations expressed as CFU g^{-1} leaf tissue.

First or second emerged leaves from romaine lettuce plants were used to enumerate the culturable bacterial populations across the leaves. Each leaf was divided into 12 testing areas (Figure 4.2), which were excised using a razor blade and transferred separately to a sterile 50-ml conical tube. Afterwards, 30 ml of PB 0.1M pH 7.0 were added and samples were sonicated as described before. Then, 100 µl were spread onto Plate Count Agar (PCA) plates, which were incubated at 30˚C for 72 hours. Finally, colonies were counted, and the total bacterial populations expressed as log CFU.

3.3.3 Stomatal conductance measurements

The stomatal conductance to water vapor (g_s) of adaxial leaf surfaces was measured at 11:00 am after 5 hours of day light, utilizing a LiCor-6400XT portable infrared gas analyzer (LI-COR Inc., Lincoln, NE, USA). For each environmental relative humidity condition, measurements were performed on 20-30 plants between 6-8 weeks old from one of the first fully expanded leaves. Temperature, light intensity and CO_2 inside the gas exchange chamber were kept constant (24.1˚C [SD= 0.6˚C], 500 µmol m^{-2} s^{-1}, 480 µmol mol^{-1}) through the measurements. The relative humidity parameter was set according to the environmental condition under evaluation, the highest RH (condition A) was achieved by low flow rate (150 µmol s^{-1}), while the medium and lowest RH (conditions B and C) were achieved by high flow rate (300 µmol s^{-1}).

3.3.4 Contact angle hysteresis measurements

One of the first or second emerged leaves (6-8 weeks-old) was harvested per plant to analyze the interaction of water with the leaf adaxial surface. Three GT and three TH lettuce plants were analyzed from each batch of growth, for a total of 12 leaves per environmental condition. To study the differences of surface hydrophobicity across the romaine lettuce leaf, every leaf was divided into 12 testing areas (Figure 3.2). Squared pieces of tissue of less than 10 mm were excised from each testing area using a razor blade. These pieces were then taped (Scotch double-sided tape) to a smooth microscope glass slide to expose the adaxial leaf surface. The water contact angle measurement was conducted by a Ramé-Hart goniometer (model 500) connected to an automated dispensing system (model 100-22, Ramé-Hart, Netcong, NJ), using the needle-in-the-sessile-drop method (112). A drop of 5 µl of deionized water was dispensed at the center of the leaf piece, then 0.25 µl were pumped into the drop every two seconds until the Advancing Contact Angle (ACA) was reached. Afterwards, 0.25 µl were pumped out until the Receding Contact Angle is reached (RCA). The contact angle hysteresis is calculated from the difference of ACA and RCA.

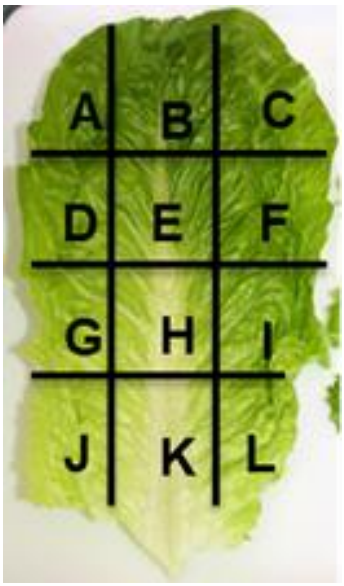


Figure 3.2 Twelve testing areas used as sampling unit for contact angle, stomatal density and stomatal traits measurements.

3.3.5 Quantification of cuticular waxes

The same leaves used for the contact angle measurements were used to perform cuticular wax extraction and quantification. Each leaf was divided in two samples: upper half (testing areas A to F) and lower half (testing areas G to L) and processed separately. Wax extraction was performed by further cutting down each half and soaking into 20ml of Hexane solution for 1 min. The leaf tissues were removed and around 17ml ± 1 mL of clear hexane with wax content was then separated into new tubes, 100 µL of n-tetracosane (1 mg/ml) was added as internal standard (ISTD) and the extracts were subsequently evaporated under a gentle stream of nitrogen. Derivatization of the waxes was performed as described previously (113). Briefly, the wax residues were re-dissolved in a mixture of 50 µL of pyridine and 100 µL of bis-N, N- (trimethylsilyl) trifluoroacetamide containing 10% trimethylchlorosilane (TMCS; Sigma-Aldrich, St. Louis, MO) and heated at 75°C for 75 min. 1 µL of each sample was injected on to Agilent 7890B GC system equipped with Agilent 7693 autosampler coupled to the Agilent 5977B MSD mass spectrometer. The Agilent DB-5MS capillary column (30 m × 0.25 mm; film thickness 0.25 µm) was used for the separation of waxes. The GC oven temperature program started from an initial temperature of 80°C (3 min hold) and ramped at 6°C min-1 up to 300°C with 3 min hold, the total run time was 42.6 min. Helium was used as the carrier gas at the flow of 1ml/min, the GC inlet and MS auxiliary temperatures were set at 220°C and 230°C respectively. The MS data were acquired within the range of 50-500 daltons.

Compound identification was performed using Agilent MassHunter Qualitative Analysis B.07.00 software, and NIST14 database by applying score threshold of >80% similarity against the NIST library spectra. The linear retention index calculations were performed by analyzing C8-C40 alkanes and retention indices were assigned to each compound. Quantification of each wax compound was performed as previously described (114). Briefly, the ratio between each peak area and the peak area of the ISTD was calculated, then each ratio was normalized by the total amount of ISTD (100µg) and by the weight (g) of the leaf pieces of each sample.

3.3.6 Stomatal density and stomatal traits measurements

Three Green Towers and three True Heart 7 week-old lettuce plants were used to characterize the stomatal traits from each environmental condition. One of the first fully expanded

leaves was harvested per plant and the stomatal traits were analyzed on the adaxial surface across the twelve testing areas of the entire leaf (Figure 3.2). To preserve the leaf tissue for the subsequent characterization of bacterial communities, internal 10 x 10 mm leaf pieces were excised to be imaged with confocal laser scanning microscopy (CLSM) using a Zeiss LSM 880 Upright Confocal (Carl Zeiss, Jena, Germany). To alleviate the uneven topography of lettuce leaves, the leaf pieces were enclosed with CoverWell Imaging Chamber Gaskets (20-mm diameter, 0.5-mm deep, ThermoFisher Scientific). Confocal z-stack images were captured using Plan-Apochromat 20x/0.8 objective, chlorophyll autofluorescence with 633nm laser and transmitted light images with differential interference contrast (DIC).

Images were analyzed using ImageJ software (https://imagej.nih.gov/ij/). Stomatal density was evaluated by recording the number of stomata in four randomly chosen fields of view (FOV, 20x objective) from each testing area across the leaf. Stomatal length and width were measured as shown in Figure 3.3; these dimensions were recorded from 50% of the stomata visualized in each FOV. As we cannot guarantee the pore aperture is fully open at the time of imaging, the stomatal length and width were used to calculate pore index (SPI), maximum stomatal pore area (pamax) and maximum theorical stomatal conductance (gsmax) as previously described (115).

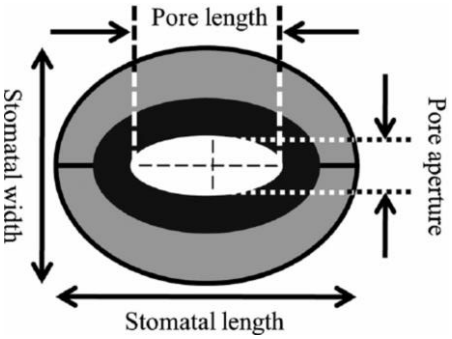


Figure 3.3. Schematic representation of the stomatal dimensions measured by image analysis. The grey area represents the guard cells, the black area the pore walls, and the internal white elliptic area corresponds to stomatal pore area (116).

3.3.7 Characterization of resident bacterial communities across the lettuce leaf surface

To correlate the distribution of leaf structures with the composition and distribution of bacterial communities, the same leaf samples (three GT and three TH) used for stomatal density and traits measurements were used for DNA extraction and amplicon sequencing. Leaf tissue from

each of the twelve testing areas was placed in a sterile 50-ml conical tube and sonicated in 30 milliliters of PB 0.1M pH 7.0 for 14 minutes. The leaf wash was then filtered twice using 0.2 and 0.1 µm polyethersulfone (PES) filter membranes (Membrane solutions, WA, USA) to collect the bacterial cells. Filters were placed in sterile Eppendorf tubes and stored at -80˚C until DNA extraction.

For the amplicon sequencing analysis, bacteria coming from the upper half (pooling testing areas A to F) and from the lower half (pooling testing areas G to L) of each leaf were sequenced separately. A total of 36 samples (12 from each environmental condition) were processed. Additionally, three samples prepared by sonication of whole leaves using entire plants (5 GT and 5 TH pooled together) grown under each condition A, B and C were included to represent the total bacterial microbiome. Also, controls for the identification of contaminant reads from the DNA extraction process with buffer extraction controls (BEC) and from the amplification method with non-template controls (NTC) were included. Finally, to perform quality control of our entire methodology from sample processing, DNA extraction, 16S rRNA amplification, MiSeq sequencing until bioinformatics processing, one mock microbial community standard (D6300) and one mock microbial community DNA standard (D6305) were included (Zymo Research, Irvine, CA, USA).

3.3.8 DNA extraction and library preparation

Frozen filter membranes were minced under liquid nitrogen in the same Eppendorf tubes using disposable pellet pestles (Fisherbrand, CA, USA). The minced filters were placed directly into the PowerBead Pro tubes of the DNeasy® PowerSoil® Pro DNA extraction Kit (Qiagen, MD, USA). DNA extraction was performed following the manufacturer's instructions, a 1600 MiniG® Automated Cell Lyser (SPEX Sample Prep, NJ, USA) was used to homogenize samples thoroughly by performing three repetitions of cycles of three minutes at 1500rpm and one-minute rest in between. Library preparation was carried out using the Illumina 16S Metagenomic Sequencing Library protocol with some modifications. A dual indexing protocol with a two-step PCR was performed to amplify the V5 through V6 region of the bacterial 16S rRNA gene and to add dual TruSeq indices. Primers for 16S sequences (underlined portion) with TruSeq adaptor tails are shown: primer 799F (5' ACACTCTTTCCCTACACGACGCTCTTCCGATCT AACMGGATTAGATACCCKG 3') was chosen to exclude chloroplast amplification (91, 92), and

primer 1114R (5’ GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT GGGTTGCGCTCGTTGC 3’) includes three mismatches with plant mitochondrial DNA (117).

The primary amplification of the 16S region was performed in 25-µl PCR reactions with 3.0 µl of genomic DNA (~ 10ng), 1X of 5X KAPA HiFi Buffer, 0.3 mM of dNTPS, 0.5 U of KAPA HiFi polymerase (KAPA Biosystems, Woburn, MA) and 0.2 µM of each primer 799F/1114R. Cycling conditions were as follows: 95°C for 5 min, followed by 30 cycles of 98°C for 20 s, 55°C for 30 s, and 72°C for 45 s, with a final extension of 72°C for 5 min. The approximately 415-bp PCR products were purified using Agencourt AMPure XP (Beckman Coulter, Inc., CA, USA) at a ratio of 0.7x to remove any products shorter than 200-bp.

The secondary amplification to add TruSeq dual indices and sequencing adaptors was performed in 25-µl PCR reactions with 3.0 µl of purified first PCR product, 1X of 5X KAPA HiFi Buffer, 0.3 mM of dNTPS, 0.5 U of KAPA HiFi polymerase (KAPA Biosystems, Woburn, MA) and 0.3 µM of each primer i5 and i7. Cycling conditions were as follows: 95°C for 5 min, followed by 10 cycles of 98°C for 20 s, 55°C for 30 s, and 72°C for 45 s, with a final extension of 72°C for 5 min. These second PCR products were also purified as described above.

Each sample was amplified three times and every replicate was prepared as a separate library to be sequenced. All 108 libraries were quantified through qPCR assays using primer 1.1 (5’ AATGATACGGCGACCACCGAGAT 3’) and primer 2.1 (5’ CAAGCAGAAGACGGCATACGA 3’), which is the method reported with more accurate prediction of sequencing coverage (118). A 10,000-fold dilution of the purified second PCR products was prepared with 10 mM Tris-HCl pH 8.0 to achieve concentrations inside the limits of the standard curve built with the KAPA Library Quantification DNA Standards 1 – 6 (KAPA Biosystems, Woburn, MA). Quantification was performed in 10-µl qPCR reactions with 5 µl of PowerUp™ SYBR™ Green Master Mix 2X, 2 µl of diluted second PCR products, and 0.2 µM of each primer 1.1/2.1. Cycling conditions were as follows: 50°C for 2 min, 95°C for 2 min, followed by 40 cycles of 95°C for 15 s, 55°C for 15 s, and 72°C for 45 s. All libraries were multiplexed into a single pool in molar equivalent concentrations. The pool was sequenced and demultiplexed at Genewiz (South Plainfield, NJ, USA) using Illumina MiSeq paired-end 250-bp sequencing.

3.3.9 Sequence data processing and analysis

Data processing and analysis was performed using packages in R (v 3.6.1) and Bioconductor (v 3.10). Reads were processed through a customized DADA2 (v 1.14.1) pipeline (95) for trimming the primer sequences, filtering based on read quality, calculating error rates, inferring amplicon sequence variants (ASVs) instead of using OTUs (119), merging paired-end reads, removing chimeras and assigning taxonomy with the SILVA reference database (v 138) (120). Contaminant sequences which were present in the buffer extraction controls (BEC) and non-template controls (NTC) were removed with the decontam package (120). Also, reads with unclassified Phylum or identified as Chloroplast or Mitochondria were filtered out. The α-diversity indices (richness and Shannon) were performed with the Phyloseq (v 1.32.0) package after 100 subsampling iterations to the smallest library size and averaging the results (96). The β-diversity unconstrained analysis was performed on weighted Unifrac distances applied to reads proportionally scaled to 1,000-bp (code adapted from tutorial: http://deneflab.github.io/MicrobeMiseq) and were then visualized using Principal Coordinates Analysis (PCoA). Experimental conditions were compared by analysis of variance using a permutation test with pseudoF ratios with the adonis() function (R vegan package). Normalization and differential abundance analysis were performed on unrarefied ASV table using DESeq2 (v 1.28.1) (121); only significant taxa ($P<0.05$ after multiple-hypothesis testing) were reported. Plots were generated with the ggplot2 (v 3.3.2) package and minor aesthetic details were arranged in Inkscape (v 0.92.4).

3.4 Results and Discussion

3.4.1 Relative humidity determines leaf properties of romaine lettuce

The pressure of the environmental conditions used to grow our plants is evidenced by the significant differences achieved on vapor pressure deficit (VPD), which significantly increased the transpiration under dryer conditions (Figure 3.4). Mechanisms of adaptations to low relative humidity (RH) conditions were observed, such as stomatal resistance through decrease of stomatal size (both length and width), stomatal density (SD), stomatal pore area (PA_{max}) and stomatal aperture (measured as stomatal conductance, g_s), as well as cuticle resistance through modification of composition of cuticular waxes (122). Additionally, modifications of leaf wettability were

observed through contact angle hysteresis (CA) measurements because the larger the CA value, the more the leaf surface tends to repel water (i.e. the more hydrophobic it is) (123, 124). Noticeably, most of the changes associated with stomatal resistance to water loss were more pronounced between plants grown under humid condition A and intermediate condition B, while plants grown under dry condition C exhibited stomatal traits between humid condition A and intermediate condition B (Table 3.2). This might reflect how the plants need to balance their responses between the protection from losing too much water and the uptake of enough carbon dioxide to drive photosynthesis (122). This hypothesis is supported by our results of photosynthesis rates which showed a decrease from humid condition A to intermediate condition B, and then showed a recovery under the dry condition C. It is important to note that our conditions were focused only on the environmental RH, and water deprivation was not intended as plants were watered at a frequency required to keep healthy lettuce plants.

Figure 3.4. Comparison of (A) vapor pressure deficit (VPD) and (B) transpiration among the three relative humidity (RH) conditions where the romaine lettuce plants were grown: humid condition A, intermediate condition B and dry condition C. Different letters indicates significant difference ($P<0.05$).

A.

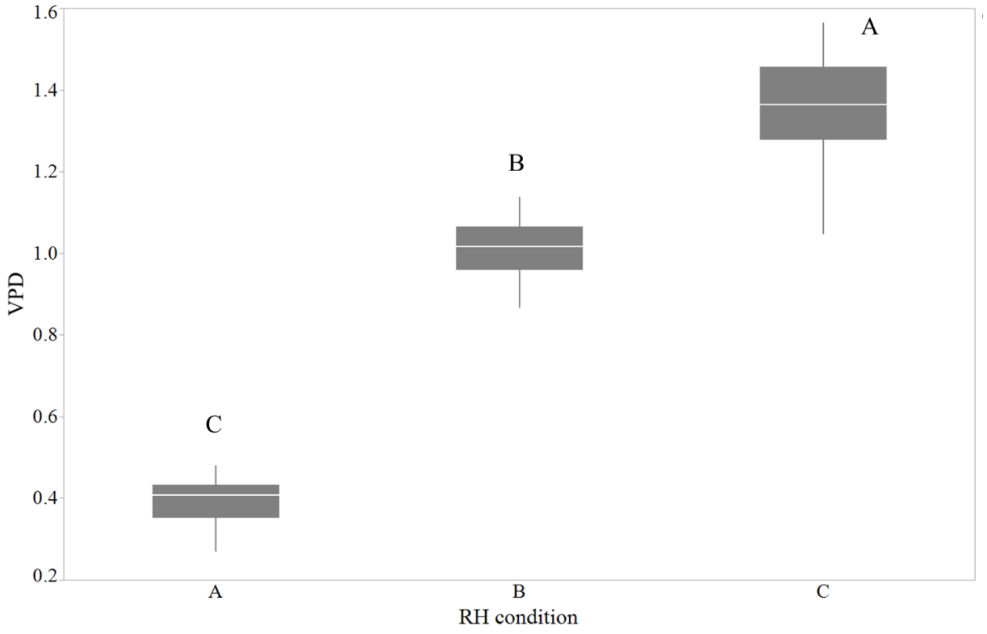
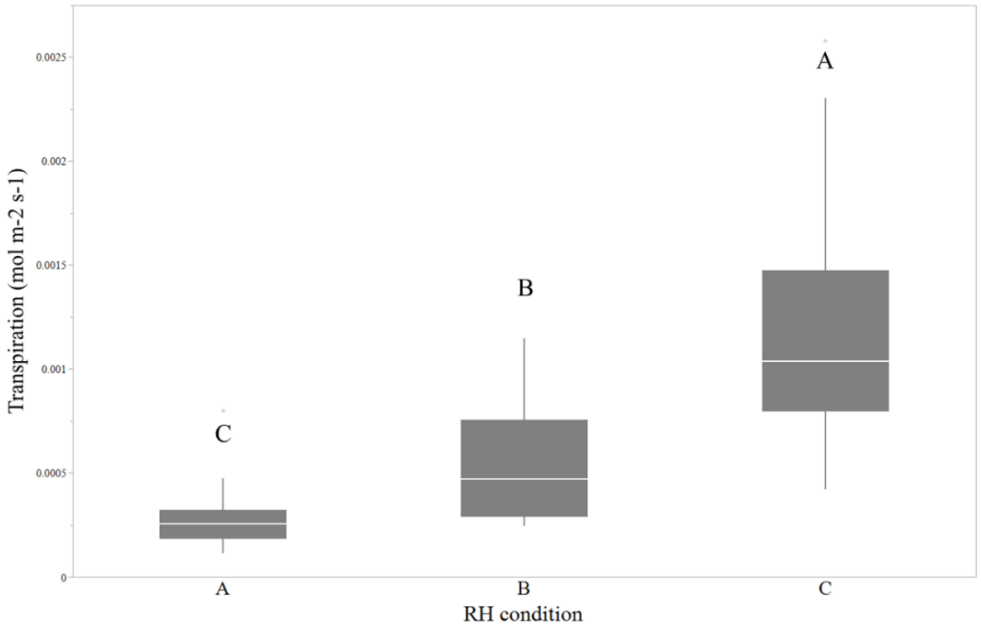


Figure 3.4 continued

B.



The increase of stomatal density observed under RH B was based on the development of smaller stomata as supported by the significant negative correlations found with stomatal length and width (Table 3.2). Higher stomatal densities were associated with a decreased leaf wettability as evidenced by the higher contact angle hysteresis (CA) obtained on leaves of plants grown under this RH (significant positive correlation, R= 0.52). Previous studies have also reported that leaf surfaces with greater concentration of stomata were the least wettable (125, 126). The higher density of stomata was also associated with a reduction in the proportion of fatty acids in leaf cuticles under RH B (Figure 3.5). This relationship might be explained by the reduction of the area that epidermal cells can occupy. This is supported by previous evidence suggesting that the epicuticular layer of guard cells differs from that of other epidermal cells in both thickness and composition with increased concentration of fluorescing phenolic compounds (127, 128).

Stomatal size had a big impact on stomatal conductance (g_s) and photosynthesis (P). Although the reduction of size of stomata under RH B produced a higher stomatal density, this compensation was insufficient, and the stomatal pore area under this RH condition was

significantly lower (Table 3.2). Also, the lowest g_s was observed under RH B which suggests the closure of stomata as a strategy to avoid water loss; this observation is supported by the theorical maximum g_s which showed that leaves from RH B have the anatomical potential to produce the highest conductance if their stomata were fully open. On the other hand, photosynthesis was positively correlated with stomatal size, stomatal pore area and stomatal aperture, therefore lettuce plants grown under humid condition A exhibited a higher photosynthetic activity. Kroupitski *et al.* (129) suggested that photosynthesis and not merely stomatal opening is the major motive force for *Salmonella* internalization in harvested iceberg lettuce. This could also imply a greater risk for internalization of food-borne human pathogens after a contamination event on romaine lettuce plants grown under humid conditions.

Additionally, a significant negative correlation was found between CA and photosynthesis (Table 3.2), which supports the idea that leaf wettability (low CA) increases photosynthesis. This photosynthetic response to wetness has been postulated as a consequence of the stomatal regulation of pore aperture (130). Additionally, leaf wettability promotes bacterial growth through the provision of free water and protection from desiccation stress (131); for that reason our culturable bacterial counts were significantly higher in lettuce plants exhibiting lower CA values. Finally, increased leaf wettability has a positive effect in reducing transpiration, which was evidenced in our lettuce plants grown under condition A (Figure 3.4 B).

Table 3.2. Properties and traits of leaves of romaine lettuce plants grown under different relative humidity (RH) conditions (A, B and C). Row legends: SD (stomatal density); SL (stomatal length); SW (stomatal width) SPI (stomatal pore index), PA_{max} (maximum stomatal pore area), CA (contact angle hysteresis); CW (cuticular wax); Alk (alkanes); FA (fatty acids); RHO (alcohols); g_s (stomatal conductance); g_{smax} (theorical maximum stomatal conductance); T (transpiration) and P (Photosynthesis).

	RH A	RH B	RH C	Correlations
Total Culturable Bacteria (Log CFU/g lettuce)	5.43 (0.45) A Cultivar: P<0.0001	3.46 (0.46) B Cultivar: P=0.0758	3.76 (0.60) B Cultivar: P<0.0001	SD: R= -0.30 P= 0.0098 SL: R= 0.38 P= 0.0010 SW: R= 0.49 P< 0.0001 PA: R= 0.50 P< 0.0001 CA: R= -0.50 P< 0.0001 P: R= 0.42 P=0.0002 T: R= -0.41 P=0.0003
SD (#/mm^2)	43.7 (6.9) B Cultivar: P=0.5166	53.3 (8.9) A Cultivar: P=0.1599	46.9 (9.2) B Cultivar: P=0.0149	SL: R= -0.56 P< 0.0001 SW: R= -0.37 P= 0.0014 PA: R= -0.53 P< 0.0001 CA: R= 0.52 P< 0.0001

Table 3.2 continued

SL (µm)	28.8 (1.1) A Cultivar: P=0.0011	26.9 (1.4) C Cultivar: P=0.0758	27.8 (1.2) B Cultivar: P=0.6883	SW: R= 0.52 P<0.0001 CA: R= -0.47 P< 0.0001 P: R= 0.31 P= 0.0073
SW (µm)	11.5 (0.6) A Cultivar: P<0.0001	10.8 (0.4) C Cultivar: P<0.0001	11.1 (0.3) B Cultivar: P<0.0001	CA: R=-0.38 P= 0.0009 P: R= 0.28 P=0.0190
SPI (x100)	1.45 (0.25) Cultivar: P=0.0090	1.52 (0.18) Cultivar: P=0.0012	1.44 (0.28) Cultivar: P=0.1049	CA: R= 0.33 P=0.0041
PA_{max} (μm^2)	261.0 (17.7) A Cultivar: P<0.0001	227.9 (16.9) C Cultivar: P=0.3779	242.4 (13.4) B Cultivar: P=0.2244	CA: R= -0.49 P= 0.0041 P: R= 0.33 P= 0.0041
CA (˚)	42.8 (8.3) C Cultivar: P=0.1035	56.7 (7.8) A Cultivar: P=0.3031	50.4 (8.3) B Cultivar: P=0.6883	P: R= -0.44 P= 0.0001
CW Alk (µg/g leaf)	4.65 (4.06) B Cultivar: P=0.0092	2.06 (2.91) B Cultivar: P=0.0004	9.33 (8.53) A Cultivar: P=0.6883	P: R= 0.27 P= 0.0242 T: R= 0.37 P= 0.0013
CW FA (µg/g leaf)	94.1 (28.4) A Cultivar: P=0.0227	97.2 (19.9) A Cultivar: P=0.1049	70.3 (16.0) B Cultivar: P=0.0425	SD: R= -0.28 P=0.0153 SPI: R= -0.37 P= 0.0015 RHO: R= 0.79 P< 0.0001
CW RHO (µg/g leaf)	95.1 (25.2) B Cultivar: P=0.0092	114.0 (32.2) A Cultivar: P=0.2389	55.2 (22.1) C Cultivar: P=0.0003	
g_s (mol m^{-2} s^{-1})	0.075 (0.039) AB Cultivar: P=0.0509	0.053 (0.030) B Cultivar: P=0.0389	0.090 (0.050) A Cultivar: P=0.0003	P: R= 0.64 P< 0.0001
g_{smax} (mol m^{-2} s^{-1})	0.629 (0.164) B Cultivar: P=0.7474	0.687 (0.169) A Cultivar: P=0.0121	0.641 (0.180) B Cultivar: P=0.0149	
T (mol m^{-2} s^{-1}) x1000	0.283 (0.132) C Cultivar: P=0.1935	0.521 (0.245) B Cultivar: P=0.0121	1.147 (0.504) A Cultivar: P=0.0012	g_s: R= 0.74 P< 0.0001 P: R= 0.31 P= 0.0087
P (µmol CO2 m^{-2} s^{-1})	8.57 (2.69) A Cultivar: P=0.0034	4.46 (2.13) B Cultivar: P=0.0007	6.92 (2.71) A Cultivar: P<0.0001	

Different letter next to values indicates significant difference (ANOVA P < 0.05) caused by RH condition. Correlations among leaf properties estimated from all RH conditions together using Row-wise method.

In the case of the response of cuticle resistance, we observed a significant increase of alkanes along with a decrease of fatty acids and alcohols in the composition of the cuticular waxes of plants grown under the driest condition C. Previous studies have reported higher amounts of total waxes extracted from leaves of *Brassica oleracea*, *Eucalyptus gunnii* and *Tropaeolum majus* plants grown under very low (20-30%) relative humidity conditions (132). Our data only include

wax compounds classified within the most commonly reported classes of cuticular waxes in lettuce (*Lactuca sativa*) plants grown under greenhouse conditions where RH conditions were not controlled (113, 133). This suggests that additional classes of compounds might be increasing in cuticles of plants grown under our low RH condition C and further analysis should be performed to assess this. Significant changes in cuticle composition were observed, with a higher relative abundance of alcohols found under our intermediate RH condition B. This agrees with the studies performed on greenhouse-grown lettuce where the primary form of wax reported was Hexacosanol (113, 133), probably because their plants were grown under similar intermediate RH conditions since their objective was not related to water stress. On the other hand, our results exhibited shifts in the proportions of wax compounds when plants were grown under extreme conditions of high and low RH. For instance, increase of alkanes under the lowest RH condition coincides with a recent study of litter flammability where they found that higher cuticular alkane concentrations were reached under aggravated drought and were associated with increased flammable potential (134).

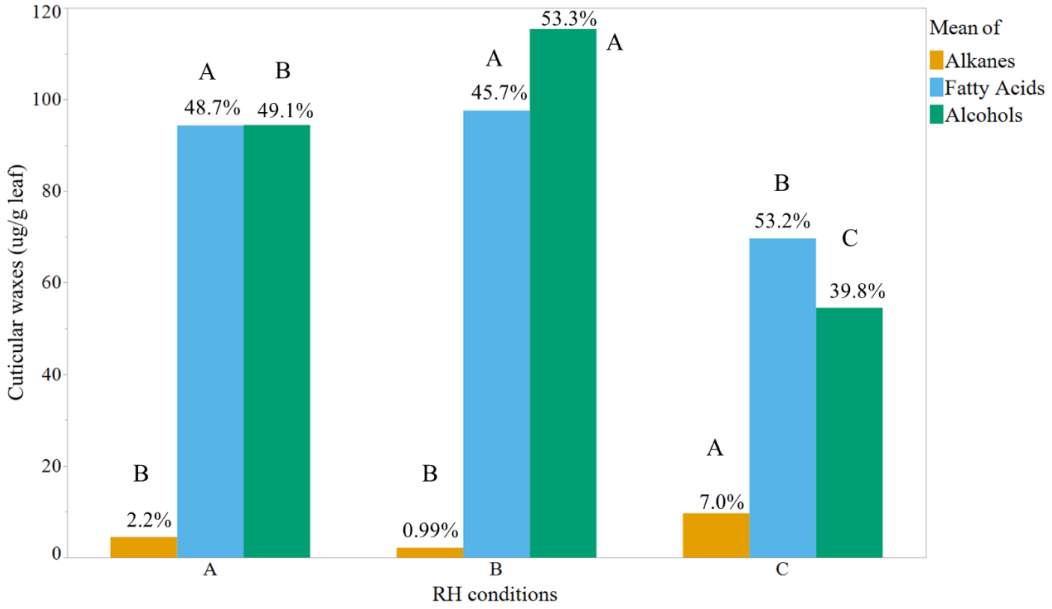


Figure 3.5. Mean concentration (µg/g leaf) of the three main classes of cuticular waxes in romaine lettuce: alkanes, fatty acids and alcohols. Bars are labeled with the proportion of each class of waxes. Different letters indicate significant difference ($P<0.05$) caused by RH conditions within each class of waxes.

3.4.2 Leaf properties are heterogeneously distributed across the leaf area

Colonization of the phyllosphere is not spatially distributed homogeneously across the entire leaf surface. Bacterial clustering around areas where there is more free water, shield from UV rays, and leaking of nutrients such as at the base of trichomes, around stomata and along veins have been proposed (135, 136). For that reason, we evaluated the distribution of leaf properties and traits across the adaxial leaf area by quantification in each of the twelve testing areas described before (Figure 3.2), to test their association with the distribution of bacterial communities.

Differences in leaf properties and traits were tested through different distributions across the adaxial leaf surface using the twelve testing areas as follows: vertical zones (Left: A, D, G, J; Middle: B, E, H, K; Right: C, F, I, L), horizontal zones (1: A-C; 2: D-F; 3: G-I, 4: J-L) and horizontal half (Upper: A-F, Lower: G-L). Major differences were observed as a gradient among horizontal zones, elucidating the upper half and lower half of the same leaf as contrasting habitats mainly under the extreme RH conditions A and C (Table 3.3). For instance, stomatal density (SD) was consistently higher in the upper half of leaves regardless of the RH condition of plant growth. This higher density was also reported from iceberg lettuce in post-harvest conditions (137). Although this difference of SD was not significantly different in plants grown under intermediate RH condition B, the stomatal pore index (SPI), which is dependent on stomatal density and size, was. Stomatal length showed a significant difference between the upper and lower half under all RH conditions, while stomatal width was not variable. Therefore, the upper halves of leaves comprised higher stomatal density with longer stomata when compared with the lower halves of leaves, as a result SPI and maximum pore area were also significantly higher in the upper halves. However, these differences across leaf surface might become less significant when plants are undergoing a process of adaptation to avoid water loss by reducing stomatal size that consequently caused increase of stomatal density.

Leaf wettability was also significantly different between the upper and lower halves of lettuce leaves regardless of the RH condition of plant growth. Upper halves consistently exhibited higher CA which evidenced a reduced wettability (Table 3.3). Contrary to some previous reports where a strong positive correlation between CA and quantity of cuticular waxes was found (133, 138), we observed significantly lower concentrations of fatty acids and alcohols in the upper halves of leaves. This finding again supports our hypothesis that increased stomatal density might be the leading factor for the reduction of leaf wettability, as previously reported (125, 126), due to

differences of epicuticular waxes in guard cells (128). On the other hand, it is noticeable how CA measurements exhibited additional significant differences across more areas of the leaf surface when the RH condition was decreasing. Leaf areas with positions that are closer to the central vein and less exposed to the surrounding environmental conditions such as horizontal zone 4 and the middle vertical zone exhibited significantly increased wettability.

To the best of our knowledge this is the first study to determine differences of leaf properties and traits across the entire adaxial leaf surface of romaine lettuce plants during pre-harvest. Previous studies have focused on differences between abaxial and adaxial leaf surface (103, 139, 140), while others have performed comparisons among only three areas (top, middle and bottom) of the leaf surface but only in post-harvest conditions using lettuce plants bought from grocery stores (137, 141). Overall, we demonstrated that leaves from romaine lettuce living plants before harvesting exhibited significant differences in the distribution and quantity of leaf properties and traits across the entire leaf surface and that the RH conditions during plant sowing and growth cycles produced significant changes in the leaf properties. Next, we evaluated how these differences caused by environmental conditions and localization on leaf surfaces might determine the structure and distribution of resident bacterial communities.

3.4.3 Sequencing of resident bacterial communities of romaine lettuce leaves

As described before, lettuce plants (cultivars Green Towers-GT and True Heart-TH) grown under three environmental conditions of high RH (condition A), medium RH (condition B), and low RH (condition C) were used for sequencing of the resident bacterial communities. The same leaves subjected to the characterization of stomatal traits using confocal microscopy were used for bacterial 16S sequencing of the upper and lower halves of each leaf separately. Therefore, six plants (3 GT and 3 TH) were surveyed from each RH treatment. A total of thirty-six samples with three technical replicates each were sequenced; additionally, for comparison with the total bacterial microbiome of the lettuce plants, three samples prepared by pooling the sonication buffer of all whole leaves from entire plants (10 plants: 5 GT and 5 TH) were included.

The V5-V7 region of the 16S rRNA was sequenced by paired-end 250bp MiSeq Illumina sequencing, yielding contigs with length sizes of 298 ± 7.1 bp. After quality filtering and removal of chimera, non-target (archaea, chloroplast and mitochondria), and contaminant sequences (from BEC and NTC), 2.1 million reads corresponding to 12,375 ASVs were obtained. Low abundance

ASVs (fewer than 2 reads in less than 10% of the samples) which are considered spurious products, were removed and 1516 significant ASVs were left. Finally, two samples were removed after we applied a minimum threshold of 1000 reads for library size (Figure 3.6), and the final data set for analysis consisted of 1.8 million reads, 1516 ASVs, an average of 14,987 and a median of 9,963 reads per sample. The quality control performed with mock microbial communities showed a good performance of our methodology. The eight standard strains were identified in the mock controls and their relative abundances were close to the expected values, with a deviation less than 15% as recommended by the manufacturer (Figure 3.7).

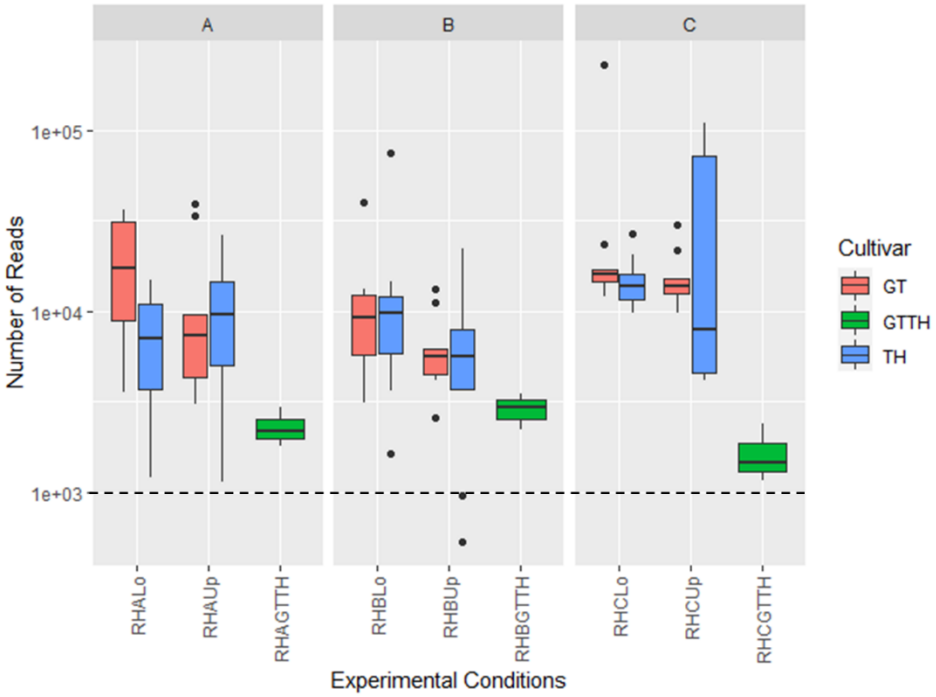


Figure 3.6. Boxplots of library sizes of sequenced samples from lower (Lo) and upper (Up) halves of leaves. Panels correspond to relative humidity (RH) conditions (A, B, and C) used for lettuce plant growth. Samples are colored by cultivar: GT (Green Towers), TH (True Heart) and GTTH refers to the pool of GT and TH plants sequenced as the total bacterial microbiome. Dashed line indicates the threshold of minimum library size applied.

Table 3.3. Properties and traits across the leaf adaxial area of romaine lettuce plants grown under different relative humidity (RH) conditions (A, B and C). Row legends: SD (stomatal density); SL (stomatal length); SW (stomatal width) SPI (stomatal pore index); PA_{max} (maximum stomatal pore area); CA (contact angle hysteresis) and CW (cuticular wax). Column legends and descending order of data: VZ (vertical zones: Left, Middle, Right); HZ (horizontal zones: 1, 2, 3, 4*); HH (horizontal half: Lower, Upper). Data are presented as mean (SD).

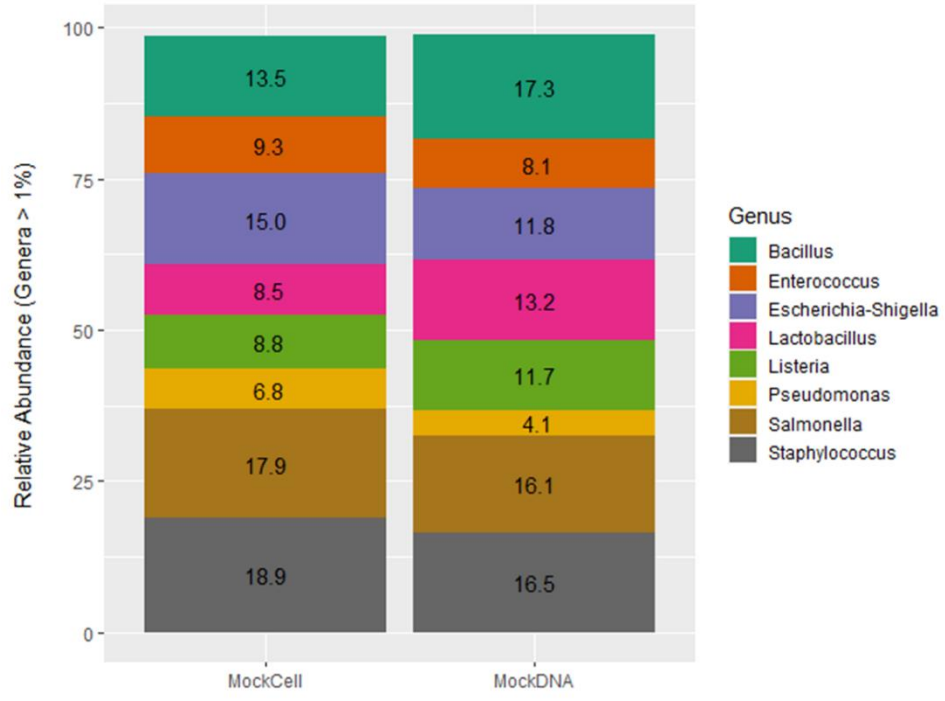
	RH A			RH B			RH C		
	VZ	HZ	HH	VZ	HZ	HH	VZ	HZ	HH
Bacteria (log CFU)	2.53 (0.38) 2.89 (0.69) 2.67 (0.42)	2.80 (0.53) 2.86 (0.63) 2.55 (0.44) 2.57 (0.47)	2.56 (0.45)^{B} 2.83 (0.58)^{A}	2.83 (0.56) 2.85 (0.60) 2.74 (0.46)	2.85 (0.43) 2.84 (0.44) 2.54 (0.47) 2.99 (0.73)	2.77 (0.64) 2.85 (0.43)	3.19 (0.71) 3.37 (0.79) 2.99 (0.51)	3..02(0.50)^{B} 3.08 (0.56)^{B} 2.83 (0.67)^{B} 3.74 (0.67)^{A}	3.30 (0.80) 3.05 (0.52)
SD (#/mm^2)	45.7 (9.5) 42.7 (13.7) 42.6 (10.7)	48.6 (10.5)^{A} 48.6 (10.1)^{A} 39.5 (9.0)^{B} 37.9 (11.9)^{B}	38.7 (10.6)^{B} 48.6 (10.2)^{A}	51.7 (14.1)^{AB} 47.4 (12.4)^{B} 54.2 (12.6)^{A}	47.7 (11.0)^{B} 55.6 (12.6)^{A} 52.5 (13.2)^{AB} 48.4 (14.9)^{B}	50.4 (14.2) 51.6 (12.4)	46.1 (12.2) 44.3 (14.4) 48.3 (13.0)	50.4 (10.8)^{A} 49.9 (11.1)^{A} 49.6 (11.7)^{A} 35.4 (13.0)^{B}	42.5 (14.2)^{B} 50.2 (10.9)^{A}
SL (µm)	28.6 (2.5) 29.1 (2.0) 28.7 (2.3)	29.3 (2.3)^{A} 29.3 (1.7)^{A} 29.0 (2.0)^{A} 27.7 (2.5)^{B}	28.3 (2.3)^{B} 29.3 (2.1)^{A}	26.8 (2.0)^{B} 28.0 (2.0)^{A} 26.4 (1.7)^{B}	28.0 (1.5)^{A} 27.2 (1.8)^{AB} 26.9 (2.0)^{B} 26.3 (2.3)^{B}	26.6 (2.2)^{B} 27.6 (1.7)^{A}	27.3 (1.8)^{B} 28.7 (2.0)^{A} 27.4 (2.0)^{B}	27.3 (1.5)^{BC} 28.8 (1.8)^{A} 27.1 (2.0)^{C} 28.0 (2.3)^{B}	27.5 (2.2)^{B} 28.1 (1.8)^{A}
SW (µm)	11.4 (0.9)^{B} 11.9 (0.9)^{A} 11.3 (0.9)^{B}	11.6 (0.8) 11.6 (1.0) 11.6 (0.8) 11.4 (1.0)	11.5 (0.9) 11.6 (0.9)	10.6 (0.8)^{B} 11.0 (0.9)^{A} 10.8 (0.8)^{AB}	10.9 (0.9) 10.7 (0.8) 10.8 (0.8) 10.6 (0.9)	10.7 (0.9) 10.8 (0.8)	11.2 (0.8) 11.0 (0.9) 11.1 (0.8)	11.0 (0.6) 11.3 (0.8) 11.0 (1.0) 11.1 (0.9)	11.1 (0.9) 11.2 (0.7)
SPI (x100)	1.49 (0.3) 1.48 (0.5) 1.38 (0.4)	1.64 (0.4)^{A} 1.65 (0.4)^{A} 1.33 (0.3)^{B} 1.18 (0.4)^{B}	1.26 (0.4)^{B} 1.65 (0.4)^{A}	1.45 (0.4) 1.46 (0.4) 1.53 (0.3)	1.46 (0.4)^{BC} 1.62 (0.4)^{A} 1.51 (0.3)^{AB} 1.33 (0.4) ^{C}	1.42 (0.4)^{B} 1.54 (0.4)^{A}	1.41 (0.4) 1.39 (0.4) 1.47 (0.4)	1.52 (0.4)^{A} 1.62 (0.4)^{A} 1.47 (0.3)^{A} 1.07 (0.3)^{B}	1.27 (0.4)^{B} 1.57 (0.4)^{A}
PA_{max} (µm^2)	256.9 (32.7)^{B} 272.3 (29.5)^{A} 255.8 (27.9)^{B}	266.4 (29.0)^{A} 268.0 (26.7)^{A} 263.7 (25.8)^{A} 248.5 (37.5)^{B}	256.1 (33.0)^{B} 267.2 (27.8)^{A}	223.1 (28.5)^{B} 241.7 (26.1)^{A} 223.9 (26.4)^{B}	240.8 (26.0)^{A} 229.3 (23.9)^{AB} 228.2 (26.6)^{B} 220.5 (32.5)^{B}	224.3 (29.9)^{B} 235.2 (25.5)^{A}	241.3 (24.7) 248.4 (29.6) 239.1 (28.6)	237.2 (20.8)^{B} 255.7 (24.9)^{A} 235.0 (28.7)^{B} 243.4 (31.5)^{B}	239.3 (30.3)^{B} 246.5 (24.7)^{A}
CA (˚)	44.4 (16.0) 40.4 (14.9) 42.1 (14.6)	44.5 (15.5) 47.2 (15.5) 42.3 (15.3) 36.0 (12.6)	39.0 (14.2)^{B} 46.0 (15.4)^{A}	59.2 (9.6) 52.9 (6.7) 58.4 (8.6)	61.0 (7.5)^{A} 59.5 (10.0)^{A} 57.9 (7.2)^{AB} 49.0 (4.7)^{B}	53.4 (7.5)^{B} 60.2 (8.6)^{A}	53.4 (8.8)^{A} 44.6 (9.8)^{B} 53.0 (8.1)^{A}	55.7 (8.3)^{A} 52.2 (11.8)^{AB} 49.3 (5.8)^{AB} 43.9 (8.5)^{B}	46.5 (7.7)^{B} 53.9 (10.2)^{A}
CW: Alkanes (µg/g leaf)	NA	NA	6.21 (4.2) 2.95 (4.54)	NA	NA	2.51 (4.0) 1.8 (2.3)	NA	NA	12.6 (14.4) 10.3 (10.1)

Table 3.3 continued

CW: Fatty acids (µg/g leaf)	NA	NA	116.6 (16.2)^{A} 74.6 (21.9)^{B}	NA	NA	101.8 (26.9) 92.9 (16.3)	NA	NA	86.8 (28.5)^{A} 61.1 (14.1)^{B}
CW: Alcohols (µg/g leaf)	NA	NA	112.4 (21.7)^{A} 78.5 (22.4)^{B}	NA	NA	126.5 (41.3) 103.2 (24.8)	NA	NA	70.7 (20.2)^{A} 42.0 (18.8)^{B}

*Horizontal zones are arranged with testing areas (Figure 3.2) as follows: 1 (A-C), 2 (D-F), 3 (G-I), 4 (J-L).
Different letter next to values indicates significant difference (ANOVA $P < 0.05$) caused by VZ, HZ or HH within each leaf parameter. NA: Not Applicable.

A.



B.

ZymoBIOMICS™ Microbial Community Standard	
Species	**Theorical composition (%) 16S Only**
Pseudomonas aeruginosa	4.2
Escherichia coli	10.1
Salmonella enterica	10.4
Lactobacillus fermentum	18.4
Enterococcus faecalis	9.9
Staphylococcus aureus	15.5
Listeria monocytogenes	14.1
Bacillus subtilis	17.4

Figure 3.7. Microbial composition (relative abundance, %) of mock community standards. A. Our results for the microbial community culture standard (D6300: MockCell) and the microbial community DNA standard (D6305: MockDNA). B. Theorical composition reported by the manufacturer (Relative-abundance deviation in average: <15%).

3.4.4 Structure of resident bacterial communities from lettuce plants grown under different relative humidity conditions

The taxonomic structure of native bacterial communities from leaves of romaine lettuce plants grown in growth chamber, combining all the RH conditions, consisted in 20 phyla and 45 classes. Most of the resident bacteria belong to 9 classes (Figure 3.8): Actinobacteria, Acidimicrobia and Thermoleophilia from phylum Actinobacteria, Gamma and Alphaproteobacteria from phylum Proteobacteria, Bacilli and Clostridia from phylum Firmicutes, Bacteroidia from phylum Bacteroidetes and Deinococci from phylum Deinococcota. As a whole, Actinobacteria was the most abundant class (42.3 ± 8.6 %) in all lettuce leaves, followed by Bacilli (17.7 ± 7.5%), Alphaproteobacteria (14.8 ± 6.9 %) and Gammaproteobacteria (8.84 ± 5.15 %). These classes and their corresponding phyla have been described as the major inhabitants of the phyllosphere across most agricultural and native plant species (2, 67, 142, 143). However, one of the main differences between field-grown and indoor-grown (growth chamber in our case) lettuce plants is the predominance of Proteobacteria in the bacterial communities from field conditions (66, 72).

Several reasons might explain why our plants grown under growth chamber settings exhibited more abundance of Actinobacteria than Proteobacteria. First, the reduced exposure of indoor-grown plants to environmental conditions of rain, aerosols, winds, contact with insects and animals. The sources of microbes under growth chamber conditions are limited to seed, soil, tap water, minimal air circulation (since the growth chamber was closed most of the time) and minimal human contact (because only 1-2 people worked inside and always with gloves on). For that reason, soil from an agricultural field in Indiana was mixed with the growing medium to increase the bacterial diversity provided to our plants. Therefore, the main source of our plants for the uptake microbes was the potting soil mix, and Actinobacteria have been reported as representative inhabitants of leaves and soil too. However, sequencing of free soil is ruled out since the underrepresentation of Acidobacteria (not abundant in our leaf and plant samples) and the increased proportion of Actinobacteria and Proteobacteria members have been reported as the main characteristics of rhizosphere and root endosphere when compared with free soil (2). Additionally, aerosols have been found to be an important source of Alpha and Gammaproteobacteria with abundant sequences of the genera *Sphingomonas* and *Pseudomonas* (144); the lack of aerosols as inoculum sources was noticed by the underrepresentation of

Pseudomonas in our plants. A final reason is that our methodology to survey the bacterial communities implied the excision of testing leaf areas which can provoke endophytic communities to be released too.

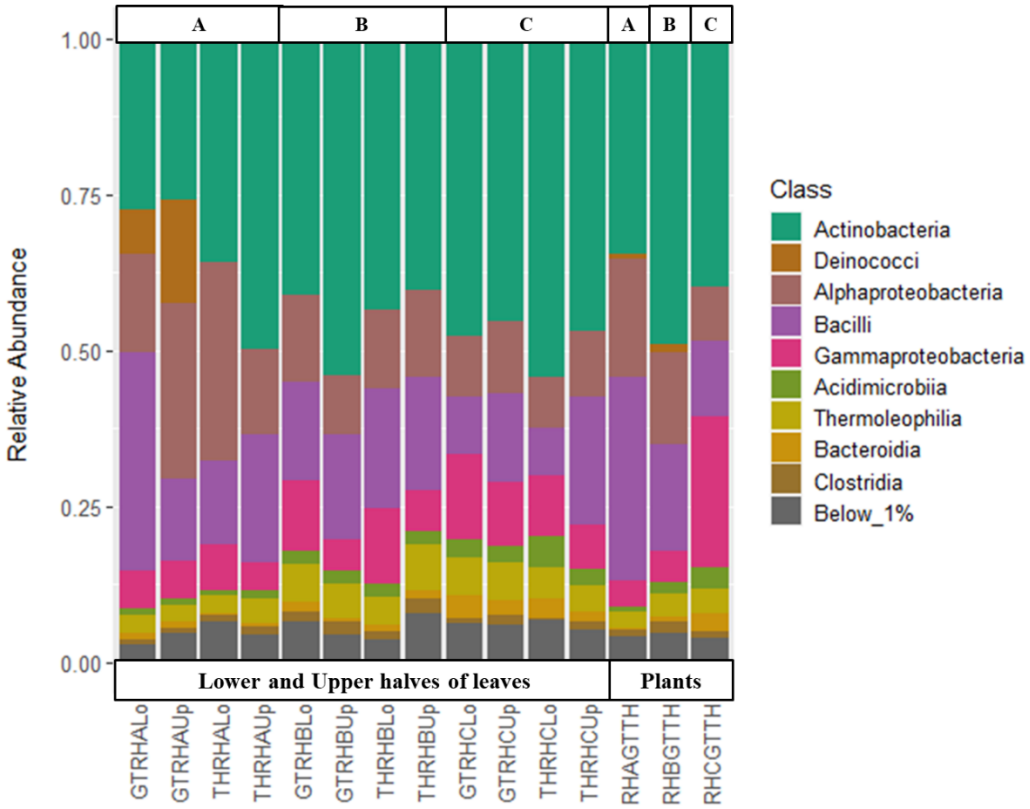


Figure 3.8. Bacterial community composition at class level of romaine lettuce plants grown under high (A), medium (B) and low (C) relative humidity (RH) conditions. Labels in x-axis correspond to samples separated by lettuce cultivar (GT-Green Tower, TH-True Heart) and half of leaf (Lo-lower, Up-upper half). Whole plants (Plants) from each RH condition were included for comparison. Classes with less than 1% average abundance across samples were grouped together and represented as Below 1%.

A summary of the average relative abundance of the bacterial communities at the class level is presented in Table 3.4. At this high taxonomic rank, no dramatic changes in the general structure of bacterial communities can be observed among the experimental conditions of RH, cultivars, and distribution on the leaf. Only two changes were noticeable (bold numbers), both under humid condition A: 1. Increase of the class Deinococci across the entire leaf surface (both halves) of GT cultivar only; 2. Increase of the class Bacilli only on the lower half of leaves, (the same increase

was observed in whole plants). As many diverse bacterial genera can be found within these classes, the significance of their increase was evaluated later in detail through differential abundance analysis at lower taxonomic ranks.

Table 3.4. Summary of the average relative abundance of bacterial communities at class level under the different experimental conditions of relative humidity (RH), cultivar (GT-Green Tower, TH-True Heart) and distribution on the leaf. Noticeable changes are in bold numbers.

Experimental conditions	Parameters			Relative abundances (%) of bacterial classes									
	RH condition	Cultivar	Distribution	Actinobacteria	Bacilli	Alpha proteobacteria	Gamma proteobacteria	Deinococci	Thermoleophilia	Acidimicrobia	Bacteroidia	Clostridia	Below 1%
GTRHALo	A	GT	Lower half	27.5	**34.8**	16.0	6.24	**6.93**	2.69	1.08	1.20	0.71	2.82
GTRHAUp	A	GT	Upper half	25.9	13.3	28.1	5.90	**16.5**	2.64	1.10	1.00	0.84	4.66
THRHALo	A	TH	Lower half	35.8	13.3	31.9	7.30	0.11	2.74	1.03	0.36	1.03	6.43
THRHAUp	A	TH	Upper half	49.7	20.6	13.6	4.58	0.12	3.96	1.30	0.41	1.43	4.31
GTRHBLo	B	GT	Lower half	41.1	15.8	13.9	11.3	0.03	6.05	2.18	1.51	1.66	6.45
GTRHBUp	B	GT	Upper half	54.0	16.9	9.55	4.80	0.07	5.48	2.24	0.67	1.86	4.51
THRHBLo	B	TH	Lower half	43.4	19.2	12.7	12.0	0.11	4.39	2.16	1.23	1.23	3.59
THRHBUp	B	TH	Upper half	40.2	18.2	13.9	6.50	0.08	7.22	2.21	1.27	2.55	7.73
GTRHCLo	C	GT	Lower half	47.6	9.02	9.90	13.7	0.02	5.99	3.05	3.83	0.54	6.35
GTRHCUp	C	GT	Upper half	45.3	14.4	11.5	10.1	0.06	5.98	2.80	2.33	1.57	6.04
THRHCLo	C	TH	Lower half	54.3	7.79	7.99	9.65	0.02	5.21	4.94	2.99	0.34	6.75
THRHCUp	C	TH	Upper half	46.8	20.4	10.6	7.15	0.06	4.37	2.54	1.40	1.29	5.29
RHAGTTH	A	GT TH	Plant	34.7	**32.5**	19.1	4.32	0.62	2.67	0.69	0.29	0.96	4.2
RHBGTTH	B	GT TH	Plant	49.0	17.1	14.6	5.04	1.40	3.58	1.88	0.99	1.67	4.74
RHCGTTH	C	GT TH	Plant	39.8	12.3	8.54	24.1	0.11	4.15	3.28	2.72	1.25	3.74

The total 45 bacterial classes identified in romaine lettuce leaves under three different RH conditions comprised 199 families and 417 genera. As a whole community, the highest relative abundance (30.9 ± 5.6%) was represented by all families that were grouped together because they accounted for less than 1% of abundance in each experimental condition. The second-most abundant family was *Streptomycetaceae* (9.89 ± 3.20%), followed by families with higher variation among the experimental conditions such as *Nocardioidaceae* (9.76 ± 7.60%), *Pseudonocardiaceae* (6.36 ± 6.53%), *Staphylococcaceae* (5.37 ± 7.49%) and *Bacillaceae* (5.09 ± 2.54%). As families with low abundances occupied the first place in the composition of resident bacterial communities, as well as high variation that was seen among RH and cultivars when all experimental conditions were quantified together, then bacterial diversity and differential abundance analysis were performed for each condition separately.

3.4.5 Impact of relative humidity (RH) condition during plant growth on the composition of resident bacterial communities in romaine lettuce

Measurements of α diversity (Richness and Shannon Index) revealed a significant difference in the diversity of resident bacterial communities explained by the RH condition where both cultivars of romaine lettuce were grown (Richness: ANOVA: $F_{2,120}$: 66.2, $P<0.0001$; Shannon Index: ANOVA: $F_{2,120}$: 65.1, $P<0.0001$). Humid condition A produced lettuce plants with significantly lower bacterial α diversity (Figure 3.9), however we previously found that plants under this humid condition also exhibited significantly higher culturable bacterial population counts (Table 3.2). Therefore, we suggest that very humid conditions during plant development promoted bacterial growth that was dominated by few taxa, causing a decline of the diversity within the communities. The same effect has been observed even on non-living surfaces where wetted materials showed this diversity reduction (145). To determine if the lettuce cultivar also explained differences in the bacterial α diversity, this factor was tested under each RH condition separately. No significant differences were found between GT and TH cultivars under humid condition A and intermediate condition B; while under dry condition C, the α diversity of bacterial communities from each cultivar was affected differently (Richness: $P=0.0257$; Shannon Index: $P=0.0354$) (Figure 3.9). This finding highlights how the plant genotype and phenotype factors have an effect on the modulation of the diversity of the bacterial microbiome under a more severe

condition of RH, which was causing a larger stress on the lettuce plants as it was evidenced previously by the VPD (Figure 3.4 A).

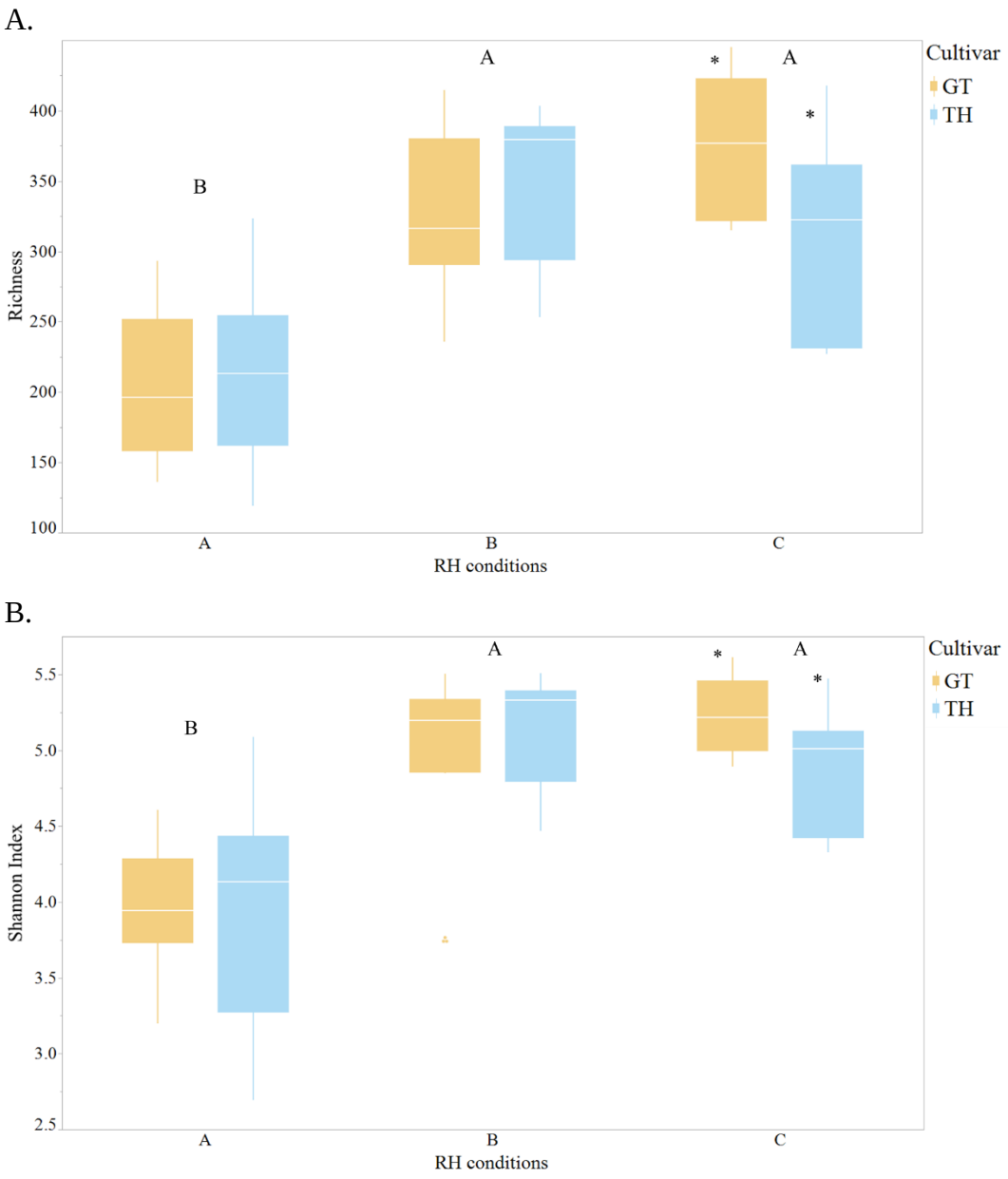


Figure 3.9. Bacterial α diversity of leaves of romaine lettuce plants grown under three different relative humidity (RH) conditions (A: humid, B: intermediate. C: dry). Boxplots are colored by the lettuce cultivar. A. Richness values. B. Shannon diversity indices.

As previously described, romaine lettuce plants suffered changes in their leaf properties and traits in response to the RH condition in which they were grown. Both cultivars expressed the same tendency of adaptations in terms of leaf wettability, stomata and cuticle resistance, however the strength of those responses was significantly different between the cultivars for some leaf properties. For instance, under the dry condition C which caused a significantly lower bacterial α diversity in cultivar TH than in GT, some leaf properties were also significantly different in TH lettuce plants such as higher stomatal density, higher stomatal conductance, higher transpiration, higher photosynthesis rates and lower bacterial population counts when compared to GT plants (p-values in Table 3.2). For that reason, the correlations between bacterial α diversity and leaf properties were evaluated across all RH conditions together and separately (Table 3.5).

Table 3.5. Correlations estimated by the Row-wise method between bacterial α diversity (Shannon Index) and leaf properties of romaine lettuce plants grown under different relative humidity (RH) conditions (A: humid, B: intermediate. C: dry).

Leaf properties	All RH conditions		RH A		RH B		RH C	
	R	P	R	P	R	P	R	P
SD (#/mm^2)	**0.3716**	**<.0001**	0.0496	0.7553	-0.0716	0.6872	**0.5043**	**0.0017**
SL (µm)	**-0.4343**	**<.0001**	-0.0084	0.9577	-0.0368	0.8365	0.0167	0.9231
SW (µm)	**-0.3160**	**0.0007**	0.1233	0.4365	0.0458	0.797	0.3651	0.0286
Bacteria (Log CFU/g lettuce)	**0.3497**	**0.0002**	0.0802	0.6138	-0.1665	0.3465	**0.4897**	**0.0024**
SPI (x100)	0.1927	0.0418	0.0934	0.5563	-0.1295	0.4654	**0.5844**	**0.0002**
PA_{max} (µm^2)	**-0.4503**	**<.0001**	0.0612	0.7003	-0.0047	0.9787	0.1739	0.3104
g_{smax} (mol m^{-2} s^{-1})	**0.2953**	**0.0016**	0.0642	0.6861	-0.1133	0.5236	**0.5499**	**0.0005**
CA (˚)	**0.4855**	**<.0001**	-0.076	0.6322	0.0267	0.8810	**0.4932**	**0.0022**
CW Alk (µg/g leaf)	0.0992	0.2982	-0.177	0.2622	**0.4457**	**0.0082**	0.0539	0.7550
CW FA (µg/g leaf)	**-0.2989**	**0.0014**	-0.1322	0.4041	-0.2342	0.1824	-0.4202	0.0107
CW RHO (µg/g leaf)	-0.1868	0.0486	-0.0907	0.5677	-0.0438	0.8059	-0.2620	0.1227
P (µmol CO2 m^{-2} s^{-1})	**-0.3536**	**0.0001**	0.1935	0.2196	0.2622	0.1341	-0.2855	0.0915
g_s (mol m^{-2} s^{-1})	0.1237	0.1939	**0.4414**	**0.0034**	0.3228	0.0626	-0.1014	0.5563
T (mol m^{-2} s^{-1}) x1000	**0.5301**	**<.0001**	**0.4150**	**0.0063**	0.3496	0.0427	0.0047	0.9784

R= correlation coefficient. P= p-value. Bold values indicate significant correlation (P<0.01)

Most of the significant correlations were observed when all the RH conditions were analyzed together, since this is the principal factor driving the biggest differences in leaf properties. The main leaf property which is showing a large positive association (R= 0.5301) with the bacterial α diversity is the transpiration rate. For that reason, plants grown under humid RH condition A which had the lowest transpiration rate, also exhibited the lowest bacterial diversity. The strength of this association gradually decreased when the RH condition was becoming dryer, because the driest condition C was causing significantly high transpiration rates evenly in all plants, allowing higher bacterial diversity. The high flow of water from stomata (transpiration) might be promoting the bacterial diversity through two possible routes: first by increasing the permeability and wettability of the leaf surface cuticle which provides leached nutrients and water to the bacteria in the phyllosphere (146, 147), and secondly by increasing the root uptake of nutrients which are beneficial for the plant and its associated microbiome (148). It is also important to notice that few associations between leaf properties and the bacterial α diversity were found within humid and intermediate conditions A and B, however under dry condition C were found several strong associations. This might be explained by our hypothesis that under this driest condition the plants are trying to balance their adaptations of resistance to water loss while maintaining the uptake of CO_2 for photosynthesis; for that reason, more variation might be found in the way that individual plants or different cultivars respond to this dry condition through modifications of their leaf properties and traits, which at the end correlate with the level of bacterial diversity promoted.

Changes in the composition of the resident bacterial communities of lettuce plants grown under different RH conditions were observed by unweighted Unifrac β-diversity patterns (Figure 3.10). The separation along the first axis (17.3%) is mainly explained by the driest condition C which promoted bacterial communities phylogenetically more distant to communities in RH conditions A and B. On the other hand, a second separation along the axis2 (13.9%) is exhibiting a succession from the humid condition A through the intermediate RH condition B. Also, it is noticeable that only under dry condition C, bacterial communities from the lower halves of leaves were more distant to communities from the upper halves. Therefore, as observed before with the bacterial α-diversity, dry condition C also promoted larger disruption on the composition (β-diversity) of the resident bacterial communities and their distribution across the lettuce leaves.

When PERMANOVA was applied to the bacterial β-diversity (weighted unifrac) of plants grown under all RH conditions, the main factor leading the overall variation was RH, accounting

for 24.8% ($F_{2,111}$: 22.8, P = 0.001). Although cultivar (P = 0.006) and distribution across the leaf (P = 0.004) were significant factors, they accounted for less than 5% of the overall variation. For that reason, the bacterial communities were then analyzed separately within each RH condition to test if the distribution across the leaf had some effect on their β-diversity.

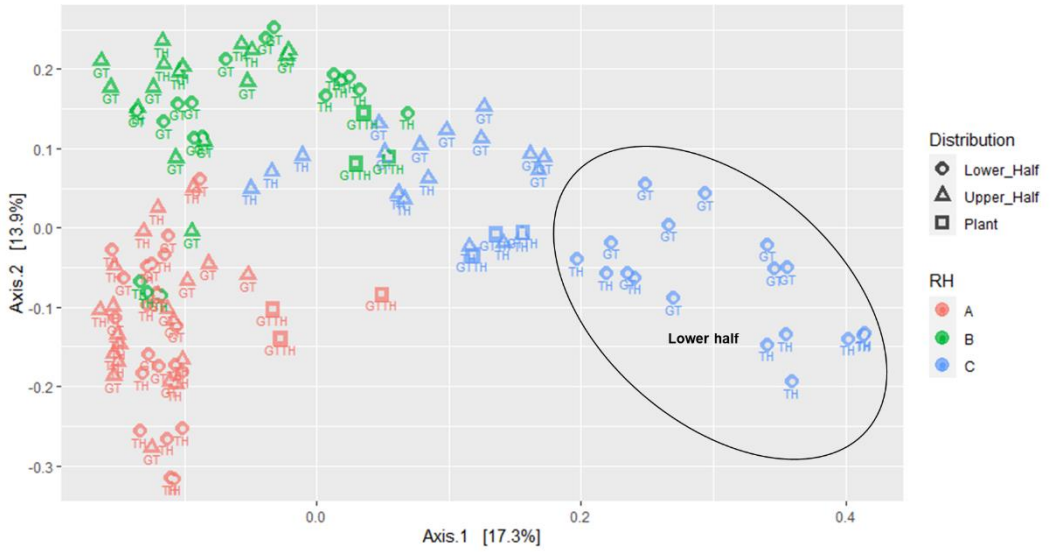


Figure 3.10 Bacterial β diversity of romaine lettuce plants grown under three different relative humidity (RH) conditions (A: humid, B: intermediate. C: dry). Visualization with principal coordinate analysis of unweighted Unifrac distances among bacterial communities. Colors indicate RH conditions as labeled in the legend. Shapes indicate the distribution where the bacterial communities come from: lower and upper half of leaf, and whole leaves from entire plant. Labels indicate the lettuce cultivar: Green Towers (GT), True Heart (TH), pool of both (GTTH).

As the RH condition used for plant growth was the main factor explaining the significant differences in the β-diversity of resident bacterial communities, differential abundance analysis (DESeq) at the family level was performed to determine how these changes were occurring (Figure 3.11). These DESeq results identified bacterial families whose differential abundance was significant (P<0.001) among the RH conditions. Interestingly, a clear succession of enriched and depleted families was observed as a gradient from humid condition A, through intermediate condition B until dry condition C.

When bacterial communities from humid condition A were compared with intermediate condition B, we observed how enriched and depleted families came from non-overlapping

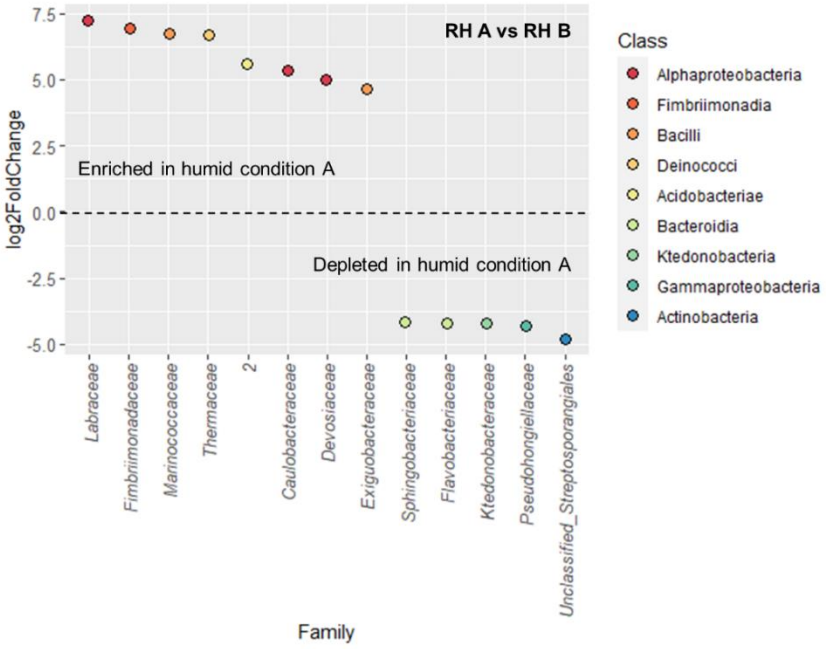
bacterial classes. For instance, families enriched by the high RH belong to classes Alphaproteobacteria, Fimbriimonadia, Bacilli, Deinoccocci and Acidobacteriae, while families depleted under high RH were from classes Bacteroidia, Ktedonobacteria, Gammaproteobacteria and Actinobacteria. One of the depleted families was *Flavobacteriaceae;* this family is known to be highly adapted to plant carbohydrate metabolism, so it is highly abundant in the phyllosphere of terrestrial plants (149, 150). Therefore, our results demonstrated that humid conditions (RH = 83 ± 7 %) can affect the presence of plant-associated bacteria which have been named as "permanent residents" for their common occurrence on many plant species and previously reported resistance to modifications of plant factors such as wax phenotype (151).

When bacterial communities from humid condition A were compared with dryer conditions B and C, we found a consistent differential abundance of families *Labraceae, Marinococcaceae,* and *Thermaceae.* These families were uniquely enriched by the humid condition and exhibited a significant decline when RH was lowered. The same was observed with family *Halomonadaceae*, which was consistently depleted under the driest condition C when compared with more humid conditions A and B. Interestingly, all these families with the exception of *Thermaceae* are described as moderately halophilic bacteria, which suggests that lettuce plants grown under humid conditions might be increasing the salt concentrations on their leaf surface or providing a more osmotically challenging environment due to the low transpiration rates. The first mechanism has been reported as part of plant adaptations to release excess salts through salt glands (152). Future studies on this subject are encouraged as they can elucidate bacterial biomarkers for salinity stress or bacterial microbiomes that are more tolerant to salt concentrations which can help plant hosts grown under environmental conditions promoting salt accumulation.

Finally, when bacterial communities from dry condition C were compared with more humid conditions A and B, we found that most of the enriched families in dry conditions belonged to the class Bacteroidia. Also, families *Spirosomaceae, Saccharimonadaceae,* and *Phaselicystidaceae* were uniquely enriched under dry conditions and exhibited a significant decline when RH was increased. Class Bacteroidia is comprised of copiotrophic bacteria, which are organisms found in environments rich in nutrients (specially carbon) and their abundance has been associated with available nutrient supplies (153). As described before, our plants grown under dry condition C exhibited the highest transpiration and photosynthesis rates, which physiologically encouraged a higher root uptake of water with nutrients, higher production of carbon sources such as glucose

and a higher amounts of leached compounds through the leaf cuticle. Therefore, these leaf properties and traits expressed by our lettuce plants under low RH conditions accompanied with enough water supply can promote the enrichment of the Bacteroidia class.

A.



B.

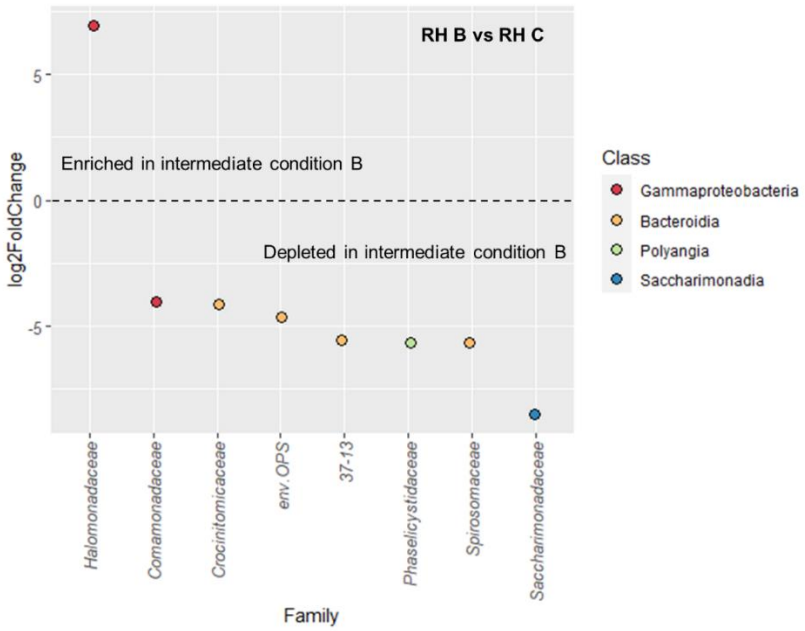
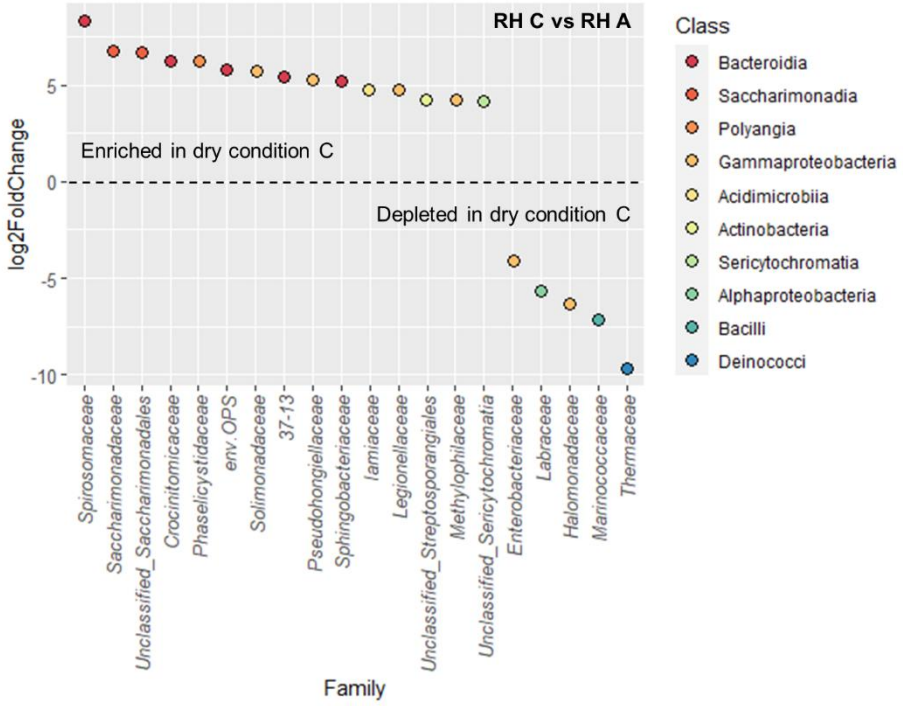


Figure 3.11. Differential abundances based on log2 fold changes of bacterial families among the RH conditions (A: humid, B: intermediate, C: dry) used for plant growth. (A) Humid vs. intermediate condition. (B) Intermediate vs. dry condition. (C) Each point represents one family

(x-axis), colored by class. Points above the dashed line represent families significantly enriched, and points below represent depleted families (log2 > 4, P < 0.001).

Figure 3.11 continued

C.



3.4.6 Distribution of resident bacterial communities across leaves from lettuce plants grown under different relative humidity conditions

Measurements of α diversity (Richness and Shannon Index) revealed a significant difference in the diversity of resident bacterial communities explained by the distribution across the leaf surface only in plants (both cultivars) grown under the driest condition C (Richness: ANOVA: $F_{1,35}$: 45.6, P<0.0001; Shannon Index: ANOVA: $F_{1,35}$: 47.7, P<0.0001). However, when cultivars were analyzed separately, additional significant differences between the lower and upper halves of leaves were found on cultivar TH under RH condition A (only Richness: P=0.0086) and condition B (only Shannon Index: P=0.0043). In all previous significant differences, upper half of the leaf exhibited a higher bacterial α diversity (Figure 3.12). This result reflects the significant correlations found between bacterial α diversity and leaf properties (Table 3.5), where SD, SPI and CA exhibited a positive strong association (R= 0.5043, 0.5844 and 0.4932, respectively) indicating that increase of these properties could explain the increase of α diversity. These leaf

properties were also found to be significantly different between lower half and upper half of leaves (Table 3.3), where upper half exhibited higher values of SD, SPI, and CA. Our data therefore suggest that leaf surface areas where there is an increase of stomatal density with consequent increase of stomatal pore index, will experience a decrease of wettability (high CA) and promote a higher bacterial α diversity.

A.

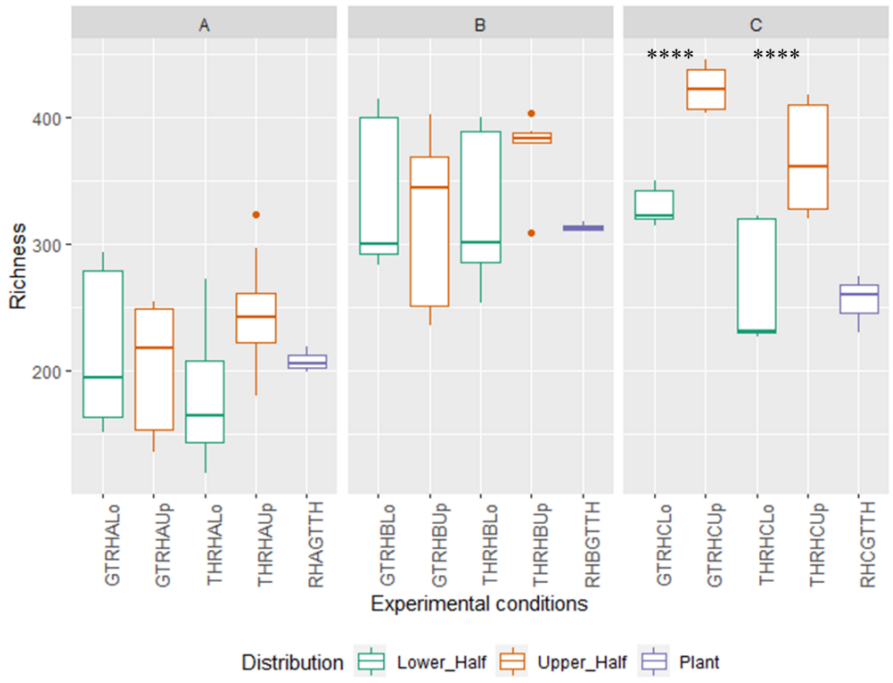
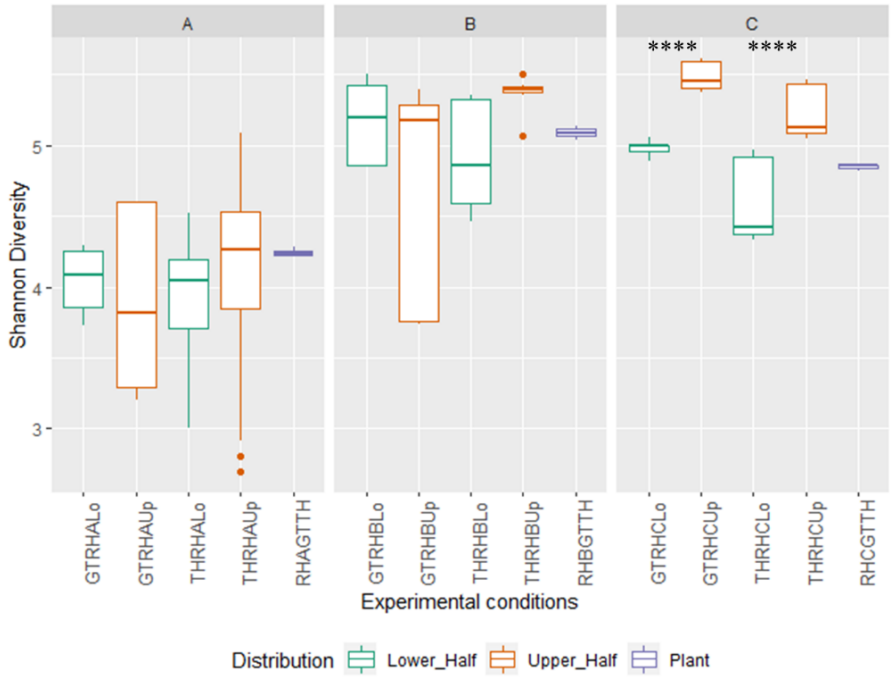


Figure 3.12. Bacterial α diversity of resident communities from different distributions across the romaine lettuce leaf (lower and upper half). A. Richness values. B. Shannon diversity indices. Panels correspond to relative humidity (RH) conditions used for plant growth (A: humid, B: intermediate. C: dry). Boxplots are colored by the distribution across the leaf (Up: upper half, Lo: lower half, Plant: whole leaves from entire plant). Experimental conditions indicate the lettuce cultivar (GT: green towers, TH: true heart, GTTH: pool of entire plants from both).

Figure 3.12 continued

B.



As observed before in Figure 3.10, changes in the composition of the resident bacterial communities of romaine lettuce plants were mainly explained by the RH condition used for plant growth. Also, a greater dispersion was only exhibited by dry condition C, for those reasons β-diversity analysis was performed exclusively for this condition to determine if that dispersion within bacterial communities was caused by the distribution on the leaf (Figure 3.13). In this case, the larger separation along the first axis (31.9%) is mainly explained by the distribution on the leaf of the resident bacterial communities. Therefore, a low level of relative humidity (condition C) promoted shifts in the composition (β-diversity) of the resident bacterial communities across the lettuce leaf surface, as well as promoting the significant differences in bacterial α diversity reported before.

When PERMANOVA was applied to the bacterial β-diversity (weighted unifrac) from all RH conditions together, the distribution across the leaf accounted for less than 2% of the overall variation ($F_{2,111}$: 91.4, P=0.001). However, distribution combined with leaf properties explained more of the variation; for instance, distribution together with the stomal size (D*SL*SW) accounted for 10.8% ($F_{1,111}$: 524, P=0.001). Also, distribution together with amounts of fatty acids and alcohols (D*FA*ALCOHOL) accounted for 14.0% ($F_{1,111}$: 682, P=0.001). Based on these

results, the distribution of the bacterial communities across the leaf surface by itself does not explain much of the variation of the bacterial composition through all RH conditions, because each RH condition caused different changes and distributions of the leaf properties. Therefore, it seems probable that localization on the leaf must be accompanied by specific leaf properties and traits happening in that particular niche to explain the bacterial β-diversity shifts.

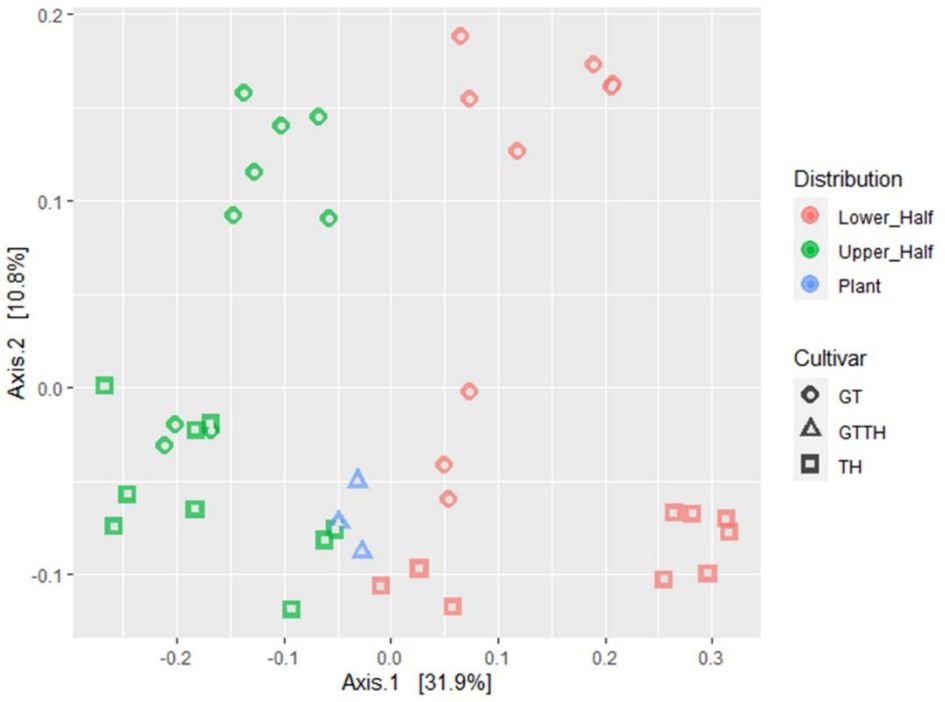


Figure 3.13 Bacterial β diversity of romaine lettuce plants grown under relative humidity (RH) condition C (dry). Visualization with principal coordinate analysis of unweighted Unifrac distances among bacterial communities. Colors indicate the distribution where the bacterial communities come from: lower and upper half of leaf, and whole leaves from entire plant. Shapes indicate the lettuce cultivar: green towers (GT), true heart (TH), pool of both (GTTH).

3.4.7 Identification of plant-associated bacterial genera linked to different levels of relative humidity during plant growth

To identify specific taxa at the genus-level or ASV-level which can be considered biomarkers for the contrasting RH conditions during plant growth, linear discrimination effect size (LEfSe) analysis was performed. This analysis detected taxa with significant differential abundance among the RH conditions through the Kruskal-Wallis rank test, and then evaluated the relevance or effect size of those taxa by Linear Discriminant Analysis (LDA) (154). As observed

in Figure 3.14, more significant genera were differentially abundant under dry condition C, which enables the discrimination of bacterial communities from plants grown under low RH conditions causing a higher vapor pressure deficit. Therefore, the increased stress caused by low levels of RH produced larger shifts in the composition of plant-associated bacterial communities as demonstrated by the PERMANOVA results of comparisons RH A vs RH B, where the condition of RH only accounted for 13.5% ($F_{1,75}$: 11.7, P=0.001) of the total variation, while in comparisons against the dry condition RH A vs RH C and RH B vs RH C accounted for 23.6% ($F_{1,77}$: 24.6, P=0.001) and 23.4% ($F_{1,69}$: 25.8, P=0.001), respectively.

These bacterial genera identified as biomarkers for levels of relative humidity (RH) during romaine lettuce growth exhibited two behaviors: 1. The genus was present under all RH conditions and exhibited a significant gradual differential abundance through the decrease/increase of RH; and 2. the genus was not present in one RH condition, then appeared and increased its abundance according to the increase/decrease of RH. The first behavior was exhibited by all biomarkers from RH condition A, for instance in the case of *Pseudonocardia*, this genus was part of the microbiome core of RH condition A exhibiting a high prevalence across plants, but once RH started to decrease its prevalence and abundance decreased too (Figure 3.15 A). Biomarkers from intermediate condition B also exhibited this type of behavior of gradual differential abundance, *Bacillus* is presented as an example of this condition (Figure 3.15 B). Finally, biomarkers from dry condition C exhibited both types of behavior, for instance *Novosphingobium* showed the first described behavior (Figure 3.15 C), while *Dyadobacter* exhibited the second type (Figure 3.15 D).

Lastly, we detected taxa that enabled the discrimination of the lower half of leaves from plants grown under dry condition C (Figure 3.16). Interestingly, these taxa identified as biomarkers for distribution in leaves, belonged to the same bacterial genera identified as biomarkers for the low level of RH (condition C). This suggests that the plant responses which were promoted by the low RH, promoted growth of these new bacterial taxa in the lower half of leaves, which also had a significantly lower α diversity, as previously described. We hypothesize that the high transpiration rates observed under this low RH condition promote a rapid root uptake of water and nutrients (122), which created a rich environment that selects for fast-growing bacterial taxa which is expected to lower the bacterial diversity (155, 156). Consequently, our biomarkers identified under condition C comprised genera from classes Bacteroidia and Gammaproteobacteria, which

are known to be copiotrophic bacteria, as well as genera known as saccharolytic bacteria such as *Saccharibacteria TM7* and *Conexibacter.*

A.

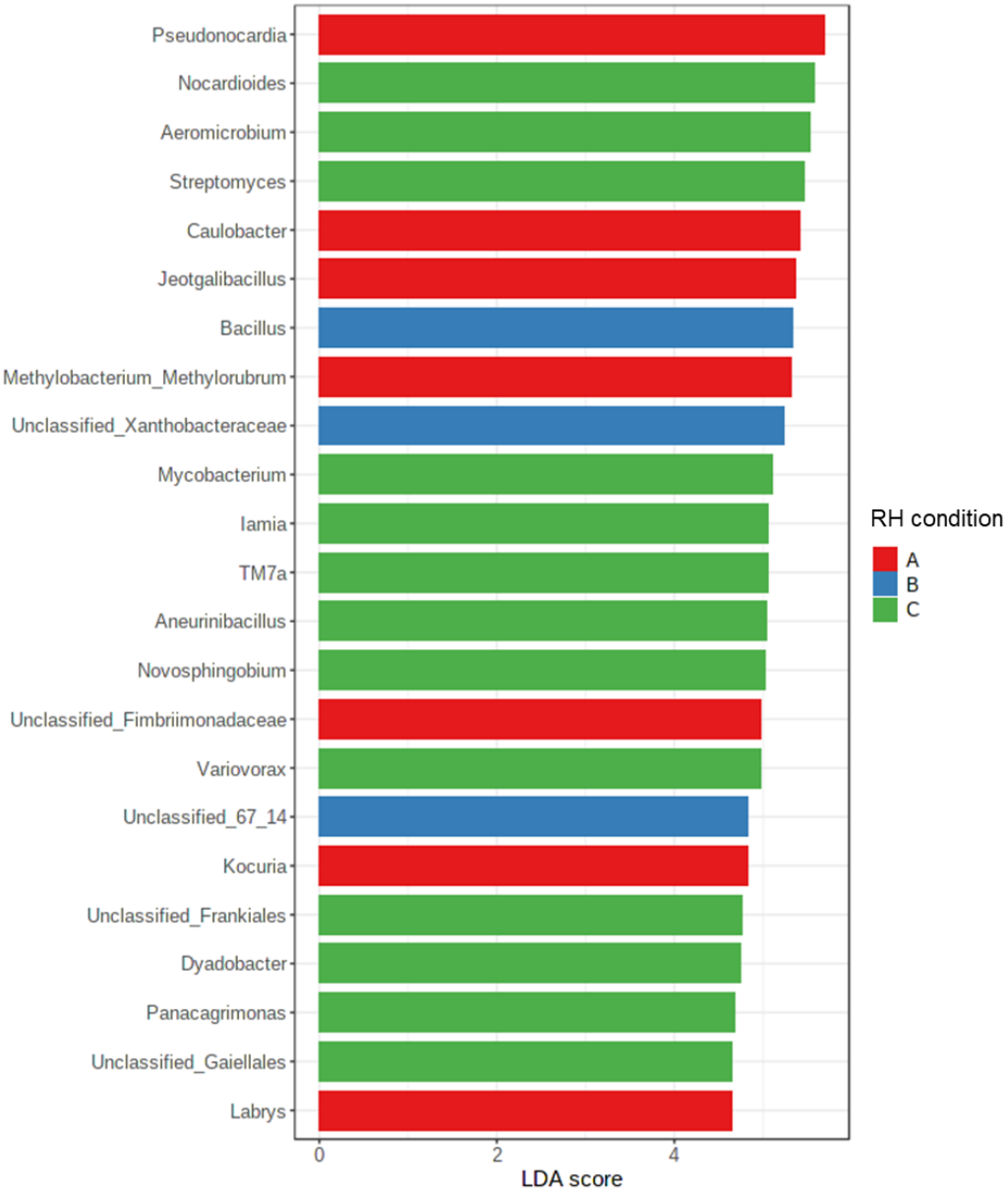


Figure 3.14. Genera identified as biomarkers of different levels of relative humidity (RH) during plant growth: humid (condition A), intermediate (condition B) and dry (condition C). Linear discrimination analysis (LDA) effect size (LEfSe) for comparisons of RH conditions. The length of the bar represents the log10 transformed LDA score. The threshold on the LDA score for discriminative features was set to 4.0, and the taxa with statistically significant change (P < 1E-08) was selected.

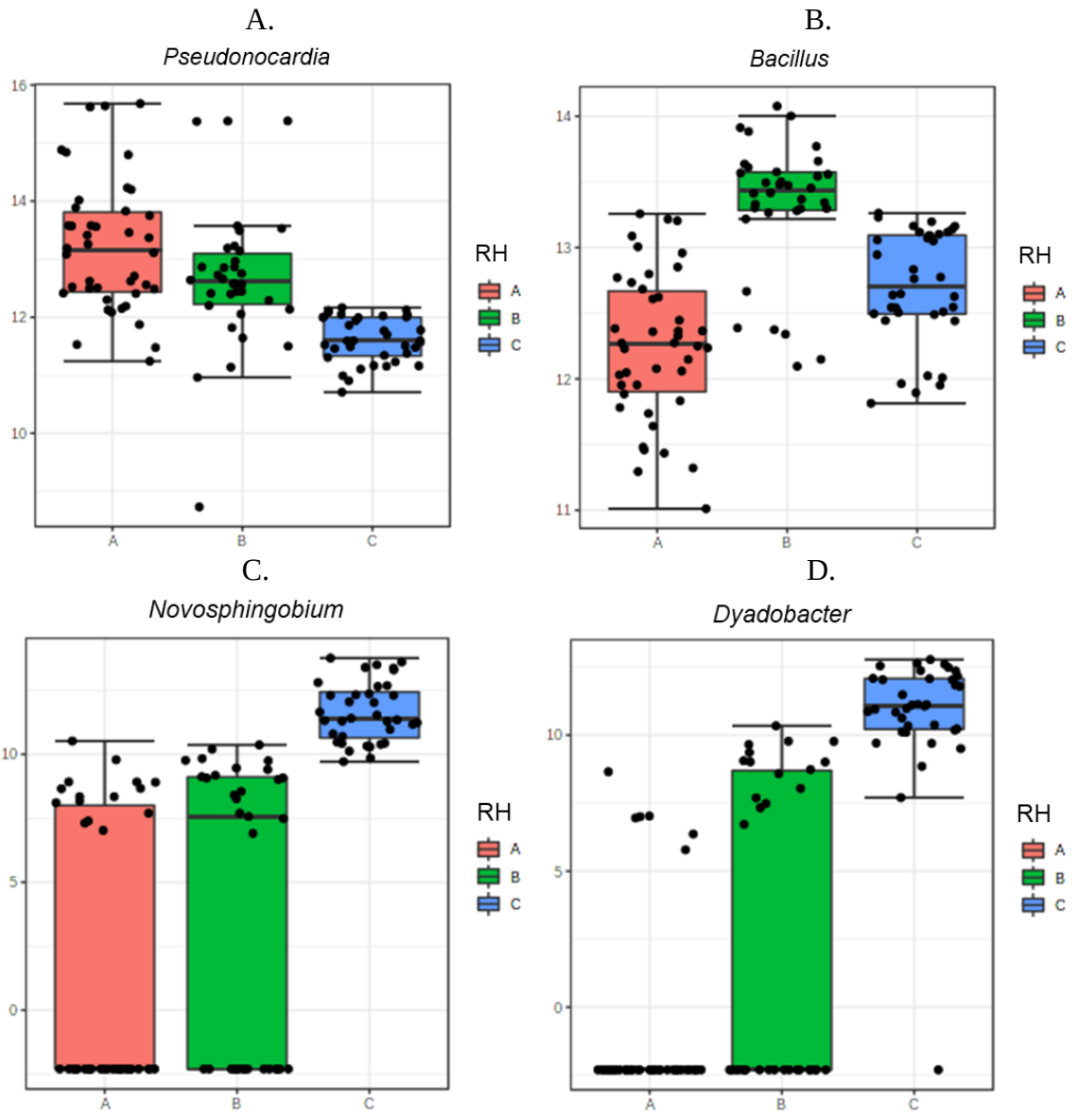


Figure 3.15. Abundance (log-transformed count) of some genera identified as biomarkers of different levels of relative humidity (RH) during plant growth. (A) Biomarker of humid condition A. (B) Biomarker of intermediate condition B. (C-D) Biomarkers of dry condition C.

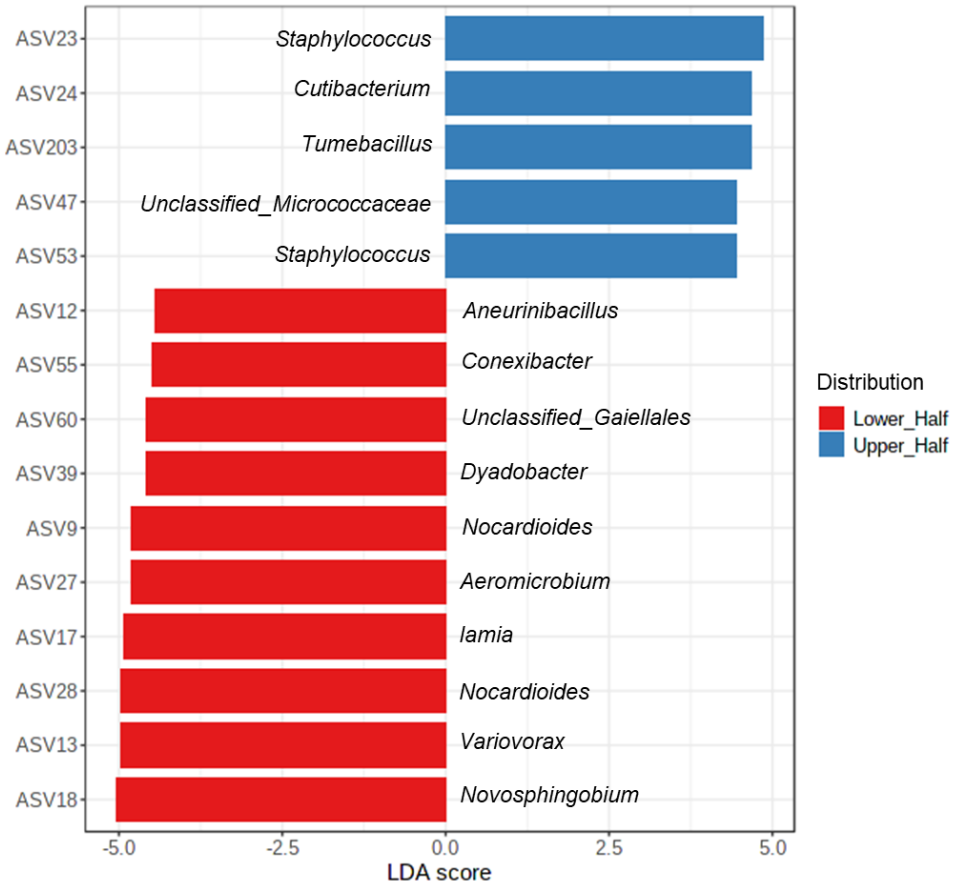


Figure 3.16. Bacterial taxa identified as biomarkers of distribution across the leaves of romaine lettuce plants grown under dry condition C. Linear discrimination analysis (LDA) effect size (LEfSe) for comparison of distribution on leaf. The length of the bar represents the log10 transformed LDA score. The threshold on the LDA score for discriminative features was set to 4.0, and the taxa with statistically significant change ($P < 0.01$) was selected.

Overall, we demonstrated that variations only in the level of relative humidity (RH) during the sowing and growth cycles of romaine lettuce plants can produce significant changes in leaf properties and traits. These leaf properties significantly varied across the leaf surface under all RH conditions. However, the distribution of leaf properties might be associated with the diversity and localization of the resident bacterial communities across the leaves, when the environmental and culturing conditions produced a heterogeneous stimulus. This heterogeneous stimulus was achieved under our condition C where the low RH produced significantly higher transpiration rates, but at the same time our plants were watered in the required amount to be healthy. Therefore, we observed plant responses that balanced resistance to water loss with the normal function of

photosynthesis. Our findings provide useful fundamental knowledge of the microbial ecology happening in romaine lettuce, which can lead to development of crop management strategies to promote a healthy and protective microbiome. Additionally, the understanding of how environmental conditions affect our crops and their microbiome is crucial to be prepared for the effects of climate change. Finally, the identification of bacterial biomarkers for environmental conditions provides tools for identification of fresh produce grown under environments that for example might be associated with greater risk of contamination with foodborne human pathogens.

3.5 Supplemental Material

Table S1. Compounds identified and quantified from cuticular waxes were grouped in three classes: alkanes, fatty acids and primary alcohols.

Class of compound (# carbons)	Compound name
Alkane C12	Dodecane, 1-fluoro-
Alkane C12	Dodecane, 4,6-dimethyl-
Alkane C14	Tetradecane
Alkane C15	Pentadecane
Alkane C16	Hexadecane
Alkane C17	Heptadecane
Alkane C19	Nonadecane
Alkane C20	Eicosane
Alkane C21	Heneicosane
Alkane C22	Docosane
Alkane C23	Tricosane
Alkane C24	Tetracosane Standard
Alkane C25	Pentacosane
Alkane C27	Heptacosane
Alkane C29	Nonacosane
Alkane C31	Hentriacontane
Alkane C34	Tetratriacontane
Alkane C35	Pentatriacontene
Fatty Acid C7	Heptanoic acid
Fatty Acid C9	Nonanoic acid
Fatty Acid C12	Dodecanoic acid (lauric acid)
Fatty Acid C14	Tetradecanoic acid (myristic acid)
Fatty Acid C15	Pentadecanoic acid
Fatty Acid C16	Hexadecanoic acid (palmitic acid)
Fatty Acid C18	alpha.-Linolenic acid
Fatty Acid C18	Petroselinic acid

Table S1 continued

Fatty Acid C18	Octadecanoic acid (stearic acid)
Fatty Acid C22	Docosanoic acid (behenic acid)
Fatty Acid C24	Tetracosanoic acid (lignoceric acid)
Fatty Acid C24	Tetracosenoic acid (nervonic acid)
Fatty Acid C26	Hexacosanoic acid
Primary alcohol C10	1-Decanol, 2-hexyl-
Primary alcohol C12	1-Dodecanol
Primary alcohol C18	1-Octadecanol
Primary alcohol C19	18-Methyl-nonadecanol
Primary alcohol C20	1-Eicosanol
Primary alcohol C22	Docosanol (Behenic alcohol)
Primary alcohol C23	1-Tricosanol
Primary alcohol C24	1-Tetracosanol
Primary alcohol C24	22-Methyltetracosanol
Primary alcohol C25	1-Pentacosanol
Primary alcohol C26	1-Hexacosanol
Primary alcohol C27	Anteiso-heptacosanol
Primary alcohol C28	1-Octacosanol

CHAPTER 4. GROWTH OF *ESCHERICHIA COLI* O157:H7 ON ROMAINE LETTUCE LEAVES IS AFFECTED BY ENVIRONMENTAL CONDITIONS OF RELATIVE HUMIDITY, LEAF PROPERTIES, AND COMPOSITION OF NATIVE BACTERIAL COMMUNITIES

4.1 Abstract

The success of a human pathogen such as *Escherichia coli* O157:H7 to grow and survive as a plant-associated microbe is modulated by its interaction with the plant host, with the resident microbial communities and its adaptability to the environmental conditions. We found that environmental conditions in which romaine lettuce plants are grown, as well as at the time of inoculation, influence their susceptibility to allow growth of *E. coli* O157:H7. Also, we were able to achieve actual growth of population numbers of the pathogen on intact leaves when a low inoculum dose and less adverse relative humidity (RH) conditions are provided, demonstrating a correlation between increase of die-off rates and high inoculum dose. The growth of *E. coli* O157:H7 (log change CFU) varied significantly ($P<0.05$) between upper and lower halves of lettuce leaves, as well as with leaf wettability defined by contact angle, which might explain the higher population counts found in the upper area. Relative humidity (RH) was the main factor modulating the fate of *E. coli* O157:H7 and the composition of resident bacterial communities on romaine lettuce. *Microbacterium* and one unclassified genus from the *Rhizobiaceae* family were found as biomarkers of bacterial communities where *E. coli* O157:H7 reached higher and lower population counts, respectively. Overall, high relative humidity levels, low inoculum of *E. coli* O157:H7 and localization on the upper halves of the leaves are demonstrated to promote higher growth of this human pathogen on romaine lettuce. Based on these results, evaluation of best strategies for crop management such as avoidance of sprinkler irrigation are encouraged, as well as the use of resident bacterial communities as potential biomarkers for *E. coli* O157:H7 contamination.

4.2 Importance

Infections with the human pathogen *Escherichia coli* O157:H7 cause 73,000 illnesses, 2,200 hospitalizations and 60 deaths annually in the United States. The annual economic cost of this foodborne disease is around $405 million dollars including deaths, medical care and lost

productivity (157). Efforts to understand what factors might influence the colonization, growth and persistence of this human pathogen on edible plants are warranted. For those reasons, we evaluated the growth of *E. coli* O157:H7 on romaine lettuce leaves under different environmental conditions of relative humidity (RH). This study demonstrates how the growth of *E. coli* O157:H7 is affected by RH, how it was different across location areas of the same lettuce leaf according to the distribution of leaf properties, and how contrasting high and low growth of the pathogen lead to specific changes in the composition of native bacterial communities.

4.3 Materials and Methods

4.3.1 Bacterial strain and culture conditions

The strain *Escherichia coli* O157:H7 B6-914, which does not produce Shiga-like toxins Stx1 and Stx2, was used for laboratory safety advantages (87). To track the inoculated pathogen in lettuce plants, this strain was transformed with the plasmid pGFP (cDNA vector, Clontech, Mountain View, CA, USA) to express green fluorescent protein (GFP) and an ampicillin resistance gene for selection (88). The presence of GFP in *E. coli* O157:H7 had no effect on its behavior when compared with the parent strain (89) and the lack of toxin genes had no influence on the bacterial growth (90).

Inoculum was prepared by streaking frozen GFP-labeled *E. coli* O157:H7 onto Luria Bertani (LB) agar supplemented with 100 µg/ml of ampicillin (LBA-Amp) and incubating at 37˚C for 24 h. An individual colony was transferred to one milliliter of LB broth with 100 µg/ml ampicillin (LB-Amp) and incubated at the lower temperature of 30˚C to acclimatize the bacterial strain for subsequent plant inoculations. Incubation of this liquid culture was carried out for 7 h in constant agitation (250 rpm) to the early stationary phase of growth, when optical density at 600 nm (OD_{600}) was around 1.5. For accurate quantification of *E. coli* O157:H7 inoculum, this liquid culture was serially 10-fold diluted in phosphate buffer (PB) 0.1M pH 7.0 and plated onto LBA-Amp. Plate count was performed after 16 h of incubation at 37˚C, while the liquid culture was stored at 4˚C. Based on this first quantification of viable culturable cells, our liquid cultures ranged between 1.0 – 9.9 x 10^8 CFU/ml, then dilutions were freshly prepared in PB 0.1M pH 7.0 to yield 10^3 and 10^6 CFU in 50 µl. These inoculum doses were plated onto LBA-Amp right before being used to inoculate lettuce plants, so the exact number of CFU placed onto lettuce leaves was known.

4.3.2 Romaine Lettuce Source

Two cultivars of romaine lettuce were used, Green Tower (GT) and True Heart (TH). Seeds for these cultivars were purchased from Urban Farmer (Westfield, IN, US) and Synergene Seed & Technology, Inc. (Salinas, CA, US), respectively. The seeds were sown in trays using a soil mix prepared in a 2:1 ratio by volume of PRO-MIX FLX growing medium (Premier Tech Horticulture, PA, USA) and field soil (sandy loam collected from the top 10 cm of a conventional agricultural field in Indiana). Three weeks after germination, seedlings were transplanted into pots (15.5 cm x 17.8 cm) containing the same soil mix with 2 grams of fertilizer (Osmocote® Smart-Release®). Plants were grown and tested between 6-8 weeks after sowing the seeds, using always the fully developed first and second emerged leaves. Plants were not subjected to water deprivation; they were watered with the frequency required to avoid the soil mix becoming dry under each environmental condition. Water was added directly on top of the soil avoiding its accumulation in the tray placed under the pots. For every environmental condition, two batches of growth (seeds of both cultivars were sown at two different dates) were performed to carried out all the experimental evaluations.

4.3.3 Growth chamber conditions for cultivation of romaine lettuce and inoculation of *E. coli* O157:H7

Seed germination, plant growth and plant inoculations were performed in an environmental growth chamber Percival AR-75L which was programmed for a 16-h light cycle (130 µmol m^{-2} s^{-1}) and 8-h dark cycle at constant temperature of 24˚C (SD= 1˚C). Three different conditions of relative humidity (RH) were evaluated: A. 83% (SD= 7.0), B. 62% (SD=9.0), C. 43% (SD= 7.4). Conditions A and B were achieved through the growth chamber performance, while condition C required the introduction of an external dehumidifier. Environmental conditions were recorded with a HOBO MX2301 Temp/RH data logger (Figure 4.1, Table 4.1).

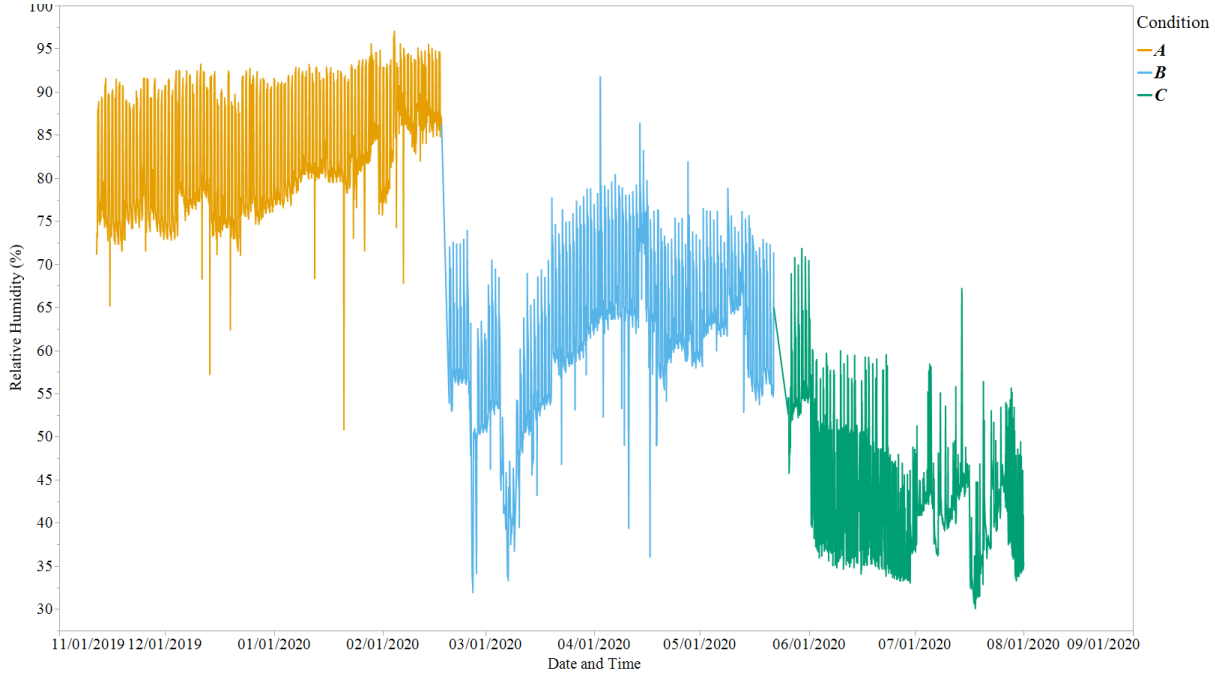


Figure 4.1. Relative humidity conditions in growth chamber used for seed germination, plant growth and experimental studies.

Table 4.1. Environmental data recorded with a HOBO MX2301 Temp/RH data logger from the growth chamber conditions A, B and C.

Condition	Planting date	Inoculation date	Relative humidity (%)				Temperature (˚C)			
			Max	Min	Median	Mean	Max	Min	Median	Mean
A	11/01/19 11/27/19 12/04/19 12/22/19	12/14/19 01/21/20 01/26/20 02/05/20	96.9	51.0	81.1	83.0	26.4	21.5	24.0	24.2
B	02/26/20 03/02/20	04/09/20 04/17/20	91.5	32.1	62.2	61.6	26.0	20.6	23.8	23.8
C	05/26/20 06/08/20	07/08/20 07/21/20	71.7	30.2	41.9	43.1	26.0	18.8	23.7	23.6

4.3.4 Greenhouse Conditions for Cultivation of Romaine Lettuce

For a subset of lettuce plants, their seed germination and plant growth were performed under greenhouse conditions, but their inoculations with *E. coli* O157:H7 were carried out under

previously described Condition A in growth chamber. Greenhouse facilities allowed light-controlled conditions operated on a daily cycle of 16 h of light and 8 h of darkness, with temperature-controlled conditions set to 22-24˚C. Relative humidity cannot be controlled, and RH measurements ranged between 20% to 50%. When plants reached the time of growth, they were transferred to the growth chamber for inoculation after a 24 h period of acclimatization.

4.3.5 Inoculation of *E. coli* O157:H7 on Romaine Lettuce Leaves

For romaine lettuce grown in growth chamber, 24 plants (6 GT and 6 TH from each batch of growth) were inoculated using two different inoculum doses (10^3 and 10^6 CFU of *E. coli* O157:H7), and two times of incubation (16 hours and 112 hours). Two leaves per plant were divided into 12 testing areas (Figure 4.2) using petroleum jelly for delimitation, then each area was drop-inoculated with 50 µl of PB 0.1M pH 7.0 containing the corresponding inoculum dose. The bacterial suspension was dispensed with a pipette and sterile tip by randomly dispersing droplets inside each area. Inoculations were performed at 6:00pm, so inoculum tested after 16 hours was exposed to 4 hours of light, followed by 8 hours of dark, and then 4 hours of light again.

Figure 4.2. Twelve testing areas used as sampling unit for contact angle, stomatal density and stomatal traits measurements.

4.3.6 Sample processing for plate counts of *E. coli* O157:H7 CFU

Half of the inoculated leaves were harvested 16 hours after inoculation and the rest of the leaves after 112 hours. To avoid disruption of the *E. coli* O157:H7 cells on the leaf surface, harvested leaves were placed onto a flat surface, testing areas were immediately excised using a razor blade and transferred separately to a sterile 50-ml conical tube. Afterwards, 30 ml of PB 0.1M pH 7.0 were added and samples were sonicated for 14 minutes using a Branson 5800 ultrasonic cleaner water bath (Branson Ultrasonics Corporation, Danbury, CT). Dilutions from wash buffer were prepared and plated onto LBA-Amp. After an incubation at 37˚C for 24 h, the plate counts were recorded to calculate the growth of *E. coli* O157:H7. The remaining wash buffer was stored at -80 ˚C until DNA extraction.

4.3.7 Statistical Analysis of *E. Coli* O157:H7 Growth

E. coli O157:H7 populations obtained from plate counts were log transformed and expressed as log CFU, or as growth change (log CFU) calculated as the logarithmic change between the number of CFU inoculated and the number of CFU after the time of inoculation on lettuce leaves. Analysis of variance (ANOVA) followed by Tukey's multiple comparison test were conducted to test differences of bacterial growth among the environmental conditions and the leaf testing areas. Zones of the leaves exhibiting outlier *E. coli* O157:H7 growth (very high or very low) were defined as testing areas with growth change (log CFU) values with a distance greater than one standard deviation from the mean.

4.3.8 Counts of Culturable Native Bacteria Across the Leaf

First or second emerged leaves from romaine lettuce plants between 6-8 weeks old were used to enumerate the culturable bacterial populations across the leaves. Each leaf was divided into 12 testing areas (Figure 4.2), all the leaf tissue was collected avoiding the central vein. Testing areas were excised using a razor blade and transferred separately to a sterile 50-ml conical tube. Afterwards, 30 ml of PB 0.1M pH 7.0 were added and samples were sonicated for 14 minutes using a Branson 5800 ultrasonic cleaner water bath (Branson Ultrasonics Corporation, Danbury, CT). Then, 100 µl were spread onto Plate Count Agar (PCA) plates, which were incubated at 30˚C

for 72 hours (111). After the time of incubation, colonies were counted, and the total bacterial populations expressed as log CFU.

4.3.9 Contact Angle Hysteresis Measurements

One of the first or second emerged leaves (6-8 weeks-old) was harvested per plant to analyze the interaction of water with the leaf adaxial surface. Three GT and three TH lettuce plants were analyzed from each batch of growth, for a total of 12 leaves per environmental condition. To study the differences of surface hydrophobicity across the romaine lettuce leaf, every leaf was divided into 12 testing areas (Figure 4.2). Squared pieces of tissue of less than 10 mm were excised from each testing area using a razor blade. These pieces were then taped (Scotch double sided tape) to a smooth microscope glass slide to expose the adaxial leaf surface. The water contact angle measurement was conducted by a Ramé-Hart goniometer (model 500) connected to an automated dispensing system (model 100-22, Ramé-Hart, Netcong, NJ), using the needle-in-the-sessile-drop method (112). A drop of 5 µl of deionized water was dispensed at the center of the leaf piece, then 0.25 µl were pumped into the drop every two seconds until the Advancing Contact Angle (ACA) is reached. Afterwards, 0.25 µl were pumped out until the Receding Contact Angle is reached (RCA). The contact angle hysteresis is calculated from the difference of ACA and RCA.

4.3.10 Quantification of Cuticular Waxes

The same leaves from the contact angle measurements were used to perform cuticular wax extraction and quantification. Each leaf was divided in two samples: upper half (testing areas A to F) and lower half (testing areas G to L) and processed separately. Wax extraction was performed by further cutting down each half and soaking in 20ml of Hexane solution for 1 min. The leaf tissues were removed and around 17ml ± 1 mL of clear hexane with wax content was then separated into new tubes, 100 µL of n-tetracosane (1 mg/ml) was added as internal standard (ISTD) and the extracts were subsequently evaporated under a gentle stream of nitrogen. Derivatization of the waxes was performed as described previously (113). Briefly, the wax residues were re-dissolved in a mixture of 50 µL of pyridine and 100 µL of bis-N, N- (trimethylsilyl) trifluoroacetamide containing 10% trimethylchlorosilane (TMCS; Sigma-Aldrich, St. Louis, MO) and heated at 75°C for 75 min. 1 µL of each sample was injected on to Agilent 7890B GC system

equipped with Agilent 7693 autosampler coupled to the Agilent 5977B MSD mass spectrometer. The Agilent DB-5MS capillary column (30 m × 0.25 mm; film thickness 0.25 μm) was used for the separation of waxes. The GC oven temperature program started from an initial temperature of 80°C (3 min hold) and ramped at 6°C min-1 up to 300°C with 3 min hold, the total run time was 42.6 min. Helium was used as the carrier gas at the flow of 1ml/min, the GC inlet and MS auxiliary temperatures were set at 220°C and 230°C respectively. The MS data were acquired within the range of 50-500 daltons.

Compound identification was performed using Agilent MassHunter Qualitative Analysis B.07.00 software, and NIST14 database by applying score threshold of >80% similarity against the NIST library spectra. The linear retention index calculations were performed by analyzing C8-C40 alkanes and retention indices were assigned to each compound. Quantification of each wax compound was performed as previously described (114). Briefly, the ratio between each peak area and the peak area of the ISTD was calculated, then each ratio was normalized by the total amount of ISTD (100μg) and by the weight (g) of the leaf pieces of each sample.

4.3.11 Stomatal density measurements

Seven week-old lettuce plants from both cultivars were used to characterize the stomatal density of plants grown under each environmental condition. One of the first fully expanded leaves was harvested per plant and the stomatal density was recorded on the adaxial surface across the twelve testing areas of the entire leaf (Figure 3.2). Leaf pieces of 10 x 10 mm were excised to be imaged with confocal laser scanning microscopy (CLSM) using a Zeiss LSM 880 Upright Confocal (Carl Zeiss, Jena, Germany). To alleviate the uneven topography of lettuce leaves, the leaf pieces were enclosed with CoverWell Imaging Chamber Gaskets (20-mm diameter, 0.5-mm deep, ThermoFisher Scientific). Confocal z-stack images were captured using Plan-Apochromat 20x/0.8 objective, chlorophyll autofluorescence with 633nm laser and transmitted light images with differential interference contrast (DIC).

Images were analyzed using ImageJ software (https://imagej.nih.gov/ij/). Stomatal density was evaluated by recording the number of stomata in four randomly chosen fields of view (FOV, 20x objective) from each testing area across the leaf.

4.3.12 Characterization of Native Bacterial Communities from Leaf Areas with Outlier *E. coli* O157:H7 Growth

Frozen wash buffer from leaf testing areas identified as outliers were used for DNA extraction. The leaf wash was thawed and filtered twice using 0.2 and 0.1 µm polyethersulfone (PES) filter membranes (Membrane solutions, WA, USA) to collect the bacterial cells. Filters were placed in sterile Eppendorf tubes and stored at -80˚C until DNA extraction.

For the amplicon sequencing analysis, leaf testing areas identified as outliers with high or low *E. coli* O157:H7 growth under the same conditions of lettuce cultivar, RH, inoculum dose and time of incubation were pooled together to be sequenced. A total of 32 samples (16 from Condition A, 8 from Condition B, and 8 from Condition C) were processed. Additionally, three samples of non-inoculated plants grown under conditions A, B and C, also the strain of *E. coli* O157:H7 used for inoculations, and controls for the identification of contaminant reads from the DNA extraction process with buffer extraction controls (BEC) and from the amplification method with non-template controls (NTC) were included. Finally, to perform quality control of our entire methodology from sample processing, DNA extraction, 16S rRNA amplification, MiSeq sequencing and bioinformatics processing, one mock microbial community standard (D6300) and one mock microbial community DNA standard (D6305) were included (Zymo Research, Irvine, CA, USA).

4.3.13 DNA Extraction and Library Preparation

Frozen filter membranes were minced under liquid nitrogen in the same Eppendorf tubes using disposable pellet pestles (Fisherbrand, CA, USA). The minced filters were placed directly into the PowerBead Pro tubes of the DNeasy® PowerSoil® Pro DNA extraction Kit (Qiagen, MD, USA). DNA extraction was performed following the manufacturer's instructions, a 1600 MiniG® Automated Cell Lyser (SPEX Sample Prep, NJ, USA) was used to homogenize samples thoroughly by performing three repetitions of cycles of three minutes at 1500rpm and pausing one minute. Library preparation was carried out using the Illumina 16S Metagenomic Sequencing Library protocol with some modifications. A dual indexing protocol with a two-step PCR was performed to amplify the V5 through V6 region of the bacterial 16S rRNA gene and to add dual TruSeq indices. Primer for 16S sequences (underlined portion) with TruSeq adaptor tails are shown: primer 799F (5' ACACTCTTTCCCTACACGACGCTCTTCCGATCT

AACMGGATTAGATACCCKG 3') was chosen to exclude chloroplast amplification (91, 92), and primer 1114R (5' GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT GGGTTGCGCTCGTTGC 3') includes three mismatches with plant mitochondrial DNA (117).

The primary amplification of the 16S region was performed in 25-µl PCR reactions with 3.0 µl of genomic DNA (~ 10ng), 1X of 5X KAPA HiFi Buffer, 0.3 mM of dNTPS, 0.5 U of KAPA HiFi polymerase (KAPA Biosystems, Woburn, MA) and 0.2 µM of each primer 799F/1114R. Cycling conditions were as follows: 95°C for 5 min, followed by 30 cycles of 98°C for 20 s, 55°C for 30 s, and 72°C for 45 s, with a final extension of 72°C for 5 min. The approximately 415-bp PCR products were purified using Agencourt AMPure XP (Beckman Coulter, Inc., CA, USA) at a ratio of 0.7x to remove any product shorter than 200-bp.

The secondary amplification to add TruSeq dual indices and sequencing adaptors was performed in 25-µl PCR reactions with 3.0 µl of purified first PCR product, 1X of 5X KAPA HiFi Buffer, 0.3 mM of dNTPS, 0.5 U of KAPA HiFi polymerase (KAPA Biosystems, Woburn, MA) and 0.3 µM of each primer i5 and i7. Cycling conditions were as follows: 95°C for 5 min, followed by 10 cycles of 98°C for 20 s, 55°C for 30 s, and 72°C for 45 s, with a final extension of 72°C for 5 min. These second PCR products were also purified as described above.

Each sample was amplified three times and every replicate was prepared as a separate library to be sequenced. All 108 libraries were quantified through qPCR assays using primer 1.1 (5' AATGATACGGCGACCACCGAGAT 3') and primer 2.1 (5' CAAGCAGAAGACGGCATACGA 3'), which is the method reported with more accurate prediction of sequencing coverage (118). A 10,000-fold dilution of the purified second PCR products was prepared with 10 mM Tris-HCl pH 8.0 to achieve concentrations inside the limits of the standard curve built with the KAPA Library Quantification DNA Standards 1 – 6 (KAPA Biosystems, Woburn, MA). Quantification was performed in 10-µl qPCR reactions with 5 µl of PowerUp™ SYBR™ Green Master Mix 2X, 2 µl of diluted second PCR products, and 0.2 µM of each primer 1.1/2.1. Cycling conditions were as follows: 50°C for 2 min, 95°C for 2 min, followed by 40 cycles of 95°C for 15 s, 55°C for 15 s, and 72°C for 45 s. All libraries were multiplexed into a single pool in molar equivalent concentrations. The pool was sequenced and demultiplexed at Genewiz (South Plainfield, NJ, USA) using Illumina MiSeq paired-end 250-bp sequencing.

4.3.14 Sequence data processing and analysis

Data processing and analysis was performed using packages in R (v 3.6.1) and Bioconductor (v 3.10). Reads were processed through a customized DADA2 (v 1.14.1) pipeline (95) for trimming the primer sequences, filtering based on read quality, calculating error rates, inferring amplicon sequence variants (ASVs) instead of using OTUs (119), merging paired-end reads, removing chimeras and assigning taxonomy with the SILVA reference database (v 138) (120). Contaminant sequences which were present in the buffer extraction controls (BEC) and non-template controls (NTC) were removed with the decontam package (120). Also, reads with unclassified Phylum or identified as Chloroplast or Mitochondria were filtered out, as well as reads from the inoculated *E. coli* O157:H7 strain. The α-diversity indices (richness and Shannon) were performed with Phyloseq (v 1.32.0) package after 100 subsampling iterations to the smallest library size and averaging the results (96). The β-diversity unconstrained analysis was performed on weighted Unifrac distances applied to reads proportionally scaled to 1,000-bp (code adapted from tutorial: http://deneflab.github.io/MicrobeMiseq) and were then visualized using PCoA. Experimental conditions were compared by analysis of variance using a permutation test with pseudoF ratios with the adonis() function (R vegan package). Normalization and differential abundance analysis were performed on unrarefied ASV table using DESeq2 (v 1.28.1) (121), only significant taxa ($P<0.05$ after multiple-hypothesis testing) were reported. Plots were generated with the ggplot2 (v 3.3.2) package and minor aesthetic details were arranged in Inkscape (v 0.92.4).

4.4 Results and Discussion

4.4.1 Comparison of growth of *E. coli* O157:H7 on leaves of romaine lettuce plants grown in greenhouse and growth chamber conditions

For this experiment, all inoculations of *E. coli* O157:H7 were performed under humid growth chamber conditions (Condition A) to provide the most optimal environment for bacterial growth and evaluate the effect of the environmental conditions in which the lettuce plants were grown. Plants grown in greenhouse conditions (Temp: 22-24˚C, RH: 20-50%) were transferred to growth chamber condition A and allowed an acclimatization period of 24 hours before performing the inoculations. Eighteen inoculations were carried out using greenhouse-grown plants and twenty-four inoculations using growth chamber condition A-grown plants. As observed in Figure 4.3, romaine lettuce plants grown in greenhouse did not support growth of *E. coli* O157:H7; even

though they were inoculated under humid conditions, all inoculations resulted in the decline of the bacterial pathogen's population as follows: Growth change mean ± SD 16h: -1.16 ± 0.62 logCFU, 24h: -1.67 ± 0.57 logCFU, 48h: -1.49 ± 0.71 logCFU. On the other hand, plants whose entire growth cycle happened under humid conditions were more susceptible to growth of *E. coli* O157:H7. Although some inoculations also resulted in decline, the average growth showed that the pathogen populations were doubling in number as follows: Growth change mean ± SD 16h: 0.25 ± 0.59 logCFU, 24h: 0.30 ± 0.36 logCFU, 48h: 0.28 ± 0.36 logCFU.

This finding suggests that growth conditions prior to the time of inoculation can influence the growth and survival of *E. coli* O157:H7 on romaine lettuce, in addition to the environmental conditions present at the time of inoculation as previously demonstrated in a field study conducted in the Salinas Valley, CA. Those field trials were inoculated with a high inoculum dose (7 log CFU/plant) and no growth of *E. coli* O157:H7 was observed, however they found significantly lower declines of the pathogen's population when inoculations happened at night when relative humidity and leaf wetness were higher compared with when inoculations happened in the morning (108). Our results highlight that the environmental conditions in which the romaine lettuce plants are grown can also influence their susceptibility to allow growth of *E. coli* O157:H7. For that reason, our next experiments evaluated the growth and survival of this pathogen on plants grown and inoculated under different RH conditions.

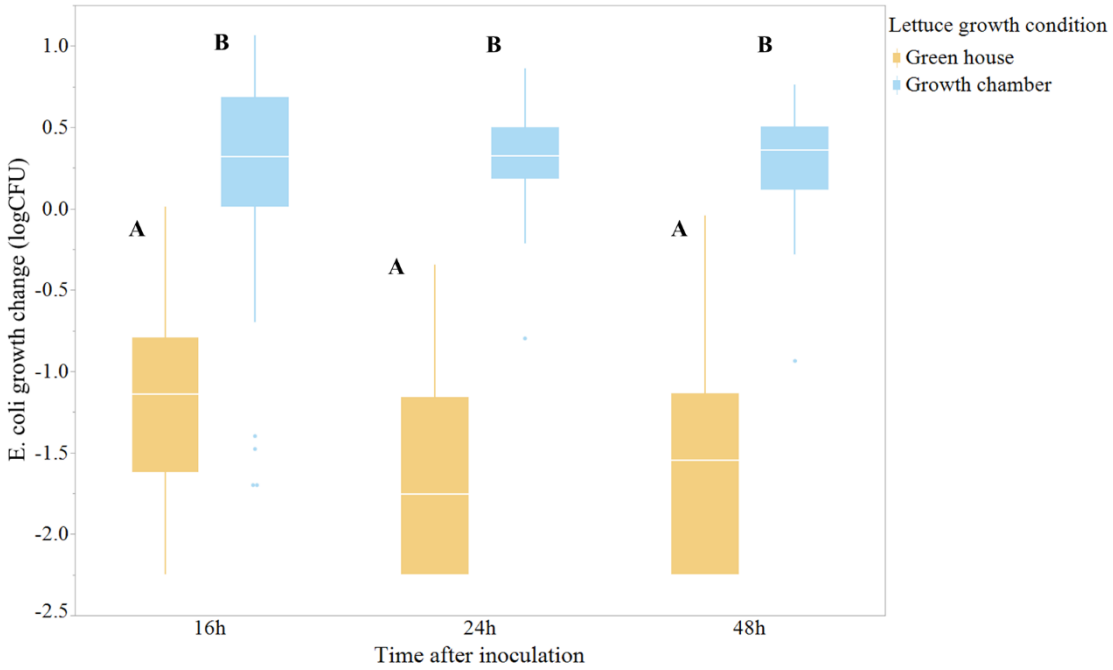


Figure 4.3. Comparison of *Escherichia coli* O157:H7 growth change after 16, 24 and 48 hours of inoculation with 10^4 CFU. Leaves of romaine lettuce plants were inoculated under humid growth chamber condition A using plants grown in greenhouse and plants grown in growth chamber condition A. Different letters within each time after inoculation indicate significant difference ($P < 0.05$).

4.4.2 Growth and survival of *E. coli* O157:H7 on leaves of romaine lettuce plants grown under different relative humidity conditions

Romaine lettuce plants were germinated from seed, grown until they were 6-8 weeks old and drop-inoculated with *E. coli* O157:H7 under the same relative humidity condition in growth chamber. Three RH conditions, A (high), B (intermediate) and C (low), were used to evaluate the fate of the human pathogen on lettuce leaves. Plants grown and inoculated under humid condition A exhibited growth of *E. coli* O157:H7 on their leaves, low inoculum of this pathogen was able to increase its population in average of 0.80 log CFU at 16 hours and 0.41 log CFU at 112 hours. On the other hand, 16 hours after infection with low inocula, intermediate RH condition B only allowed the preservation of the same inoculated population number with no significant growth, while low RH condition C caused a decline in average of 1.13 log CFU (Table 4.2). Additionally, under circumstances where *E. coli* O157:H7 exhibited decline in its population such as 112 hours or high inoculum, its survival was always greater in the humid condition A (Figure 4.4). It is also noticeable that at 16 hours, intermediate condition B and low condition C showed significantly

different *E. coli* O157:H7 log CFU, demonstrating a faster decrease rate at the lowest RH. However, at 112 hours this rate becomes equivalent resulting in similar remaining log CFU (Table 4.2).

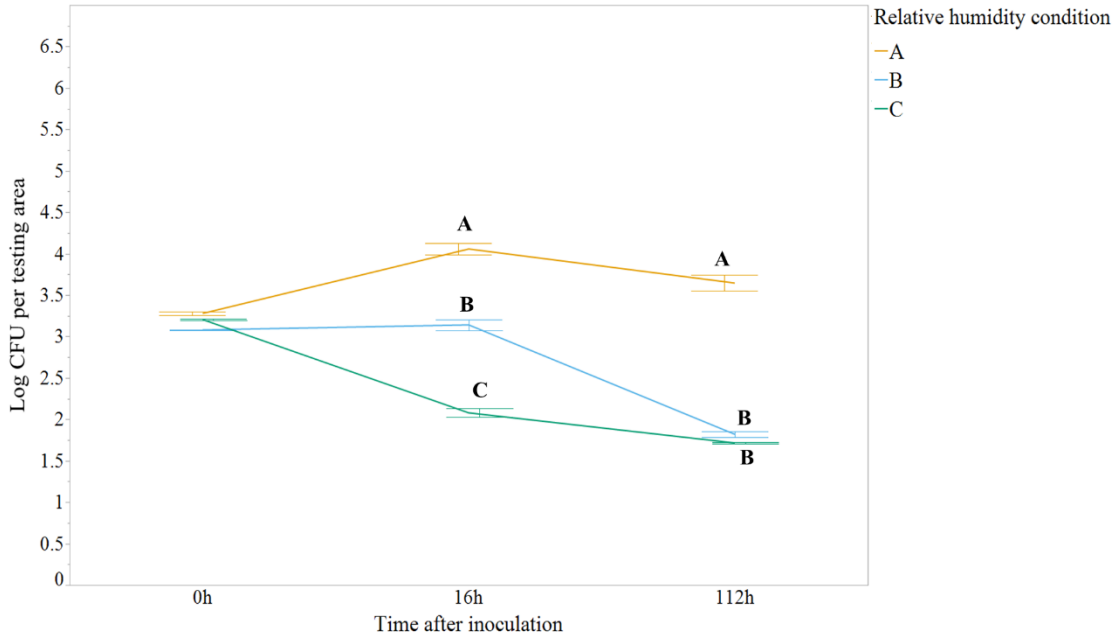
Table 4.2. *Escherichia coli* O157:H7 populations (log CFU) after inoculation on leaves of romaine lettuce plants germinated, grown and inoculated under high RH (condition A), intermediate RH (condition B) and low RH (condition C)

RH condition-Time after inoculation	Low inoculum			High inoculum		
	N	*E. coli* O157:H7 (log CFU)[a]	p-value	N	*E. coli* O157:H7 (log CFU)[a]	p-value
A- 0h	48	3.28 ± 0.13		48	6.25 ± 0.11	
B- 0h	48	3.08 ± 0.04	NA[b]	48	6.08 ± 0.04	NA[b]
C- 0h	36	3.20 ± 0.06		36	6.20 ± 0.06	
A- 16h	96	4.06 ± 0.65 (A)		144	6.36 ± 0.72 (A)	
B- 16h	144	3.14 ± 0.81 (B)	<0.0001	144	5.56 ± 0.90 (B)	<0.0001
C- 16h	144	2.08 ± 0.64 (C)		144	4.78 ± 1.24 (C)	
A- 112h	120	3.65 ± 1.03 (A)		144	4.86 ± 1.23 (A)	
B- 112h	144	1.82 ± 0.43 (B)	<0.0001	144	2.81 ± 0.91 (B)	<0.0001
C- 112h	144	1.72 ± 0.14 (B)		144	2.56 ± 0.79 (B)	

[a] Values are means ± standard deviations of log(CFU). Values within same time after inoculation followed by a different letter are significantly different ($P < 0.05$). [b] Not applicable.

Survival in the phyllosphere is highly influenced by the surrounding environmental conditions since this is an exposed habitat. Seasonal differences have been observed in the fate of *E. coli* O157:H7 on field-grown lettuce, counts of this pathogen were higher in fall than in spring when irrigated with contaminated water containing 10^7 CFU (158). The same phenomenon was observed when leafy greens in the mid-atlantic region of US were surveyed for the presence of *Salmonella enterica*, there was a higher recovery of this pathogen in fall than in spring (159). Those research results added to the fact that 74% of the *E. coli* O157:H7 outbreaks linked to lettuce consumption have occurred during summer and early fall seasons (21), highlight the important role of the environmental conditions on modulating the growth of *E. coli* O157:H7 on plants.

A.



B.

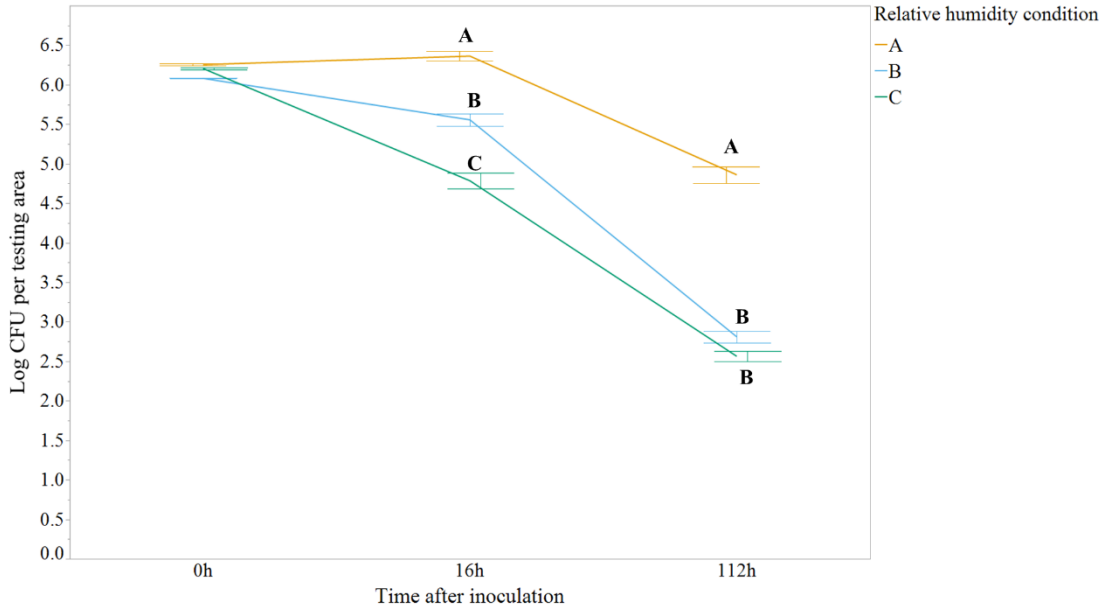


Figure 4.4 Growth and survival of *Escherichia coli* O157:H7 populations after drop-inoculation of low inoculum dose (A) and high inoculum dose (B). Inoculations were performed on leaves of romaine lettuce plants germinated, grown and inoculated under high RH (condition A, yellow line), intermediate RH (condition B, blue line) and low RH (condition C, green line). Each time point represents the mean *E. coli* O157:H7 log CFU with SE bars.

Under field conditions many environmental factors are playing a role at the same time, such as temperature, relative humidity, wind speed, rainfall, solar radiation, among others. Additionally, these factors are constantly fluctuating which makes it difficult to determine the effect of each factor on the growth and survival of *E. coli* O157:H7 on lettuce leaves. For that reason, although experiments under controlled environmental conditions might not entirely represent what is happening in the field, they are very useful to study each factor individually. This type of research leads to an understanding of which environmental factors have a stronger effect on the plant, on the pathogen and on the native microbial communities which are the main living components of this food safety issue. Our results indicate that with a constant temperature and light intensity, the growth and survival of *E. coli* O157:H7 on lettuce leaves is highly affected by the relative humidity in which the lettuce plants are grown and inoculated.

The greatest contrast of *E. coli* O157:H7 populations on lettuce leaves in our study was observed at 16 hours when actual growth was happening at the more humid condition A, while a die-off rate of more than 1 log CFU was happening at the driest condition C. This research demonstrates that growth of this pathogen on intact leaves of romaine lettuce plants is possible when a low inoculum dose and less adverse relative humidity conditions are provided. Additionally, 100% and 77.8% of *E. coli* O157:H7 log CFU (when low and high inoculum doses were used, respectively) survived after 112 hours of inoculation in humid condition A, while only around 45% of *E. coli* O157:H7 log CFU survived at this time in intermediate and dry conditions (B and C). Similar results have been reported for survival of *Salmonella enterica* subsp. *enterica* when iceberg lettuce plants were inoculated under humid conditions (85-90% RH) compared with dry conditions (45-48% RH); however, this study did not find the same significant difference for *E. coli* O157:H7, maybe because their plants were only inoculated and not grown under these contrasting conditions (160).

Most of the studies on field-grown lettuce plants did not achieve growth of *E. coli* O157:H7 on their leaves, instead die-off rates ranging between 0.4 to 1.64 log MPN per day have been reported (101). Moyne *et al.* also obtained a reduction by 3 to 4 log CFU/leaf after one day of performing inoculations in growth chamber under harsh environmental conditions such as low RH of 30% and moderate temperature of 22˚C (98). Our study complements these previous efforts that low RH does not support growth of *E. coli* O157:H7 on lettuce leaves, but also demonstrates how the same lettuce cultivars, grown under the same temperature regime might become susceptible

and support increases in the pathogen population when only the relative humidity condition is modified. Clearly, water activity of substrates is critical for growth and survival of bacterial pathogens. Previous reports show that the required values for *S. aureus, L. monocytogenes, B. cereus, E. coli* O157:H7, and *Salmonella Typhimurium* are 0.86, 0.92, 0.93, 0.95, and 0.95, respectively (161).

4.4.3 Effect of inoculum level on growth and survival of *E. coli* O157:H7 on leaves of romaine lettuce plants

Two inoculum levels of *E. coli* O157:H7 were used for the drop-inoculations of testing areas of romaine lettuce leaves: low level (ca. 3 log CFU) and high level (ca. 6 log CFU). As observed in Figure 4.5, the inoculum level applied to plants caused a significant difference in the growth or die-off rates of *E. coli* O157:H7. Lettuce plants inoculated with a low level of pathogen CFU exhibited a higher growth change leading to increase of populations at 16 hours in both environmental conditions A and B (0.80 ± 0.59 and 0.06 ± 0.81, respectively [mean log CFU ± SD]). On the other hand, high inoculum level caused a decline of the populations at 16 and 112 hours under all environmental conditions (16h-B: -0.52 ± 0.90, 16h-C: -1.42 ± 1.24, 112h-A: -1.37 ± 1.23, 112h-B: -3.27 ± 0.90 , 112h-C: -3.64 ± 0.80). The only exception was at 16 hours under humid condition A where a small increase in population was observed 0.12 ± 0.73, which was still significantly lower than the increase observed (0.80 ± 0.59 log CFU) when a low inoculum level was applied under the same conditions.

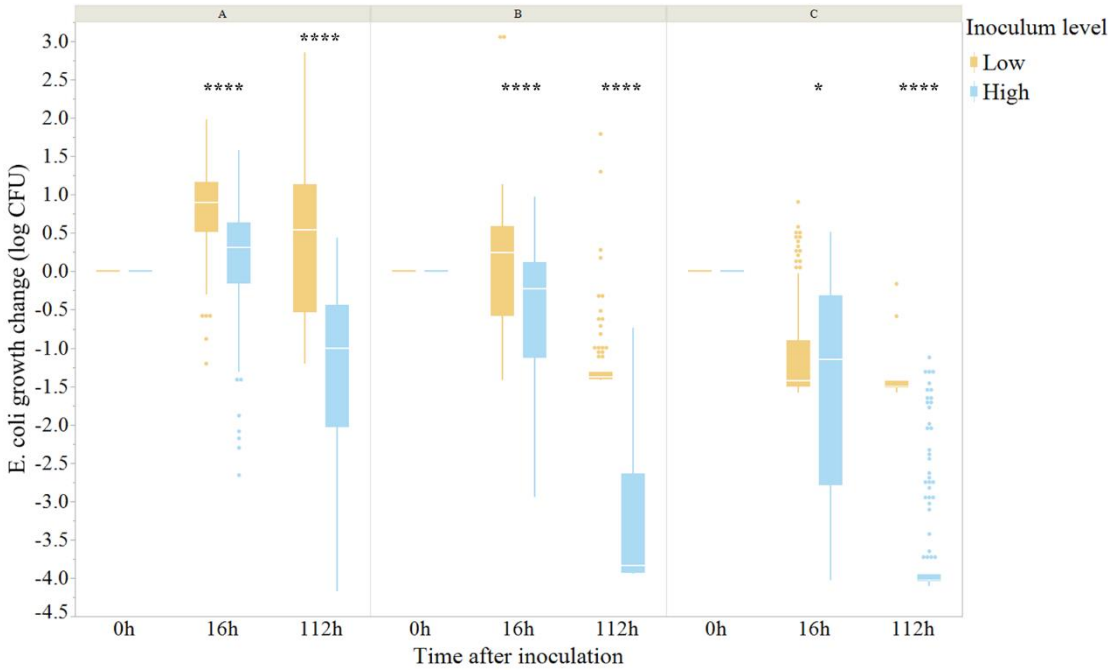


Figure 4.5. Growth and die-off rates of *Escherichia coli* O157:H7 when was inoculated using low level (ca. 3 log CFU) and high level (ca. 6 log CFU) of inoculum. Growth changes were calculated 16 and 112 hours after inoculation, under relative humidity conditions A, B and C. Within each time after inoculation, significant differences caused by inoculum level are *<0.05, ****<0.0001.

Although only low inoculum levels of *E. coli* O157:H7 were able to grow and increase their populations on lettuce leaves, the risk of foodborne disease determined by the number of CFU is significant for both levels of inoculum. After 16 and 112 hours post inoculation, both inoculum levels allowed the survival of more than 10 CFU (Figure 4.6), which has been defined as the infectious dose of this pathogen in humans (162). Moreover, a higher percent (53-59%) of *E. coli* O157:H7 log CFU were recovered after 112 hours when low inoculum levels were used, compared with only 41-46% when high inoculum levels were used. Despite that higher survival rate for low inoculum levels, their total number of CFU at 112 hours is getting very low (1.82 and 1.72 log CFU) under intermediate condition B and dry condition C, respectively. This finding suggests that times after 112 hours should be tested to determine if even though low inoculum doses exhibit higher growth rates, they also persist for shorter time periods on lettuce plants when compared with high inoculum doses.

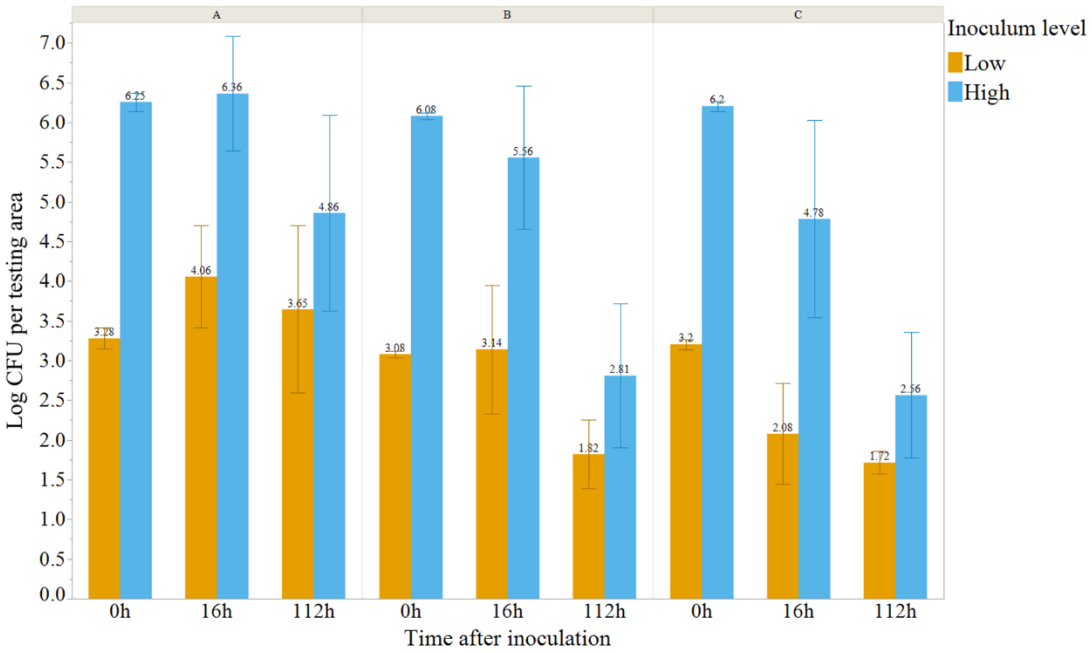


Figure 4.6. *Escherichia coli* O157:H7 populations (log CFU) on lettuce leaves after 16 and 112 hours post inoculation with low level (ca. 3 log CFU) and high level (ca. 6 log CFU) of inoculum, under relative humidity conditions A, B and C. Mean with error bars corresponding to 1 SD.

Correlation of die-off rates increasing when inoculum dose is higher was also reported by McKellar *et al.* (163), who performed data analysis and modeling using several data sets from different studies of survival of *E. coli* O157:H7 in field lettuce (102, 103, 164). This phenomenon could also explain why all studies performed under field conditions only found a decline of the pathogen's population. This type of study usually employs irrigation as contamination method, therefore high inoculum doses are used to spread enough bacteria on plants and later be able to detect them as high volumes of buffer are needed to prepare samples from entire plants. In contrast, our research study employed a more localized approach using testing areas across lettuce leaves, in that way we were able to evaluate very low inoculum doses which are more likely to occur in natural events of contamination (101). Therefore, comparing our results with previous reports, studies using high inoculum doses might be overestimating the die-off rates of *E. coli* O157:H7 on lettuce plants, and overlooking the fact that this pathogen can grow and increase its population when conducive environmental conditions are present. Also, the use of high inoculum doses might be overestimating the time that this pathogen can persist on lettuce plants, since high numbers of

E. coli O157:H7 cells, although are not increasing in number, may take more time to be completely undetectable compared with low numbers of cells.

4.4.4 Comparison of growth of *E. coli* O157:H7 across the leaf area of romaine lettuce plants

The inoculation approach applied in this study allowed us to compare the growth of *E. coli* O157:H7 on twelve different testing areas (Figure 4.2) across each lettuce leaf. Comparisons of the growth and die-off rates were performed using different arrangements across the leaf surface. Growth and die-off rates were compared between three vertical zones (right, middle and left areas), four horizontal zones (by grouping testing areas as follows: [A, B, C], [D, E, F], [G, H, I], [J, K, L]), as well as the upper and lower halves of the leaf (by grouping testing areas as follows: upper [A-F] and lower [G-L]). Significant differences in the fate of *E. coli* O157:H7 were found between the upper and the lower halves of the leaves. This was observed under conditions where the pathogen's population was increasing or declining less than 2.0 log CFU, such as Condition A at 16 and 112 hours, and 16 hours of Condition B and C (Figure 4.7). On the other hand, after 112 hours under conditions B and C, when the *E. coli* O157:H7 population was very reduced with a decrease higher than 2.0 log CFU, the significant difference between the leaf horizontal halves was lost. This finding suggests that some leaf surface properties could be differentially developed between upper and lower halves and might affect the successful growth and initial decline of *E. coli* O157:H7 populations.

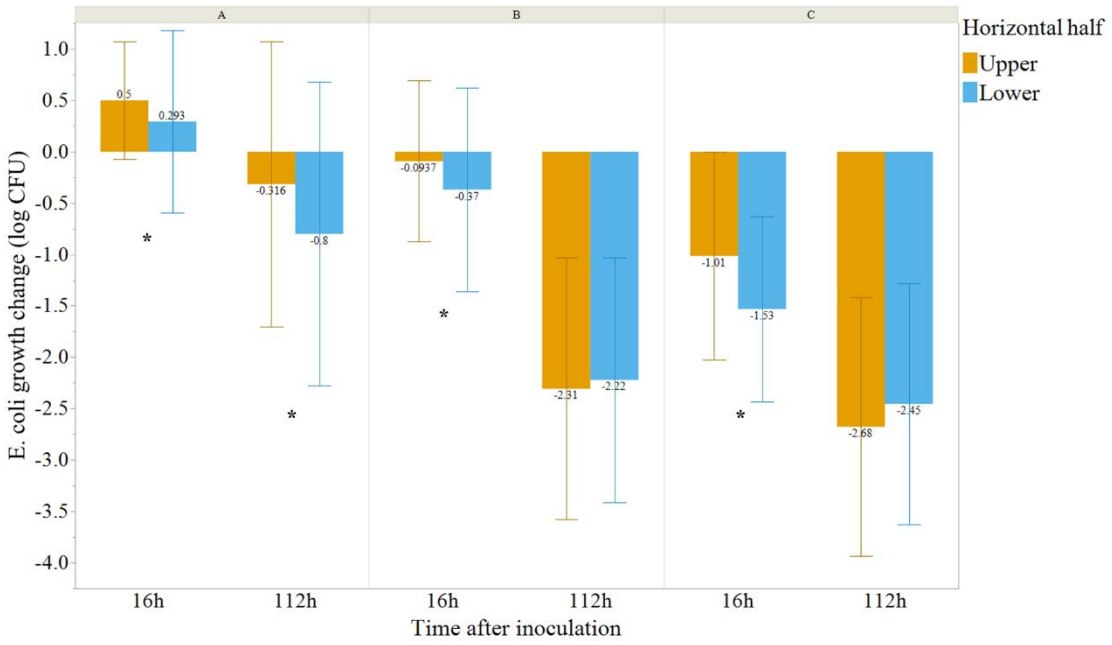


Figure 4.7. Comparison of growth and decline of *Escherichia coli* O157:H7 populations across the lettuce leaf surface of upper and lower halves of leaves. Growth changes were calculated 16 and 112 hours after inoculation, under relative humidity conditions A, B and C. Within each time after inoculation, significant differences caused by horizontal half are *<0.05. Mean with error bars corresponding to 1 SD.

In previous studies, exposed bacterial populations on outer leaves were demonstrated to decline more rapidly than protected populations on inner or shaded leaves (54, 103). However, this is countered by the fact that outer leaves get higher contamination when plants are exposed to contaminated sources such as irrigation water (158, 165). Previous observations of higher susceptibility of exposed populations to decline was not found in this study; on the contrary, the most exposed area of the leaves (upper half) exhibited higher *E. coli* O157:H7 growth and lower initial decline of its population. Few studies have approached the relationship between the leaf area localization and the growth of foodborne bacterial pathogens. For *Salmonella typhimurium* maximal attachment was observed on surfaces localized near the petiole. The leaf topography examined by scanning electron microscopy (SEM) was found to be most complex near the petiole (141). However, our sample preparation method involved sonication to also release attached bacterial cells. Additionally, in this study inoculation with *Salmonella* was performed by dipping cut leaf pieces from lettuce heads bought from grocery stores into the bacterial inoculum, these post-harvest conditions are not comparable with our pre-harvest conditions and drop-inoculations.

Therefore, other factors such as chemical constituents, trichome or stomata density, thickness of the leaf, nutrient level or enhanced wettability should be analyzed to identify those differences in pre-harvest contamination with *E. coli* O157:H7 between upper and lower halves of the lettuce leaves.

To evaluate some of those differences between upper and lower half of the leaves, measurements of culturable bacterial populations, contact angle hysteresis, cuticular wax content, and stomatal density were determined from plants grown under relative humidity conditions A, B and C. As observed in Table 4.3, *E. coli* O157:H7 populations after 16 hours of inoculation were significantly different between upper and lower halves of the leaves in all RH conditions. As previously mentioned, the localization of the pathogen on the leaf surface seems to play a protective role by slowing down the start of the population's decline. To determine if the localization on the leaf also affected the resident culturable bacteria, total plate counts were performed, and no significant difference was found between upper and lower halves of the leaves of plants grown under the same RH condition. However, it is noticeable how only the resident bacterial populations of the lower half of the leaves were significantly affected by the RH condition, exhibiting an increase in their numbers when RH decreased, while the counts of resident bacteria of the upper half did not. This finding suggests that the resident bacteria might find an alleviation from the harsh environmental conditions in the lower leaf area, which promotes a higher growth; while the resident bacteria still present in the upper leaf area might be striving to survive so their population size remains invariant.

The effect of the lower RH is evidenced by the significant difference in the wettability of the leaf surfaces between the humid condition A and the dryer conditions B and C. The larger the contact angle hysteresis value, the more the leaf surface tends to repel water, which means that it is becoming more hydrophobic. The upper half area of the leaves always exhibited larger contact angle hysteresis values than the lower half area under all RH conditions where the plants were grown. This significant difference denotes the higher exposure to environmental conditions that the upper half of lettuce leaves have. Although this effect is also observed in leaves of plants grown under humid condition A, both the upper and lower half areas presented contact angles significantly lower than those from plants grown under dryer conditions B and C. Reduction of leaf contact angle values under humid conditions has been reported, for instance a study of 68 plant species among trees, shrubs and herbs showed this kind of surface wettability increase from

the dry period (May to June) to the rainy period (July to August) in China (166). Leaves of plants growing in dry, open or alpine habitats exhibited more water repellency than those in rainy, shady, or tropical habitats (167, 168). This phenomenon has been explained as an important strategy to decrease rainfall interception and increase throughfall to the soil in arid and semi-arid environments (169).

As wettability is generally attributed to the chemical composition and the geometrical microstructures of the solid surfaces (123), cuticular wax content was measured from the upper and lower areas of the leaves. The identified compounds were separated into the three more abundant classes of components of plant cuticles: alkanes, fatty acids and primary alcohols (see Table S1). As observed in Table 4.3, no significant differences were found between the upper and lower halves of leaves when content of alkanes was evaluated across all RH conditions. On the contrary, amounts of fatty acids and alcohols were significantly higher in the lower leaf area of plants grown under the extreme environmental conditions of low RH (condition C) and high RH (condition A). However, leaves of plants grown under medium RH (condition B) did not develop differences across the surface area in terms of their cuticular wax content, culturable bacterial populations, and stomatal density. Previous studies have found that the contact angle increased with an increase in the leaf wax content (124, 132, 170, 171), however the correlation was not always significant and the contact angle did not decrease after wax removal using acetone for a number of plant species (170). Therefore, as our results did not demonstrate a positive relationship between contact angle hysteresis and cuticular wax content, we also support the previously reported suggestion that the leaf contact angle is more dependent on the complexity of wax structure, surface roughness, and surface heterogeneity rather than on absolute amount (126). Further studies using scanning electronic microscopy (SEM) might reveal more differences between the upper and lower halves of romaine lettuce leaves which can be associated with the significantly different growth of *E. coli* O157:H7 obtained here.

Stomata are known as preferential sites for bacterial attachment on leaf surfaces (172), as well as routes of internalization (45). Stomatal density was significantly higher in the upper half of leaves from plants grown under the extreme environmental conditions of low RH (condition C) and high RH (condition A). These conditions also resulted in significantly higher *E. coli* O157:H7 population counts, contact angle values, and lower contents of fatty acids and alcohols in the cuticular waxes on the upper half of the leaves. However, it is noticeable that plants grown under

medium RH (condition B) did not exhibit significant differences across the leaf surface in their stomatal density or cuticular wax content, but still showed higher contact angles and *E. coli* O157:H7 population counts in the upper halves of leaves. These findings suggest that the most extreme relative humidity conditions caused changes in the leaf properties, and these changes occurred more drastically in the leaf surface area which is more exposed. Also, within each of our three experimental RH conditions, decreased leaf wettability (evidenced as higher values of contact angle hysteresis) was the leaf property which was most strongly associated with slowing down the decline of *E. coli* O157:H7 counts.

Table 4.3. *Escherichia coli* O157:H7 counts after 16 hours of inoculation with 6 log CFU, culturable bacterial counts, and leaf surface characteristics from upper and lower halves of leaves of plants grown at different RH conditions. ANOVA tests between leaf halves within each RH condition A, B and C.

Condition RH-Leaf half	*E. coli* O157:H7 (Log CFU)[a]	Culturable bacteria (Log CFU)	Contact Angle Hysteresis (°)[a]	Cuticular wax content µg/leaf g (%) Alkanes-FAs-Alcohols			Stomatal density (stomata /mm^2)[a]
A-UpperHalf	6.56 A	2.83	46.0 B	3.0 (1.8)	74.6 (47.8)	78.5 (50.4)	48.6
A-LowerHalf	6.17 Z	2.56 Y	39.0 Y	6.2 (2.7)	116.6 (49.7)	112.4 (47.7)	38.7 Z
B-UpperHalf	5.83 B	2.85	60.2 A	1.8 (0.99)	92.9 (47.2)	103.2 (51.8)	51.6
B-LowerHalf	5.28 Y	2.77 Y	53.4 Z	2.5 (0.99)	101.8 (44.3)	126.5 (54.7)	50.4 X
C-UpperHalf	5.14 C	3.05	53.9 A	10.3 (8.6)	61.1 (55.0)	42.0 (36.5)	50.2
C-LowerHalf	4.43 X	3.30 Z	46.5 Z	12.6 (6.1)	86.8 (51.6)	70.7 (42.3)	42.5 Y
No. of plants	12	20	20		20		6
ANOVA (leaf half by RH)	<0.01 <0.001 <0.001	NS[b]	<0.05 <0.05 <0.01	NS[b]	<0.001 NS[b] <0.05	<0.01 NS[b] <0.01	<0.0001 NS[b] <0.0001

[a] Different letter indicate significant difference ($P < 0.05$) caused by RH condition between samples of upper half (A-C) and lower half (Z-X). [b] Not significant.

Overall, the decrease of relative humidity caused *E. coli* O157:H7 populations to decrease in both the upper and lower halves of leaves. However, we found that the lower wettability in the upper area of leaves might be preventing the droplets from spreading so they retain their shape, and therefore dry more slowly (173). This phenomenon might be increasing the time the inoculum drops stay on the leaves, and the prolonged presence of water may increase the microbial metabolic activity and growth, as well as increasing the rate of leaching of nutrients by diffusion across the leaf cuticle (174). These reasons may account for why the populations of *E. coli* O157:H7 were able to grow more (in the case of low inoculum dose) or decline more slowly (in the case of high inoculum dose) in the upper leaf areas.

4.4.5 Sequencing of resident bacterial communities from leaf testing areas where *E. coli* O157:H7 has been inoculated

In previously described experiments, plants grown under three environmental conditions of high RH (condition A), medium RH (condition B), and low RH (condition C) were inoculated with *E. coli* O157:H7 on twelve testing areas per leaf. A total of 576 inoculations were performed for each RH condition, and four schemes combining low and high inoculum doses and two testing times after inoculation were evaluated: 3 log CFU-16 hours, 6 log CFU-16 hours, 3 log CFU-112 hours, and 6 log CFU-112 hours. From those results, leaf testing areas exhibiting outlier growth of *E. coli* O157:H7 (very high or very low) were selected for sequencing of the resident bacterial communities (Table 4.4). Thirty-two samples inoculated with the pathogen were sequenced, as well as three control samples which were pools of 10 non-inoculated plants (5 GT and 5 TH) from each of the three environmental conditions A, B and C.

The V5-V7 region of the 16S rRNA was sequenced by paired-end 250-bp MiSeq Illumina sequencing, yielding contigs with length sizes of 298 ± 5.6 bp. After quality filtering and removal of chimeric, non-target (archaea, chloroplast and mitochondria), and contaminant sequences (from BEC and NTC), 1.5 million reads corresponding to 9,163 ASVs were obtained. Low abundance ASVs (fewer than 2 reads in less than 10% of the samples) were removed and 941 significant ASVs were left. After the reads coming from the inoculated *E. coli* O157:H7 strain were filtered out from all samples, experimental conditions which yielded high growth (Hi) of the pathogen resulted in lower library sizes than conditions where low growth (Lo) was obtained (Figure 4.8). Conditions where plants were inoculated with 6 log CFU and tested after 16 hours showed the

lower number of reads. We found that the percent of *E. coli* O157:H7 reads was significantly correlated (P<0.0001) with the inoculum used (R=0.73) and the *E. coli* O157:H7 population counts obtained at the time of testing after inoculation (R=0.64) (Table 4.4). However, the final number of reads sequenced from the resident bacterial populations was correlated only with the *E. coli* O157:H7 log CFU obtained (P=0.0024, R=-0.52) and not with the initial inoculum (P=0.1). This finding suggests that the sequencing of resident bacterial populations is not affected by the methods of our research, indeed it is affected by the fate of *E. coli* O157:H7 on lettuce leaves. Finally, samples #5 and #21 were removed after we applied a minimum threshold of 400 reads for library size, and the final data set for analysis consisted of 1.2 million reads, 941 ASVs, an average of 13,285 and a median of 3,359 reads per sample.

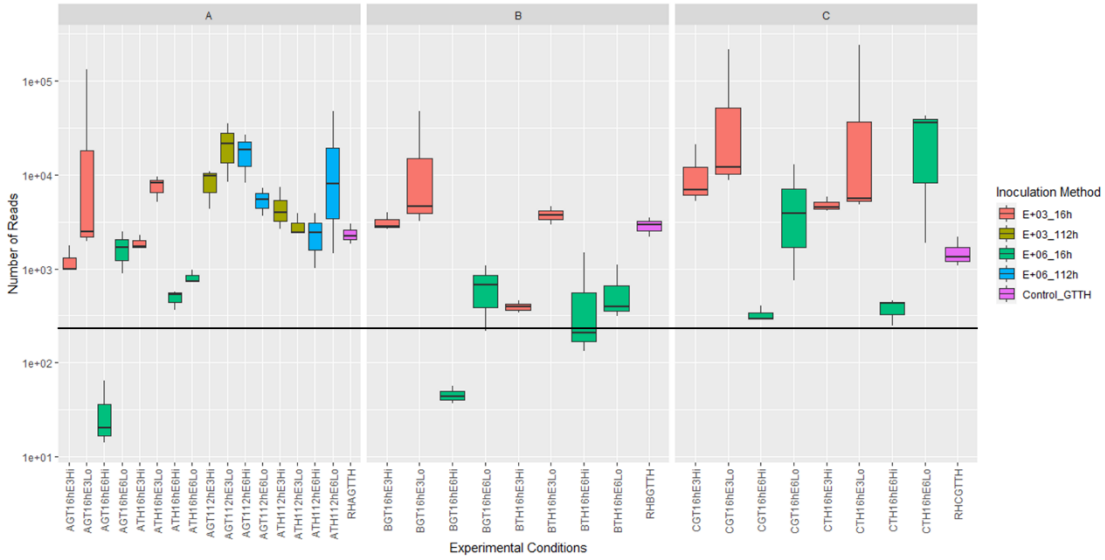


Figure 4.8. Library sizes of the 32 sequenced samples from leaf testing areas that exhibited outlier growth of *Escherichia coli* O157:H7. Final numbers of reads after sequences of inoculated *E. coli* O157:H7 were removed. Panels correspond to relative humidity conditions A, B, and C. Inoculation method describes the inoculum dose and the time of testing after inoculation. Control_GTTH refers to the pool of non-inoculated plants for each humidity condition. Black line indicates the threshold of minimum library size applied.

Table 4.4. Samples for sequencing of leaf testing areas that exhibited outlier growth of *E. coli* O157:H7. High and low outliers were selected from 16 experimental conditions of lettuce cultivar, testing time after inoculation, and inoculum dose. Average of population counts (log CFU), growth change (log CFU) and percent of reads of *E. coli* O157:H7 for each outlier sample is presented. Also, the average growth change obtained for all inoculations performed under each experimental condition is given for comparison with outliers.

Sample	RH	Lettuce cultivar [a]	Time after inoculation	Inoculum (Log CFU)	*E. coli* outlier growth	*E. coli* (Log CFU)	*E. coli* growth change of outliers[b]	*E. coli* growth change of all inoculations[b]	% *E. coli* reads
1	A	GT	16h	3.00	High	4.42	1.72	0.61	80.3
2	A	GT	16h	3.00	Low	2.69	-0.25		2.2
3	A	TH	16h	3.36	High	4.77	1.40	1.00	70.2
4	A	TH	16h	3.37	Low	3.22	-0.15		22.0
5	A	GT	16h	6.17	High	6.74	0.57	0.11	99.8
6	A	GT	16h	6.13	Low	4.09	-2.04		95.1
7	A	TH	16h	6.32	High	7.28	0.95	0.14	98.6
8	A	TH	16h	6.30	Low	4.68	-1.62		98.6
9	A	GT	112h	3.11	High	5.07	1.96	0.29	69.5
10	A	GT	112h	3.54	Low	2.28	-1.26		9.8
11	A	TH	112h	3.35	High	5.13	1.78	0.54	91.9
12	A	TH	112h	3.56	Low	2.82	-0.74		78.6
13	A	GT	112h	6.16	High	6.32	0.16	-1.23	59.8
14	A	GT	112h	6.18	Low	2.79	-3.39		88.7
15	A	TH	112h	6.33	High	6.57	0.24	-1.51	96.9
16	A	TH	112h	6.30	Low	2.65	-3.66		91.8
17	B	GT	16h	3.05	High	3.87	0.82	0.01	54.3
18	B	GT	16h	3.05	Low	1.80	-1.25		34.2
19	B	TH	16h	3.11	High	5.19	2.08	0.11	98.1
20	B	TH	16h	3.12	Low	2.08	-1.04		19.1
21	B	GT	16h	6.06	High	6.81	0.75	-0.37	99.7
22	B	GT	16h	6.11	Low	3.62	-2.42		98.8
23	B	TH	16h	6.03	High	6.57	0.45	-0.67	99.4
24	B	TH	16h	6.11	Low	3.81	-2.31		98.6

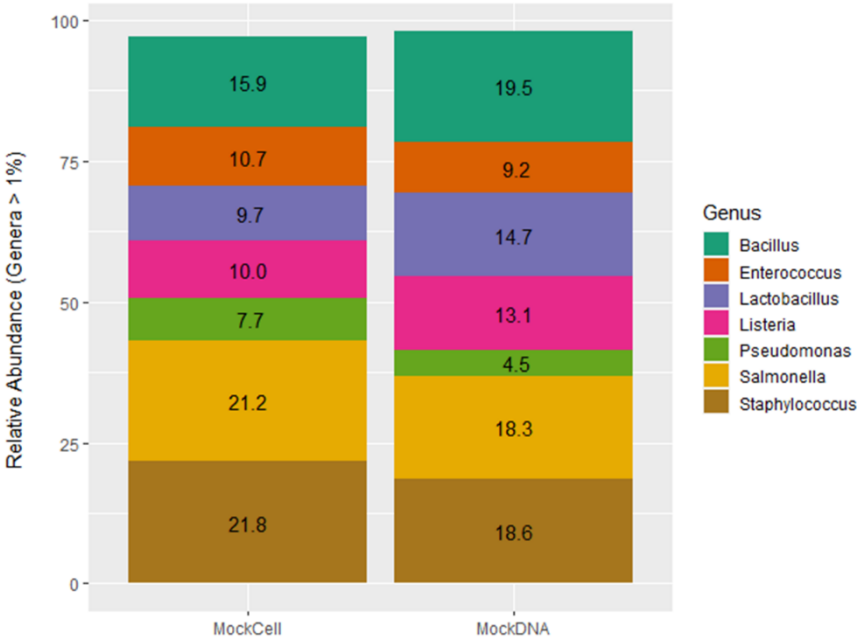
Table 4.4 continued

25	C	GT	16h	3.18	High	3.58	0.40	-1.20	17.3
26	C	GT	16h	3.18	Low	1.75	-1.43		3.0
27	C	TH	16h	3.23	High	3.65	0.42	-1.05	15.2
28	C	TH	16h	3.24	Low	1.75	-1.49		0.4
29	C	GT	16h	6.14	High	6.45	0.30	-1.57	99.0
30	C	GT	16h	6.17	Low	3.18	-3.00		97.4
31	C	TH	16h	6.26	High	6.51	0.26	-1.27	99.0
32	C	TH	16h	6.24	Low	2.88	-3.37		87.4

[a]GT: Green towers, TH: True Heart. [b] *E. coli* growth change is calculated as the logarithmic change between the inoculated CFU and the CFU obtained after the time of testing.

Therefore, it is important to highlight that the small library sizes obtained across all our samples are not a direct effect of the inoculation of *E. coli* O157:H7, instead low bacterial loads inhabiting our plants is suggested because even our controls of non-inoculated plants (pool of 10 entire plants per RH) yielded fewer than 5,000 reads. This suggestion is also supported by the culturable bacterial counts reported above in table 4.3, which were around 3 log CFU. Low numbers of bacterial populations in lettuce plants grown indoors, under laboratory and growth chamber conditions have also been reported before (72, 73, 175), which is mainly attributed to the limited microbial sources (only seed, soil and water) under experimental conditions compared to field conditions (wind, rain, dust, animal and insect contact). Additionally, the quality control performed with mock microbial communities showed a good performance of our methodology. Seven out of eight standard strains were identified in the mock controls and their relative abundances were close to the expected values, with a deviation less than 15% as recommended by the manufacturer (Figure 4.9). The standard strain of *E. coli* included in the mock controls was not reported because it is the same ASV as our strain of *E. coli* O157:H7 and it was removed by our bioinformatics pipeline.

A.



B.

ZymoBIOMICS™ Microbial Community Standard	
Species	**Theorical composition (%) 16S Only**
Pseudomonas aeruginosa	4.2
Escherichia coli	10.1
Salmonella enterica	10.4
Lactobacillus fermentum	18.4
Enterococcus faecalis	9.9
Staphylococcus aureus	15.5
Listeria monocytogenes	14.1
Bacillus subtilis	17.4

Figure 4.9 Microbial composition (relative abundance, %) of mock community standards. A. Our results for the microbial community culture standard (D6300: MockCell) and the microbial community DNA standard (D6305: MockDNA). B. Theorical composition reported by the manufacturer (Relative-abundance deviation in average: <15%).

4.4.6 Structure of romaine lettuce leaves bacterial microbiome after *E. coli* O157:H7 has been inoculated

The taxonomic structure of native bacterial communities from leaves of romaine lettuce plants grown in growth chambers and used for inoculation of *E. coli* O157:H7, combining all the RH and inoculation conditions, consisted of 15 phyla and 39 classes. Most of the resident bacteria belong to 8 classes (Figure 4.10): Actinobacteria, Acidimicrobia and Thermoleophilia from

phylum Actinobacteria, Gamma and Alphaproteobacteria from phylum Proteobacteria, Bacilli and Clostridia from phylum Firmicutes, and Bacteroidia from phylum Bacteroidetes. The average relative composition of both samples inoculated with *E. coli* O157:H7 (IP) and non-inoculated plants (NIP) was similar at the class level: Actinobacteria (47-49%), Bacilli (20%), Thermoleophilia (3.2-4.0%), Acidimicrobia (1.9%), Clostridia (1.1%), Bacteroidia (0.7-1.1%), and the group of remaining classes below 1% (4.5%). However, a change of relative abundance of Alpha and Gammaproteobacteria was observed, for instance Alphaproteobacteria were more abundant (13.8%) than Gammaproteobacteria (6.6%) in non-inoculated plants, while in inoculated plants their relative abundances were 8.9% and 9.3%, respectively. A previous study also found that considering the highly diverse phylum Proteobacteria as one group did not reveal of significant differences resulting from abiotic stresses on plants. However, when relative abundances of Gammaproteobacteria and Alphaproteobacteria were considered separately, the latter were found to be significantly reduced by elevated temperature, while Gammaproteobacteria were increased (176).

The total of 39 bacterial classes comprised 162 families and 313 genera. The average composition at the family level was compared between inoculated samples which exhibited a high outlier growth of *E. coli* O157:H7, inoculated samples which exhibited a low outlier growth, and non-inoculated plants. For the three cases, the highest relative abundance (28-32%) was represented by all families that were grouped together because they accounted for less than 1% of abundance in each experimental condition. The second-most abundant family in all three cases was *Streptomycetaceae* (14-17%), but from third place and on the family abundances varied. For instance, *Staphylococcaceae* (10.3% and 9.9%) occupied this third place for leaves with no inoculation and high growth of the pathogen, while *Micrococcaceae* (10.1%) got this place for leaves with low growth of *E. coli* O157:H7.

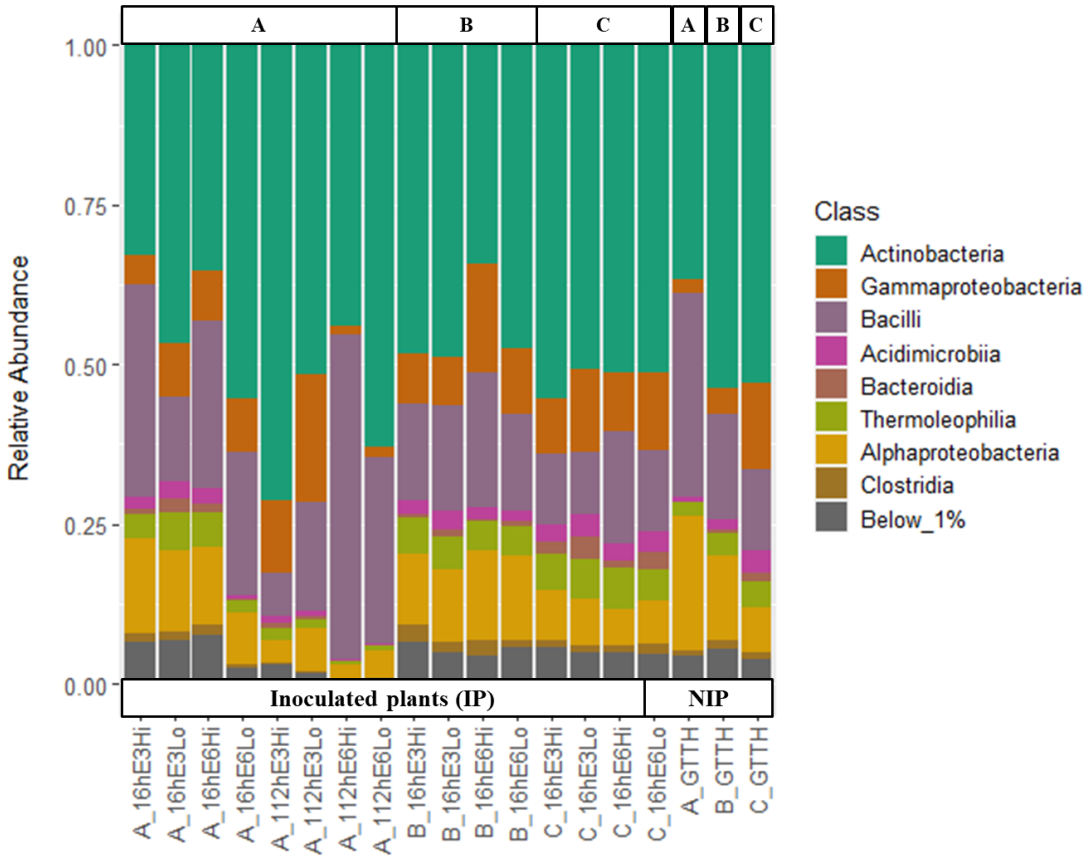


Figure 4.10. Bacterial community composition at class level of romaine lettuce plants grown and inoculated with *Escherichia coli* O157:H7 under high (A), medium (B) and low (C) relative humidity conditions. Non-inoculated plants (NIP) were included for comparison. Classes with less than 1% average abundance across samples were grouped together and represented as Below 1%. Labels in x-axis correspond to the experimental conditions describe above in table 4.4.

Differential abundance results ($P<0.05$) at the family level between inoculated plants and non-inoculated plants from the three RH conditions together (Figure 4.11) confirmed the significant depletion of families mostly from the Alphaproteobacteria class after the inoculation of *E. coli* O157:H7. Members of the Alphaproteobacteria class are predominant in phyllosphere microbiotas across many different type of plants (177). This class is identifiable by its very low level of nutrient requirements for growth, which allows them to succeed in habitats with characteristics of oligotrophy (limitation in nutrients and fluctuating physicochemical stresses) such as the phyllosphere (150). The maximal growth efficiency of oligotrophs is reached when the resources are very limited due to their ability to scavenge under harsh conditions (178); this could suggest that the inoculation of *E. coli* O157:H7 on lettuce leaves might lead to the depletion of

Alphaproteobacteria observed, due to compounds produced by the metabolism of living cells of this human pathogen or by the cell degradation of its dying cells.

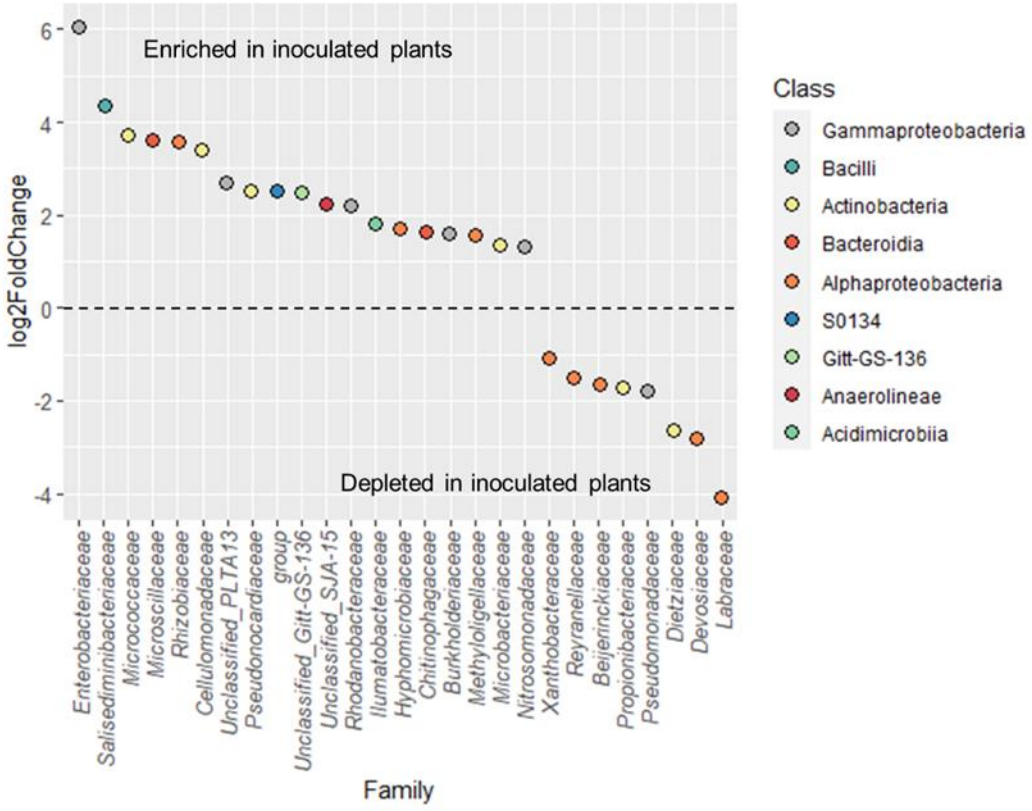


Figure 4.11. Differential abundances based on log2 fold changes of significantly enriched or depleted families between inoculated and non-inoculated plants. Each point represents one family (x-axis), colored by class. Points above the dashed line represent families significantly enriched, and points below represent depleted families.

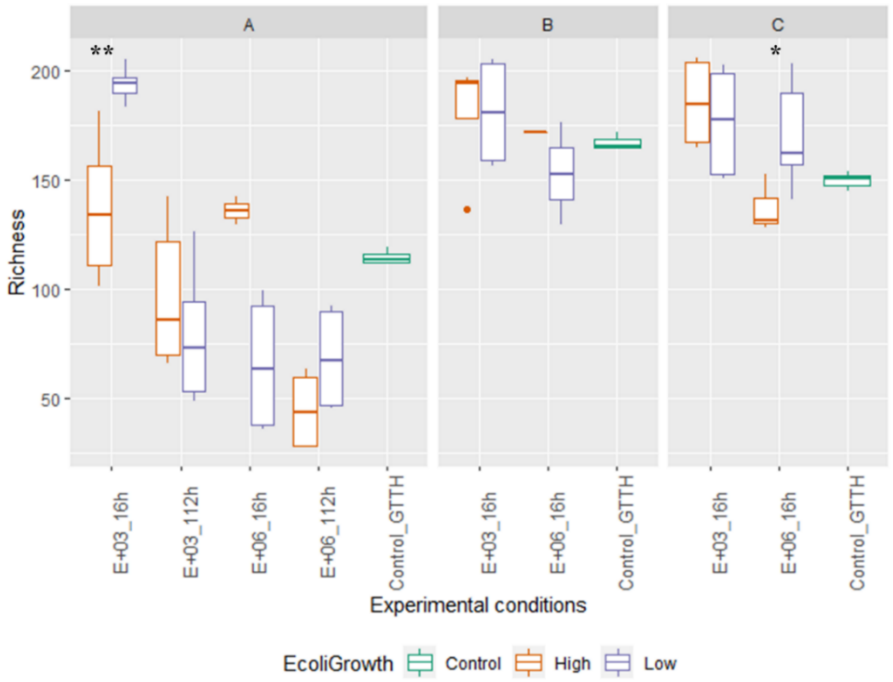
4.4.7 Impact of inoculation of *E. coli* O157:H7 on the romaine lettuce leaves bacterial microbiome

Measurements of α diversity (Richness and Shannon Index) revealed a significantly lower diversity of bacterial communities in romaine lettuce plants grown and inoculated under humid condition A when compared with intermediate condition B and dry condition C; these last two conditions were not found to be significantly different (Figure 4.12). This result was observed in both non-inoculated plants ($P<0.0001$) and inoculated plants (ANOVA: $F_{2,76}$: 28.3, $P<0.0001$). As demonstrated before in section 4.4.2, the higher population counts of *E. coli* O157:H7 were obtained after inoculation under Condition A and, in addition, the resident bacterial population

counts were positively associated with the human pathogen counts (R = 0.204, P = 0.0103). Therefore, we can conclude that humid conditions during plant development promoted bacterial growth overall. These findings suggest that the increase of bacterial populations caused by high levels of RH is dominated by few taxa, and for that reason the α diversity was lowered. The same effect has been observed even on non-living surfaces where wetted materials showed this diversity reduction (145). Finally, both measurements of alpha diversity were negatively correlated with *E. coli* O157:H7 population counts (Richness: R = -0.36, P = 0.0012 and Shannon: R = -0.26, P = 0.0190), which indicates a medium to small inverse relationship between the diversity of resident bacterial communities with the fate of the human pathogen on lettuce leaves. However, the correlation was lost when samples from each RH condition were analyzed separately.

Bacterial α diversity was then evaluated within each RH condition separately to test differences between leaf areas where growth of *E. coli* O157:H7 was high or low. Under most of the experimental conditions no significant difference was observed, with the exception of Richness values of leaves tested 16 hours after inoculation with 3 log CFU under humid condition A (P=0.0051) and dry condition C (P=0.0389) (Figure 4.12 A). As well as the Shannon index of leaves tested 16 hours after inoculation with 3 log CFU only under humid condition A (P=0.0163). These observed differences could be explained by the strong negative correlations found, only under RH condition A after 16 hours of inoculation with 3 log CFU, between both measurements of alpha diversity and *E. coli* O157:H7 population counts (Richness: R = -0.85, P = 0.0005 and Shannon: R = -0.76, P = 0.0040). Therefore, we might conclude that the association between the richness and evenness of resident bacterial communities with the fate of *E. coli* O157:H7 on lettuce leaves is not only defined by their microbial interactions but it is also affected by additive factors such as environmental conditions, initial number of invasive cells of the pathogen, time of interaction, etc. In fact, regardless the level of growth of *E. coli* O157:H7, plants from humid condition A exhibited a significantly lower alpha diversity after 112 h of inoculation than after 16 h (Richness $F_{1,28.8}$: 18.4, P=0.0002, Shannon index: $F_{1,31.3}$: 15.3, P=0.0005).

A.



B.

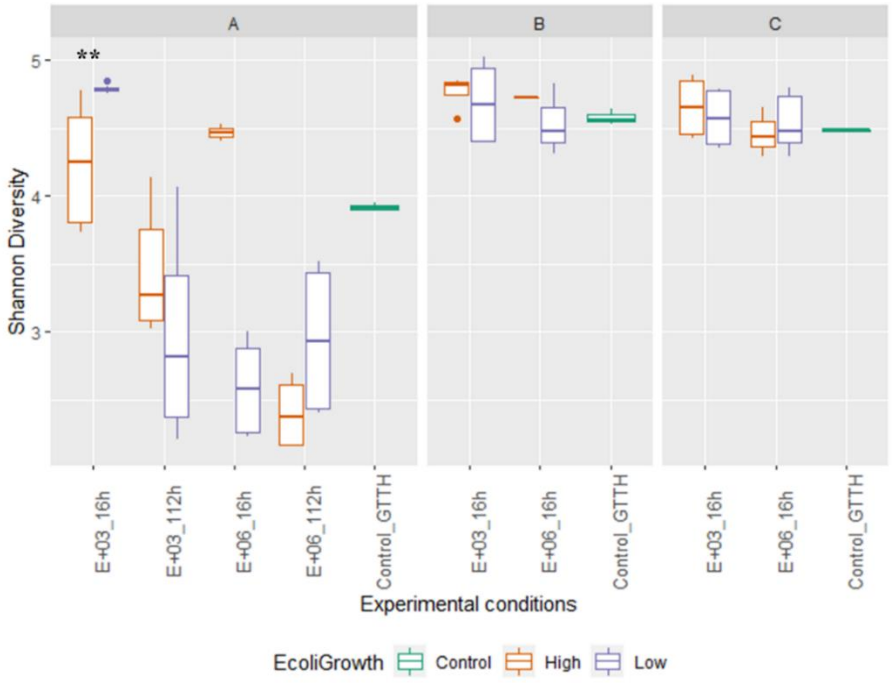


Figure 4.12. Bacterial α diversity of leaves inoculated with *E. coli* O157:H7. Panels correspond to the relative humidity levels used for the different experimental conditions. A. Richness values. B. Shannon diversity indices. Experimental conditions are inoculum of *E. coli* O157:H7 (E+03 = 3 log CFU, E+06 = 6 log CFU), testing time after inoculation (16h, 112h), and non-inoculated plants from both cultivars (Control_GTTH).

Changes in the composition of the bacterial communities were observed by unweighted Unifrac β-diversity patterns. When all RH conditions were tested together, the biggest separation along the first axis (33%) is caused by the time that *E. coli* O157:H7 has been inhabiting the lettuce leaves (Figure 4.13). It is observed how bacterial communities at 112 hours after inoculation of the pathogen were more different than at 16 hours under the same RH condition A. Additionally, the β-diversity changes occurred after 112 hours of inoculation of *E. coli* O157:H7 produced a greater differentiation of the lettuce resident bacterial communities between the plant cultivars Green Towers and True Heart. On the other hand, a second separation among the three RH conditions at 16 hours after inoculation was observed along the axis2 (7%), exhibiting a subtle succession from the more humid condition A through the intermediate RH condition B to the dry condition C. Also, it is noticeable that bacterial communities from non-inoculated plants (Control_GTTH) clustered close to the bacterial communities from plants tested after 16 hours of inoculation under all RH conditions. This demonstrates that 16 hours after inoculation of *E. coli* O157:H7, there is not enough disturbance of the resident communities even though the pathogen population counts were not significantly different from those after 112 hours (Table 4.4). However, some *E. coli* growth changes (log change CFU) at 112 hours exhibit a slight decline, which might suggest that the bacterial communities' shifts occur after *E. coli* O157:H7 has been a resident for a longer time and begins to die; this finding supports our previous idea that the release of compounds from dying cells might affect certain bacterial taxa present on lettuce leaves.

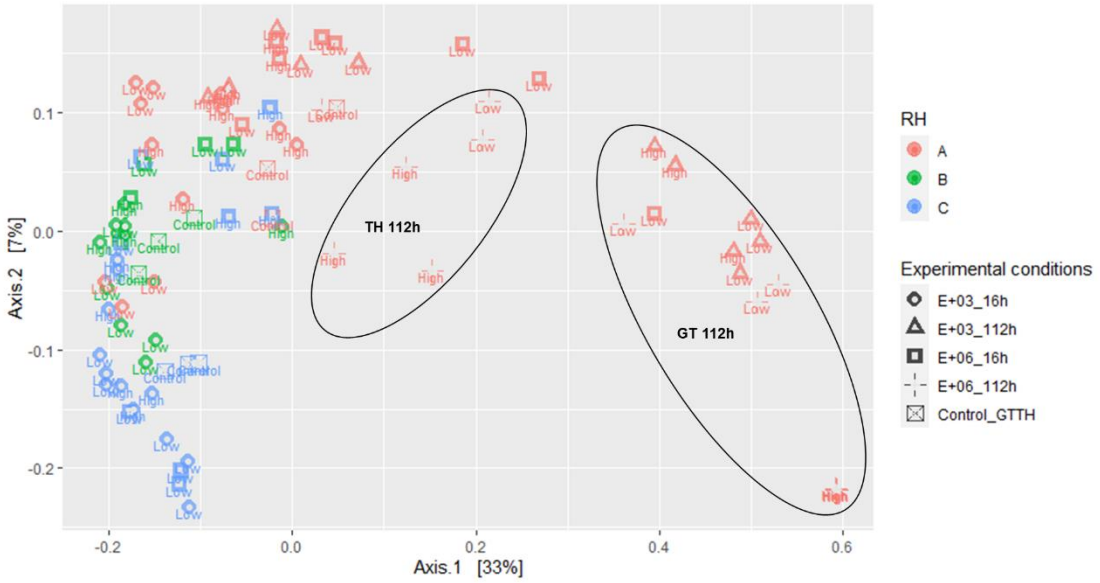


Figure 4.13 Bacterial β diversity of leaves inoculated with *E. coli* O157:H7 under the three RH conditions. Visualization with principal coordinate analysis of unweighted Unifrac distance among bacterial communities of all inoculations. Colors indicate RH conditions as labeled in the legend. Shapes indicate experimental conditions of inoculum of *E. coli* O157:H7 (E+03 = 3 log CFU, E+06 = 6 log CFU), testing time after inoculation (16h, 112h), and non-inoculated plants from both cultivars (Control_GTTH). Labels High and Low on each sample indicates the level of growth of *E. coli* O157:H7.

When PERMANOVA was applied to the bacterial β-diversity (weighted unifrac) of leaves inoculated under all RH conditions, the main factors leading the overall variation were RH and testing time after inoculation of *E. coli* O157:H7, accounting for 24.4% ($F_{2,63}$: 154.6, P=0.001) and 11.5% ($F_{1,63}$: 145.5, P=0.001), respectively; while the level of growth of *E. coli* O157:H7 accounted for only 2.0% ($F_{1,63}$: 26.2, P=0.001). For that reason, the bacterial communities were then analyzed separately within each RH condition to test if the levels of *E. coli* growth have some effect on their β-diversity. In this analysis, *E. coli* growth level accounted for 14.7% ($F_{1,63}$: 41.3, P=0.001), 14% ($F_{1,63}$: 39.5, P=0.001), 8.8% ($F_{1,63}$: 5.1, P=0.001), 15.5% ($F_{1,63}$: 25.5, P=0.001) of the overall variation under RH conditions A 16 hours, A 112 hours, B 16 hours and C 16 hours, respectively. However, other factors such as cultivar and inoculation dose accounted for similar, and in some cases a greater fraction of the variation observed in each RH condition.

Overall, our results suggest that RH is the major factor shaping the resident bacterial communities, and the growth level reached by *E. coli* O157:H7 after inoculation accounted for a small but still significant part of the communities' variation. Non-inoculated plants grown under

RH condition A exhibited significantly lower values of α-diversity, and they were more susceptible to changes in richness and evenness after the introduction of the pathogen than plants with higher α-diversity grown under less humid conditions B and C. However, the inoculation of *E. coli* O157:H7 evaluated after 16 hours did not cause a greater level of disturbance in the bacterial composition of these less diverse communities than was caused on more diverse communities from RH conditions B and C. This finding could suggest that bacterial diversity did not confer community stability against the introduction of *E. coli* O157:H7, as suggested in previous studies (72, 73). However, 112 hours after inoculation, we observed a significant disturbance of bacterial communities under humid condition A; future studies should evaluate if this effect is also present under intermediate and dry conditions which exhibited higher bacterial α-diversity. It would also be useful to perform an evaluation of cultivars with contrasting values of bacterial α-diversity under the same environmental condition to determine their resistance to change after the introduction of *E. coli* O157:H7. Finally, as part of the variation of the resident bacterial communities is explained by the level of growth reached by *E. coli* O157:H7 on lettuce leaves, differential abundances at a lower taxonomic level were investigated.

4.4.8 Identification of genera with differential abundances in native bacterial communities of romaine lettuce leaves after the inoculation of *E. coli* O157:H7

To evaluate if higher levels of growth of *E. coli* O157:H7 were associated with some specific taxa, differential abundance analysis using DESeq2 was performed within each RH condition to avoid the differences caused by the environment. Comparisons were performed first between non-inoculated and inoculated plants, and then between inoculated plants exhibiting high and low growth of *E. coli* O157:H7, so a best interpretation of the changes observed can be made. Under RH condition A at 112 hours (Figure 4.14 A), genera *Salinicoccus, Microbacterium, Actinomycetospora,* and one unclassified genus from the *Rhizobiaceae* family were significantly ($P<0.05$) enriched in lettuce leaves after the inoculation of *E. coli* O157:H7. We also observed two genera from class Alphaproteobacteria that were significantly depleted after the inoculation of the pathogen, specifically genera *Afipia* and *Labrys* appear to be the most affected by the introduction of a foreign microorganism. When inoculated plants exhibiting contrasting growth of *E. coli* O157:H7 were compared (Figure 4.14 B), *Microbacterium* was significantly enriched when high population counts of the pathogen were achieved; while genera from class Alphaproteobacteria

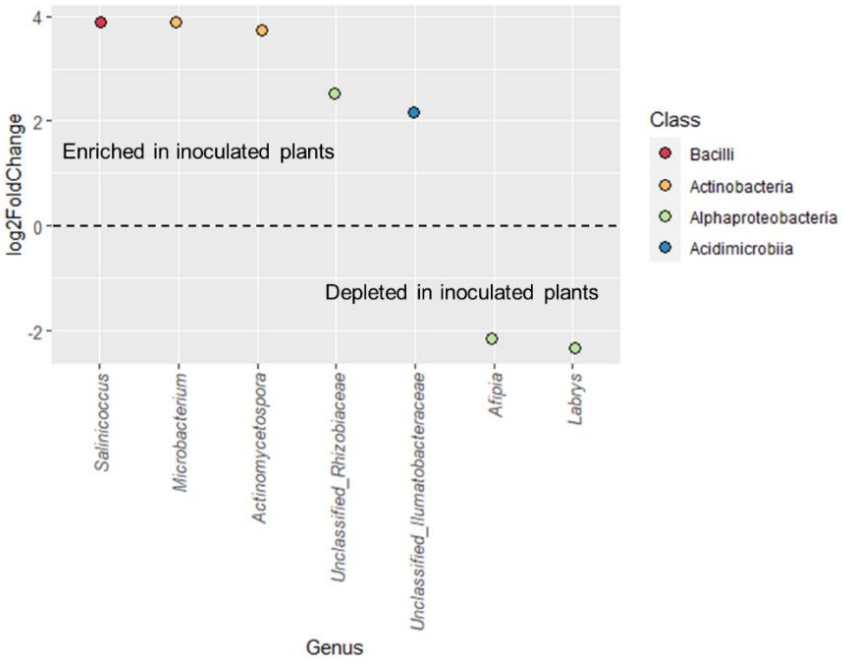
such as unclassified *Rhizobiaceae, Caulobacter, Methylobacterium-Methylorubrum and Phenylobacterium* exhibited a significantly larger decline when exposed to higher numbers of *E. coli* O157:H7 CFU. A clear tendency of Alphaproteobacteria to decrease in response to the presence and the increase of population of the human pathogen was observed. Microbial competition can be the result of several factors such as overlap of nutrient requirements, production of antimicrobial compounds, limitation of space, increase of osmotic pressure for overconcentration, etc. (53, 59). We hypothesize that the observed changes in abundances of Alphaproteobacteria might be caused by an accumulation of secondary metabolites from *E. coli* O157:H7 which disrupt the low-nutrient requirements of this slow-growing bacterial class; similar to what has been observed in soils where nutrient enrichment selected for more copiotrophic and fast-growing microbes rather than for oligotrophs (179, 180).

Under RH condition B after 16 hours of inoculation (Figure 4.15), only one genus was enriched in inoculated plants, and more dramatic was the depletion of mainly Actinobacteria. However, no genera with significant differential abundance were found between inoculated plants with high and low growth of *E. coli* O157:H7. As previously observed on the β-diversity plot (Figure 4.13), the resident bacterial communities after 16 hours of inoculation did not exhibit big disturbances from communities of non-inoculated plants. However, RH condition C showed a larger dispersion in the β-diversity plot, therefore genera with significant differential abundance were found between inoculated plants with high and low growth of *E. coli* O157:H7, even at 16 hours. But none of these genera corresponds with the differences observed between non-inoculated and inoculated plants. Finally, the identification of possible bacterial ASVs as biomarkers of lettuce leaves with high population counts of *E. coli* O157:H7 was performed using LEFSe (154). Bacterial communities from non-inoculated and inoculated plants (high and low growth) were compared, the threshold on the LDA score for discriminative features was set to 4.0, and ASVs exhibiting statistically significant change ($P < 0.05$) were selected (Figure 4.17). ASV58 corresponded to an unclassified genus of *Rhizobiaceae* family and was defined as a biomarker for bacterial communities where *E. coli* O157:H7 reached low population counts on romaine lettuce leaves. On the other hand, ASV57 classified as *Microbacterium* was identified as a biomarker for bacterial communities where the human pathogen succeeded and reached high population counts. Interestingly, *Microbacterium* (genus from the Actinobacteria class) has been also reported as significantly enriched after the inoculation of *E. coli* O157:H7 on lettuce plants grown in

laboratory conditions containing microbiota from field-grown plants (73). Therefore, *Microbacterium* is suggested as a possible indicator of heavy contamination of romaine lettuce leaves with *E. coli* O157:H7.

Overall, disturbances of the resident bacterial communities and biomarkers were identified after 112 hours of inoculation with *E. coli* O157:H7. Significant changes in abundance of certain taxa were detectable starting at 16 hours, but as demonstrated with our β-diversity analysis, these changes were not enough to modify the entire structure of the communities. However, a longer exposure of the resident communities to the pathogen caused a larger disturbance of their organization, which was associated with contrasting levels of growth of *E. coli* O157:H7 and the presence of specific ASVs. Our results demonstrate that shifts in the composition of plant-associated microbiota caused by the introduction of a foreign microbe not only can be studied on field-grown plants as suggested by William *et al.* (73), but also on plants grown in a growth chamber, as long as the environmental conditions support bacterial growth.

A.



B.

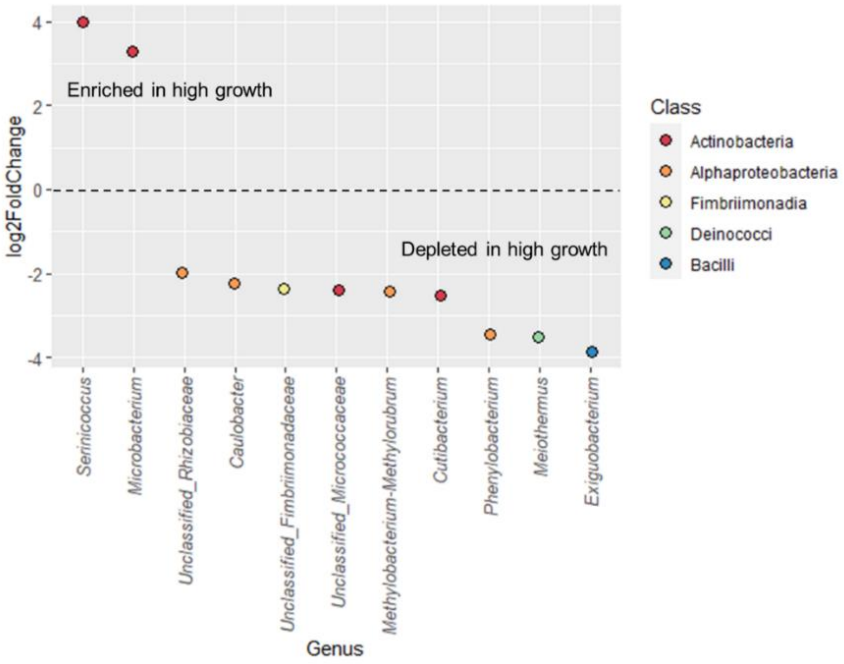


Figure 4.14. Bacterial genera with differential abundances based on log2 fold changes under RH condition A. Abundances were compared between: (A) non-inoculated and inoculated plants with *Escherichia coli* O157:H7 (16 and 112 hours after inoculation together). (B) inoculated leaf areas with high and low growth of *E. coli* O157:H7 at 112 hours (note: no significant genera were found at 16 hours). Each point represents one genus (x-axis), colored by class. Points above the dashed line represent genera significantly enriched, and points below represent depleted genera

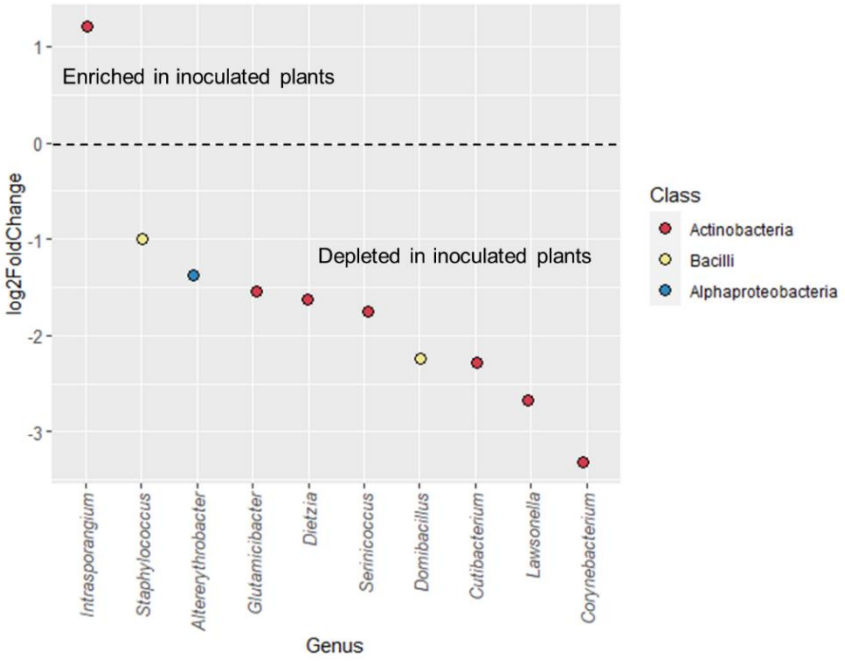
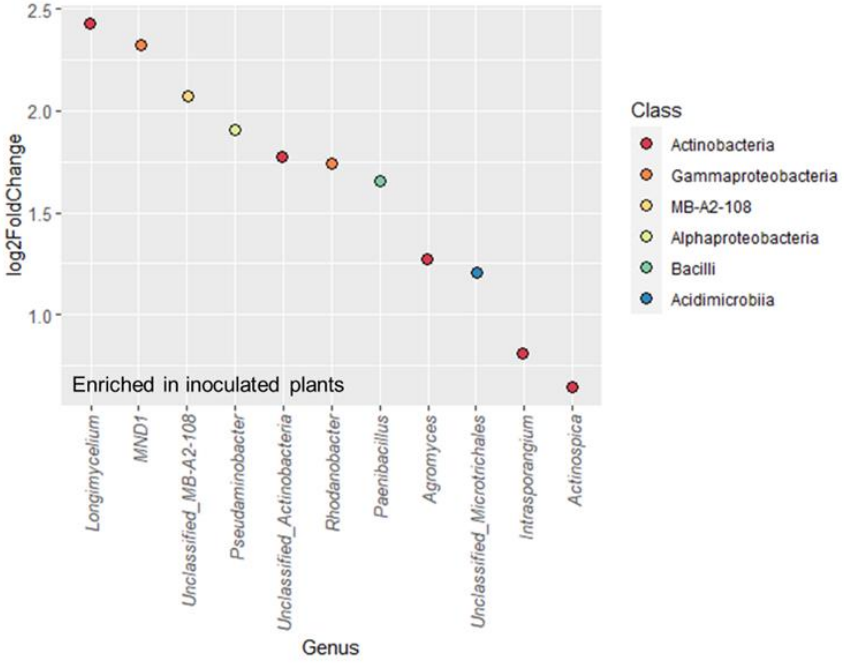


Figure 4.15. Bacterial genera with differential abundances based on log2 fold changes under RH condition B. Abundances were compared between non-inoculated and inoculated plants with *Escherichia coli* O157:H7 at 16 hours after inoculation. No significant genera were found between leaf areas with high and low growth of *E. coli* O157:H7 at 16 hours. Each point represents one genus (x-axis), colored by class. Points above the dashed line represent genera significantly enriched, and points below represent depleted genera.

A.



B.

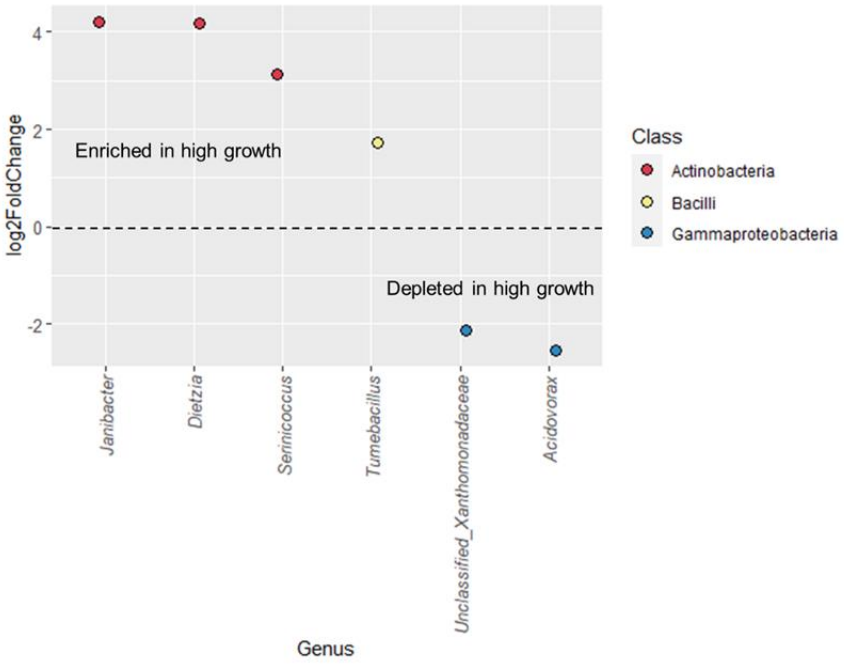
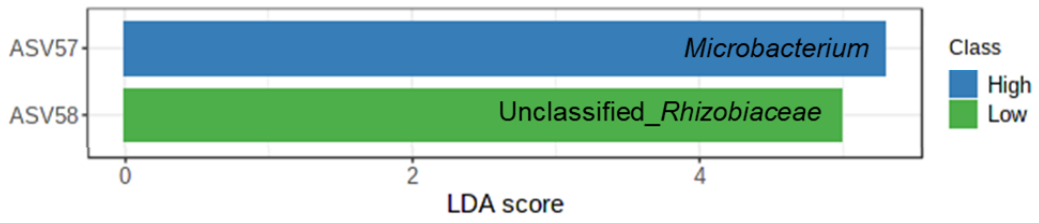


Figure 4.16. Bacterial genera with differential abundances based on log2 fold changes under RH condition C. Abundances were compared between: (A) non-inoculated and inoculated plants with *E. coli* O157:H7 at 16 hours after inoculation. (B) inoculated leaf areas with high and low growth of *E. coli* O157:H7 at 16 hours after inoculation. Each point represents one genus (x-axis), colored by class. Points above the dashed line represent genera significantly enriched, and points below represent depleted genera.

A.



B.

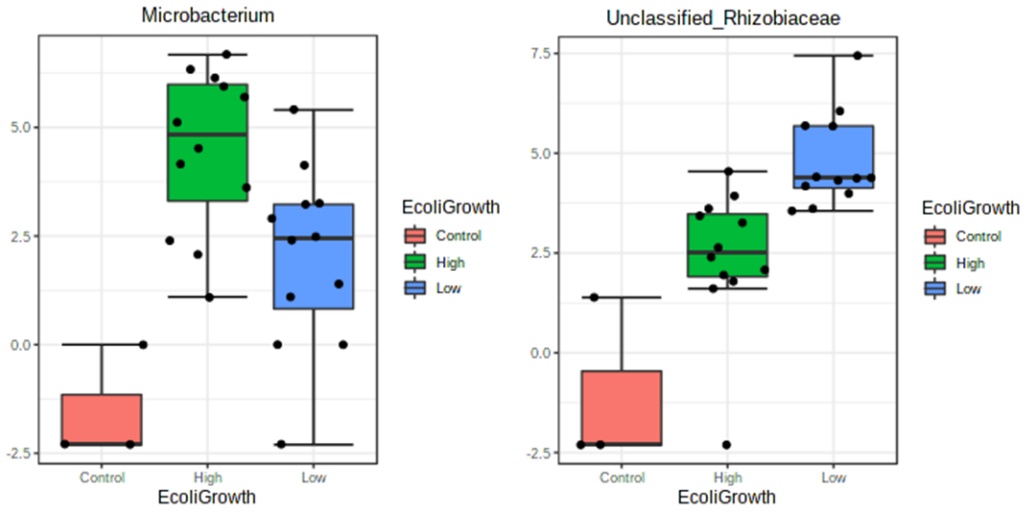


Figure 4.17. Taxa (ASVs) identified as biomarkers of high and low growth of *E. coli* O157:H7 on romaine lettuce leaves. (A) Linear discrimination analysis (LDA) effect size (LEfSe) for inoculated samples under RH condition A after 112 hours. The length of the bar represents the log10 transformed LDA score. The threshold on the LDA score for discriminative features was set to 4.0, and the taxa with statistically significant change ($P < 0.05$) was selected. (B) Boxplot showing abundances (log-transformed counts) of the biomarkers between control (non-inoculated leaves) and leaves with high and low growth of *E. coli* O157:H7.

CHAPTER 5. CHARACTERIZATION OF CULTURABLE BACTERIAL COMMUNITIES OF ORGANIC AND CONVENTIONAL ROMAINE LETTUCE USING ELASTIC LIGHT SCATTERING TECHNOLOGY

5.1 Abstract

This study investigated the culturable bacterial communities of organic and conventional romaine lettuce available in grocery stores, and whether the laser optical scatter technology designated BARDOT (Bacterial Rapid Detection using Optical Scatter Technology) could be used as a rapid classification method of the nine most abundant bacterial genera of this leafy green. The culturable bacterial population counts and bacterial richness (number of Operational Taxonomic Units - OTUs) were significantly different between organically and conventionally labeled lettuce ($p<0.01$). The composition of their bacterial communities also differed: organic samples contained a total of 41 genera of which 20 were exclusive, while conventional contained a total of 24 genera of which only 3 were exclusive.

Three scatter image libraries were built to classify the nine most abundant bacterial genera isolated from romaine lettuce, which covered around 70-76% of the total culturable bacterial population. The training parameters achieved classifiers at the genus level with positive predictive values (PPVs) between 90.6-99.8%. This was validated with blind samples that resulted in sensitivity and average classification accuracy values above 90% for both pure and mixed cultures. The sensitivity and classification accuracy per genus when new lettuce samples were tested, showed values between 51.9-79.1% and 42.9%-100%, correspondingly. Our results show the potential of BARDOT technology to characterize culturable bacterial communities from environmental samples, and even to study microbial interactions. However, specific challenges such as overlapping scatter patterns from strains of different bacterial genera, and the need of continuously feeding the libraries with new scatter images of colonies isolated from new samples, were identified.

5.2 Importance

Studies of biological control of colonization of human pathogens on fresh produce require identification and preservation of the native culturable bacterial communities. The identification

process involves multiple steps after their isolation such as biochemical tests, polymerase chain reaction (PCR) and Sanger sequencing. We characterized the culturable bacterial communities of romaine lettuce and applied the elastic light scattering technology to classify their nine most abundant bacterial genera. This study demonstrates the potential and challenges of BARDOT as a tool to identify entire culturable bacterial communities from environmental samples and suggests improvements to mitigate the limitations.

5.3 Introduction

When human-pathogenic bacteria land on a new ecological niche such as the plant phyllosphere, they need to overcome barriers such as plant defenses (29, 46, 181), harsh environmental conditions (2, 3, 182) and microbial competition with the resident microbiota (53, 59). Several studies have been approaching the question of the role of microbial populations as inhibitors or enhancers of the growth of human pathogens on plants (56, 57, 183), which is a current global food safety concern.

The characterization of bacterial communities can be performed through culture-dependent or culture-independent methods such as Next Generation Sequencing (NGS) (68, 74, 77). Although NGS offers a better approximation to the real microbial diversity, it does not allow preservation of the identified bacteria for future evaluations of potential phenotypes. Additionally, in the field of food safety, both big and small fresh produce growers aim to assure the quality of their products; hence, the availability of more simple, rapid, accessible and inexpensive technologies to identify culturable bacterial populations is desirable.

The identification of culturable bacterial populations usually involves multiple steps after their isolation such as the establishment of pure culture, the performance of biochemical tests or polymerase chain reaction (PCR) and Sanger sequencing. These methods add more time to the identification process and in the last case demand access to resources of molecular biology and external sequencing services. A laser light-scattering sensor called BARDOT (Bacterial Rapid Detection using Optical Scatter Technology) was created as a direct, real-time and label-free identification method (78). Compared with other direct microbial identification methods such as matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS) (79), BARDOT is applied directly on the bacterial colonies growing onto the culture media and preserves the viability of the microorganisms.

The BARDOT system uses a 635-nm laser, which interrogates the whole volume of the colony thus generating a scatter pattern influenced by refractive indices, extracellular material, density and individual shape of the bacterial cells (184). These patterns have been used as a fingerprint to be compared against a pre-trained scatter-signature image database which can detect specific targets such as foodborne bacterial pathogens (78, 81–83), differentiate serovars (84), serogroups (85), and virulence gene-associated mutant colonies (86). However, a broader application such as the characterization of entire culturable bacterial communities from environmental samples has not been tested yet using the BARDOT technology.

Therefore, the objectives of this study were the characterization of the culturable bacterial communities associated with organic and conventional romaine lettuce, which is a leafy green frequently related with foodborne disease outbreaks (9). Secondly, the evaluation of the BARDOT technology as a tool to provide real-time differentiation and classification of the nine most abundant bacterial genera in romaine lettuce.

5.4 Materials and Methods

5.4.1 Sample Collection and Processing

Twenty-four romaine lettuce heads were purchased from grocery stores in Lafayette, Indiana and Chicago, Illinois, USA from November 2015 to December 2016. Twelve conventional and twelve organic lettuce heads representing nine brands were handled by avoiding skin contact during their transportation from store to laboratory. All samples were processed the same day or the day after they were purchased, and they were stored in a refrigerator (4˚C) when needed.

For each lettuce head, 25 grams of intact, non-washed leaves were collected and blended into 225 milliliters of phosphate buffer (PB) 0.1M pH 7.0 to isolate the culturable bacterial communities from the phylloplane and endosphere. Then, 10-fold dilutions were performed and spread onto Plate Count Agar (PCA) plates, which were incubated at 30˚C for 72 hours (185).

5.4.2 Culture-Dependent Characterization

After the time of incubation, colonies were counted, and the total bacterial populations were expressed as CFU g^{-1} leaf tissue. Separated colonies were phenotypically classified based on color, form, elevation and margin. Representative isolates from different observed morphologies

were transferred onto a new plate, and the region V3-V6 of their 16S rRNA gene was amplified by colony PCR using the primers 338F-ACTCCTACgGGAGGCAGCAG and 1046R-AGGTGSTGCATGGcTGTCG (93, 186). To assure amplification, colonies were dissolved in 20µl of sterile PCR water, then were heated at 99˚C for 5 minutes using a thermocycler, and finally applied a short centrifugation to settle out cell debris. From this supernatant, 1µl was used to perform the PCR protocol of 95˚C for 2 min, followed by 30 cycles of 95˚C for 30 sec, 53˚C for 30sec, and 72˚C for 45 sec, with a final elongation of 72˚C for 7 min. Amplicons were paired-end sequenced with Sanger technology and assigned taxonomy to the genus level through BLAST 2.6.0 (187) and RDP classifier 2.12 (188).

5.4.3 Analysis of Culturable Bacterial Communities

Contigs from Sanger sequencing were analyzed using QIIME (189) and Uclust (190). The total culturable population was clustered into OTUs based on a threshold of similarity of 97%. A phylogenetic tree was built using RAxML 7.2.8 and R 3.4.2 with the GTRGAMMA method. To evaluate the differences between organic and conventional lettuce plants, pairwise t-tests were applied using JMP® Version Pro 14.

5.4.4 Collection of Colony Scatter Patterns

A prototype of a commercial laser light-scattering sensor built by Hettich GmbH, was used to collect scatter patterns. From the nine most abundant bacterial genera isolated from romaine lettuce leaves, 29 strains were selected (Table 5.2) as representatives of the 20 most abundant OTUs which include 76.7% of the total isolated strains. These selected strains were screened through the BARDOT system to determine constant plate reading parameters and incubation times which provide a well resolved scatter pattern for the different bacterial genera inhabiting romaine lettuce (Table 5.1).

Table 5.1. Standardized reading parameters of plates and libraries for classification of bacterial communities from romaine lettuce using the BARDOT system

Plate	**Library 21-22 hours**	**Library 32 hours**	**Library 72 hours**
Agar thickness: 1.9997 Agar volume: 20 ml Plate exposure: 601 Radius: 1,300 Colony diameter: 0.3-1.2mm Colony roundness: 0.3	Incubation time: 21 and 22 hours Exposure: Auto150 Zoom: 2X	Incubation time: 32 hours Exposure: Auto175 Zoom: 2X	Incubation time: 72 hours Exposure: Auto150 Zoom: 2X

Table 5.2. Nine most abundant culturable bacterial genera from romaine lettuce; 29 strains were selected to develop the scatter image libraries as representatives of each of the 20 most abundant OTUs

Genera	**Number of OTUs (97%)**	**Selected strains**	**Source (lettuce type)**
Pseudomonas	6	556Ps 532Ps 143Ps 39Ps 100Ps 159Ps	Organic Organic Conventional Conventional Conventional Organic
Arthrobacter	4	526Ar 165Ar 445Ar 510Ar	Organic Organic Organic Organic
Microbacterium	1	441Mi 410Mi 284Mi 289Mi	Organic Organic Organic Organic
Erwinia	2	483Er 630Er 518Er	Organic Conventional Organic
Massilia	1	128Ma 402Ma	Conventional Organic
Duganella	2	96Du 382Du	Conventional Organic

Table 5.2 continued

Pantoea	1	7Pa 114Pa 499Pa	Conventional Conventional Organic
Curtobacterium	1	536Cu 586Cu	Organic Conventional

For collection of scatter patterns, pure bacterial cultures of each strain were grown overnight (16 hours) in LB broth, at 30˚C with shaking. After the incubation period, cultures were serially diluted (10-fold) in 0.1M phosphate buffer pH 7.0, surface spread on plate-count agar (PCA) and incubated at 30˚C resulting in fewer than 150 colonies per plate. Scatter patterns from each strain were collected in at least 3 independent experiments with a minimum of 50 images per strain.

5.4.5 Development and Training of Scatter Image Libraries

For identification of the most abundant culturable bacterial genera of romaine lettuce, these were split into three groups according to their growth rates. Three libraries at the genus level were developed: Library A was composed of 17 strains of *Pantoea* spp., *Erwinia* spp., *Pseudomonas* spp., *Massilia* spp., *Duganella* spp., *Flavobacterium* spp. and *Arthrobacter* spp which were incubated for 21-22 hours. Library B was composed of 15 strains of *Arthrobacter* spp., *Duganella* spp., *Curtobacterium* spp., *Microbacterium* spp., *Massilia* spp., *Flavobacterium* spp., *Erwinia* spp. and *Pseudomonas* spp. incubated for 32 hours. Library C was composed of 13 strains of *Microbacterium* spp., *Arthrobacter* spp., *Duganella* spp. and *Flavobacterium* spp for 72 hours of incubation.

The training of the libraries was performed under the linear support vector machine (SVM) classifier (191) using built-in image analysis software (192). The libraries were trained by computing a vector of 390 features for each scatter image, including circular/ring features using magnitudes of Zernike moments (Order=20), granular/roughness features using Haralick gray-level co-ocurrence matrices (Distance=1, Levels=4) and spatial frequency component of images using a Fourier transform based method (78, 191, 193).

5.4.6 Validation of Libraries

Validation was carried out by testing the classification performance of the libraries with three types of samples: blind samples of pure cultures, blind samples of mixed cultures and new lettuce samples. Validation with blind samples was carried out by collecting new scatter patterns from the same strains used to build the libraries, to test their correct classification. Pure cultures of blind samples were grown as described above. Mixed cultures of blind samples resulted by mixing three liquid pure cultures of different strains into fresh LB broth and then growing them together during the overnight period. For the final validation, eight (four conventional, four organic) new lettuce heads were bought from stores and 215 bacterial colonies were isolated, classified using the BARDOT system with the developed libraries and corroborated their identity through Sanger sequencing as previously described.

5.4.7 Image and Feature Analysis

Training libraries, classification of validation sets, and analysis of scatter pattern differences were performed by using the image classifier software (192). Ten-fold cross-validation (CV) matrices were generated to evaluate the quality of the classifiers for each library. Computed positive predictive value (PPV) reports the probability that a colony truly belongs to the classifier (bacterial genus or group of genera) assigned by the BARDOT system (85, 194). The classification accuracies of the libraries were calculated, using the validation sets, as the average percent of colonies per strain and per classifier that were correctly classified. The sensitivity for each developed library was calculated as the proportion of the bacterial population from blind samples (pure and mixed cultures) and from new lettuce samples that produced a true result. Limitations of the BARDOT technology were shown by calculating the PPV of the conflicting classifiers and the classification error when they are used for classification. The measurement of biophysical characteristics of the colonies such as height and diameter was performed using a phase-contrast microscope with an integrated colony morphology analyzer (ICMA) by averaging 10-20 colonies per plate (86, 195). The analysis of number of rings was performed in MATLAB by converting 10 scatter patterns per strain from polar to cartesian coordinates, then using the function “findpeaks” on the extracted line data. Statistical analysis was performed in R using one-way analysis of variance (ANOVA) with Tukey’s multiple comparisons (P value < 0.05).

5.5 Results

5.5.1 Culturable Bacterial Communities from Conventional and Organic Romaine Lettuce

Total culturable bacterial population counts from organic romaine lettuce leaves ranged between 1.9 x 10^4 CFU g^{-1} and 8.9 x 10^6 CFU g^{-1}, while conventional leaves ranged between 5.0 x 10^3 CFU g^{-1} and 5.8 x 10^5 CFU g^{-1}. Bacterial counts from organically grown romaine lettuce were significantly higher than those from the conventionally grown lettuce ($p<0.01$).

A total of 645 isolates were selected, 347 coming from organic romaine lettuce leaves and 298 from conventional leaves. Sequencing of the region V3-V6 of the 16S rRNA showed that these bacterial isolates represented 44 genera from four phyla: *Proteobacteria, Actinobacteria, Bacteroidetes* and *Firmicutes,* and eight classes: Actinobacteria, Alphaproteobacteria, Bacilli, Betaproteobacteria, Cytophagia, Flavobacteria, Gammaproteobacteria and Sphingobacteriia. The total culturable bacterial communities from both types of romaine lettuce were clustered into 69 de novo OTUs based on the 97% threshold of similarity (Figure 5.1).

Pseudomonas (31.7%, 47.0%) and *Arthrobacter* (16.4%, 13.1%) were the most abundant genera in both organic and conventionally grown romaine lettuce, correspondingly. Differences in the less abundant genera of organic and conventional culturable bacterial communities were observed (Table 5.3), for instance *Microbacterium* showed to have higher populations in organic lettuce leaves while *Massilia* predominated in conventional ones. Organic communities were more diverse, since they were composed by 41 bacterial genera, while conventional communities were composed of only 24 genera. Additionally, more culturable bacterial genera appeared to be unique to the organic leaves than to the conventional leaves (Figure 5.2). This higher bacterial richness of organic romaine lettuce leaves was corroborated by the significant difference in number of OTUs when compared with conventional leaves ($p<0.01$).

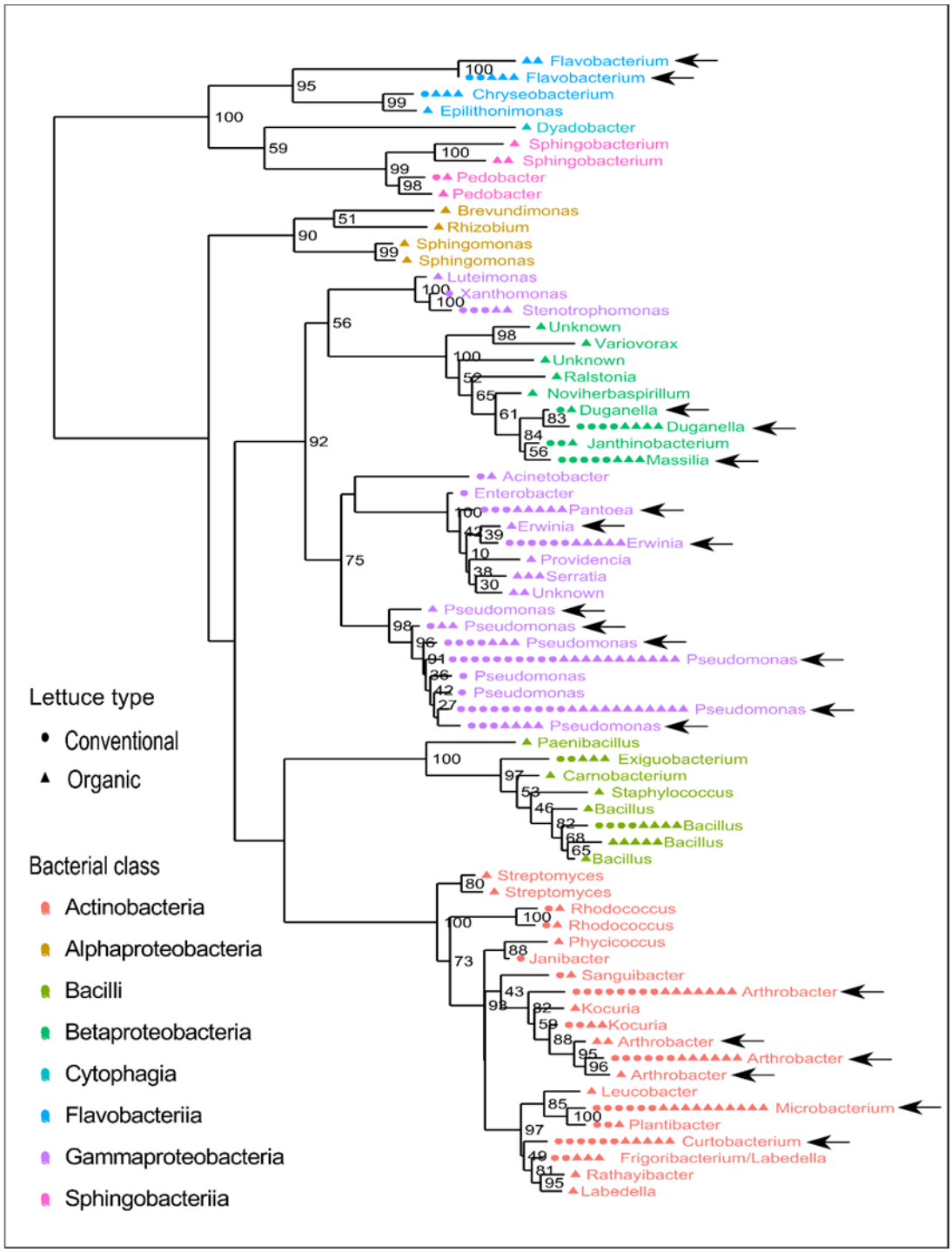


Figure 5.1. De novo OTU clustering (similarity threshold of 97%) based on sequences of region V3-V6 of the 16S rRNA. The 645 isolates were clustered into 69 OTUs, classified at genus level and analyzed by the maximum likelihood GTRGAMMA method and bootstrap analysis with 1,000 replications. Colors represent the different bacterial classes, (●) abundance from conventional lettuce, (▲) abundance from organic lettuce, (←) 20 most abundant OTUs.

Table 5.3. Comparison of relative abundance of the genera representing around 80% of the culturable bacterial communities of organically and conventionally grown romaine lettuce.

ORGANIC		**CONVENTIONAL**	
Genus	**Relative abundance (%)**	**Genus**	**Relative abundance (%)**
Pseudomonas	31.7	*Pseudomonas*	47.0
Arthrobacter	16.4	*Arthrobacter*	13.1
Microbacterium	8.9	*Massilia*	5.7
Bacillus	4.3	*Frigoribacterium*	3.7
Erwinia	4.3	*Duganella*	3.4
Flavobacterium	3.2	*Erwinia*	3.4
Duganella	2.6	*Pantoea*	3.4
Pantoea	2.3	*Curtobacterium*	3.0
Chryseobacterium	2.0	*Microbacterium*	2.7
Sphingomonas	1.7	*Stenotrophomonas*	2.3
Massilia, Curtobacterium	1.4	*Bacillus*	2.0

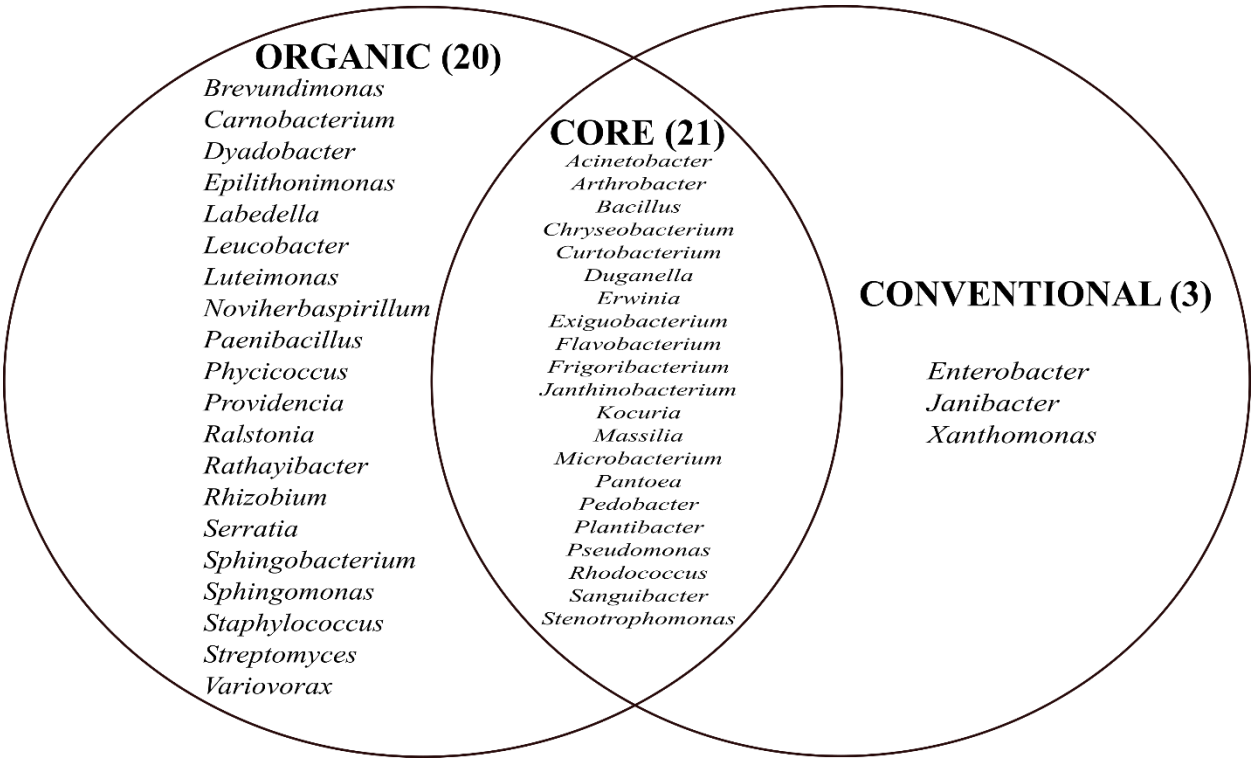


Figure 5.2. Bacterial genera isolated from leaves of organically and conventionally grown romaine lettuce. Intersection of the circles shows the bacterial genera shared between the two types.

5.5.2 Light-Scattering Libraries

A total of 8,418 scatter patterns were collected to develop the three light-scattering libraries (Table 5.4). Library A was composed of three classifiers containing 4052 scatter images. Library B was composed of seven classifiers containing 2599 scatter images. Library C was composed of 5 classifiers containing 1767 scatter images. Some classifiers were composed of more than one genus due to the similarity of their scatter patterns. Only in the case of the classifier Erwinia/Pseudomonas included in Library A, the development of a secondary library to differentiate specifically these two genera increased the power of the classifiers and the classification accuracy. Each library was trained to obtain classifiers with PPV above 90% when generating the CV matrices (Figure S1 in supplemental material).

Table 5.4. Structure of developed libraries: classifiers at genus level, number of strains and number of scatter images included

Classifiers	Strains	# Scatter Images
Library A (21-22 hours)		
Pantoea	499Pa, 114Pa, 7Pa	846
Erwinia/Pseudomonas	483Er, 630Er, 518Er, 556Ps, 532Ps, 143Ps, 39Ps, 100Ps, 159Ps	2,227
Massilia/Duganella/Flavobacterium/Arthrobacter	402Ma, 96Du, 382Du, 196Fl, 165Ar	979
TOTAL	17	4052
Library B (32 hours)		
Arthrobacter	165Ar, 445Ar	307
Duganella/Arthrobacter	96Du, 382Du, 526Ar	507
Curtobacterium/Microbacterium/Arthrobacter/Massilia	536Cu, 586Cu, 284Mi, 510Ar, 128Ma	865
Massilia	402Ma	184
Flavobacterium	196Fl, 378Fl	382
Erwinia	630Er	175
Pseudomonas	39Ps	179
TOTAL	15	2599
Library C (72 hours)		
Microbacterium	284Mi, 289Mi	420
Arthrobacter	LT3-4-20, LT2-4-40	193
Duganella	382Du	143
Flavobacterium	215Fl, 196Fl, 378Fl	430
Microbacterium/Arthrobacter	441Mi, 410Mi, 526Ar, 510Ar, LT4-4-22	581
TOTAL	13	1767

5.5.3 Validation of Light-Scattering Libraries

Blind samples and new lettuce samples were used to validate the classification accuracy and sensitivity of the developed libraries. A total of 15,356 new scatter patterns were collected from blind samples, which were the same bacterial strains included in our libraries grown in pure (6,739 images) and mixed (8,617 images) cultures. The sensitivity of the generated libraries when applied to the blind population ranged between 94.5-96.6% for pure cultures and 93.0-96.2% for mixed cultures. Moreover, the average classification accuracy of our classifiers ranged between 93.7-96.1% for pure cultures and 91.4-96.0% for mixed cultures. However, some specific classifiers had lower classification accuracy such as Flavobacterium in Library C with 86.6% and 87.5% in pure and mixed cultures; or specific strains such as *Massilia* 402Ma in Library B with 73.7% in mixed cultures (Table 5.5). The strain 402Ma showed a significant difference in classification accuracy between pure (97.4%) and mixed (73.7%) cultures, which was caused by the development of a few colonies with different morphology and scatter pattern when co-cultured overnight with other bacterial strains (Figure 5.3).

Table 5.5. Sensitivity (%) of developed libraries (calculated as proportion of the bacterial population from blind samples of pure and mixed cultures that produced a true result) and classification accuracy (%) of classifiers (average ± SEM) included in each library when applied to blind samples.

	PURE CULTURES			**MIXED CULTURES**		
Library	**Sensitivity**	**Classification accuracy of classifiers**	**Classifier with the lowest classification accuracy**	**Sensitivity**	**Classification accuracy of classifiers**	**Classifier with the lowest classification accuracy**
A	95.8	96.1 ± 0.5	Erwinia/Pseudomonas 95.3	96.2	96.0 ± 1.1	Massilia/Duganella/Flavobacterium/Arthrobacter 93.8
A Erw-Pseudo	94.5	95.3 ± 2.3	Pseudomonas 93.0	93.0	92.7 ± 0.9	Erwinia 91.8
B	96.6	96.4 ± 0.7	Erwinia 92.6	94.0	92.2 ± 3.5	Massilia 73.7
C	95.3	93.7 ± 3.6	Flavobacterium 86.6	92.8	91.4 ± 3.9	Flavobacterium 87.5

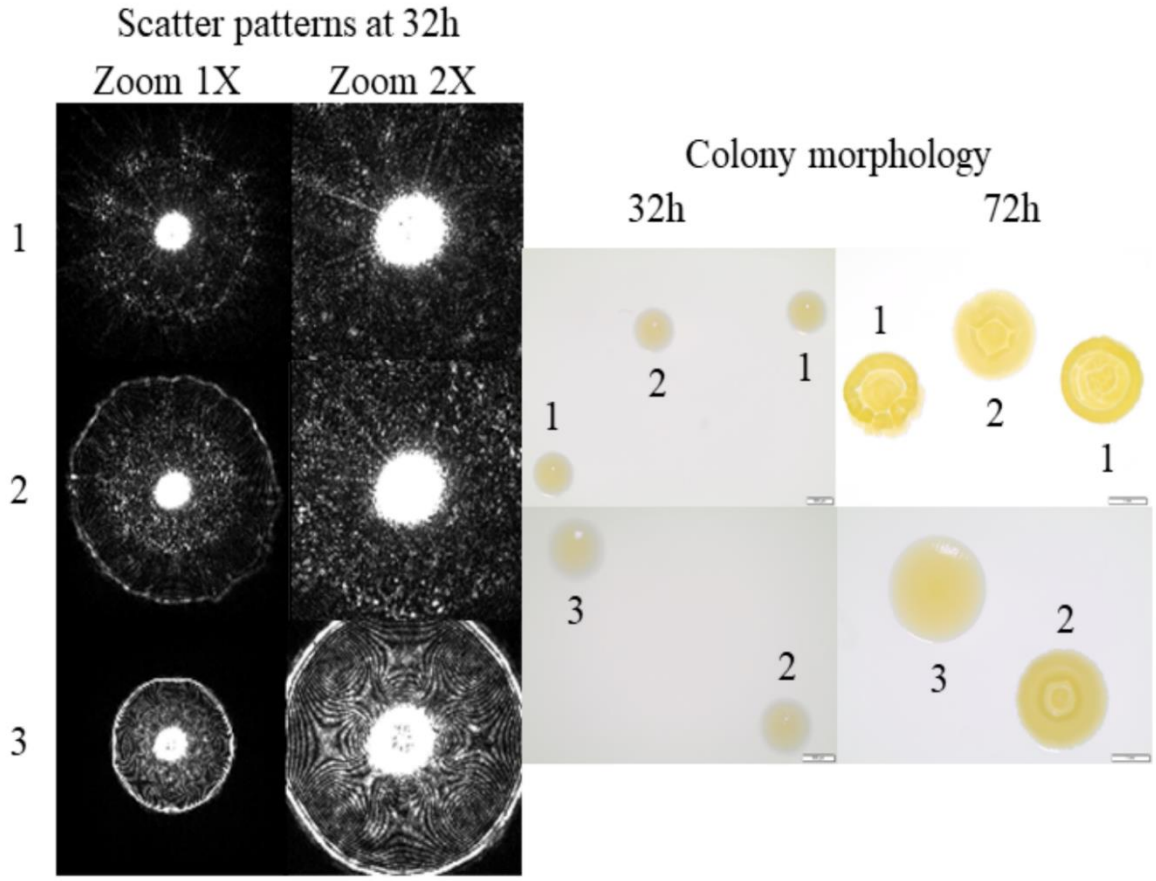


Figure 5.3. Different scatter patterns and colony morphologies produced by *Massilia* strain 402Ma when grown on PCA for 32 hours. Scatter patterns and colony morphologies number 1 and 2 are produced in pure cultures, while number 3 are produced in few colonies from mixed cultures. Colonies grown for 72h (scale bar= 1mm) exhibited the morphological differences, since colonies at the BARDOT reading time (32h) are not fully developed (scale bar=500µm).

The generated libraries were used to classify bacterial strains isolated from new lettuce heads bought from grocery stores. This validation showed that approximately 29.8% of the culturable bacterial population of romaine lettuce belong to bacterial genera which are not included in our current libraries, and these strains will be unknown for the technology. For a fair evaluation of the technology as a method of characterization of culturable bacterial communities from new samples, the unknown strains were not included in the validation process. Under these conditions, the sensitivity of the libraries applied to new samples ranged between 51.9-79.1%, and the classification accuracy per genus between 42.9%-100%. Overall, the sensitivity of the classification of culturable bacterial communities from romaine lettuce using the BARDOT technology with our current libraries is expected to be around 78.8% (Table 5.6).

Table 5.6. Percent of unknowns (bacterial colonies of genera not included in our libraries), sensitivity per library (% of colonies correctly classified at each time/library), and overall sensitivity of the classification using BARDOT technology when the developed libraries are applied to culturable bacterial communities from new lettuce samples.

Library	**Unknown**	**Validation excluding unknowns**		
		Number of BARDOT readings	**Sensitivity per library**	**Classification accuracy per genus**
A	20.7	93	79.1	83.2 ± 9.9
B	36.8	60	66.7	76.9 ± 9.1
C	30.8	26	51.9	41.0 ± 23.1
Overall	29.8	257	78.8	74.5 ± 7.4

5.5.4 Bacterial Genera with Overlapping Scatter Patterns

The broad scope of this environmental study allowed us to identify strains from different bacterial genera which under our experimental conditions produce very similar scatter patterns. The overlapped features among their scatter patterns represented a limitation for the current BARDOT technology.

The bacterial genera with non-differentiable scatter patterns were merged as one classifier in the respective library as shown in Table 5.4. One example in the library B is the classifier Curtobacterium/Microbacterium/Arthrobacter/Massilia, which is composed by similar scatter patterns of five bacterial strains (536Cu, 586Cu, 284Mi, 510Ar, 128Ma) belonging to four different genera. Figure 5.4A shows their morphology seen with a stereoscope, their colony profiles (height and diameter) measured through ICMA, their scatter patterns collected with BARDOT and their number of rings visualized through the 1-D cross-sections when their colonies have grown on PCA for 32 hours. Figures 5.4B, 5.4C and 5.4D show how the colony height, diameter and number of rings were significantly different among these five strains, for instance 284Mi and 510Ar exhibit colonies with lower colony height, smaller diameter and lower number of rings. However, their aspect ratios calculated as height/diameter were not significantly different (Figure 5.4E). Additionally, when these merged bacterial genera were tested as separate classifiers, the BARDOT technology was not able to find differentiable features among them since their classification error rates were up to 52.4% (Table 5.7).

Table 5.7. Average of height, diameter and aspect ratio of colonies grown for 32 hours from bacterial strains with overlapping scatter patterns (average ± SEM). Classification error (calculated as percent of colonies wrongly classified) when the merged bacterial genera are tested as separate classifiers to classify colonies growing in pure and mixed culture.

Classifier	**Strain**	**Average colony height (µm)**	**Average colony diameter (µm)**	**Average colony aspect ratio**	**Classification error (%) in pure culture**	**Classification error (%) in mixed culture**
Massilia	128Ma	103.8 ± 3.1	625.8 ± 29.9	0.17 ± 0.003	6.5	16.5
Microbacterium	284Mi	43.9 ± 2.0	302.4 ± 8.9	0.14 ± 0.003	50.9	51.5
Arthrobacter	510Ar	46.0 ± 3.4	277.4 ± 18.8	0.17 ± 0.002	10.8	52.4
Curtobacterium	536Cu	94.8 ± 1.9	588.8 ± 10.1	0.16 ± 0.001	14.1	34.2
	586Cu	120.5 ± 1.9	737.2 ± 13.0	0.16 ± 0.001		

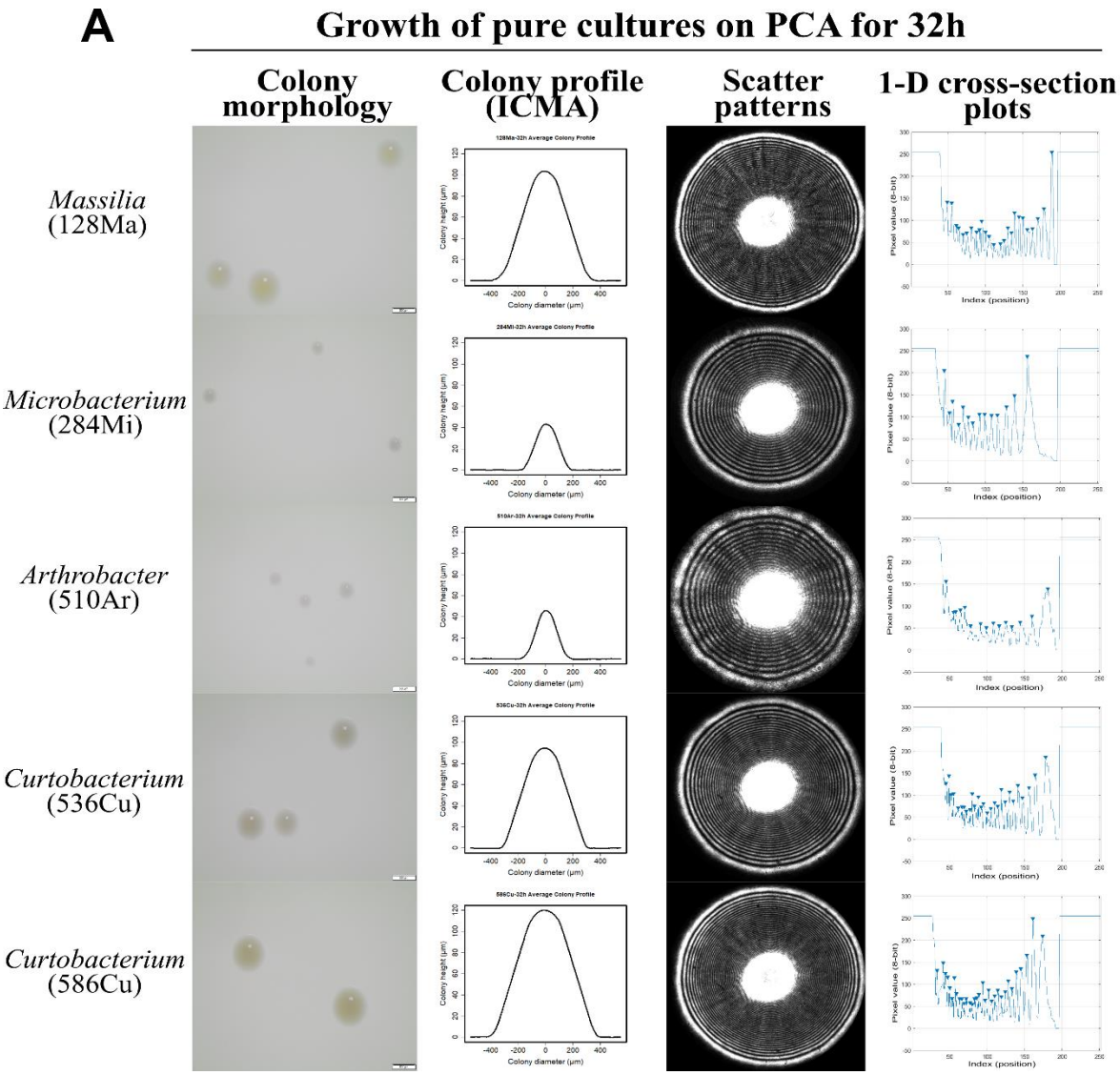


Figure 5.4. Analysis of colony morphology, colony profile (height and diameter), aspect ratio (calculated as height/diameter) and number of rings of bacterial strains of different genera with overlapping scatter patterns when grown for 32 hours on Plate Count Agar (PCA). A. Pictures of colony morphologies through a stereoscope (scale bar= 500µm). Colony profiles based on height and diameter measurements using a phase-contrast microscope with an integrated colony morphology analyzer (ICMA, the line graphs represent averages for 10-20 colonies per plate). Scatter patterns collected with BARDOT were used to plot the 1-D cross-sections to show the number of rings. B, C, D, E. Tukey's multiple comparisons to test differences in colony height, diameter, number of rings and aspect ratio, respectively.

Figure 5.4 continued

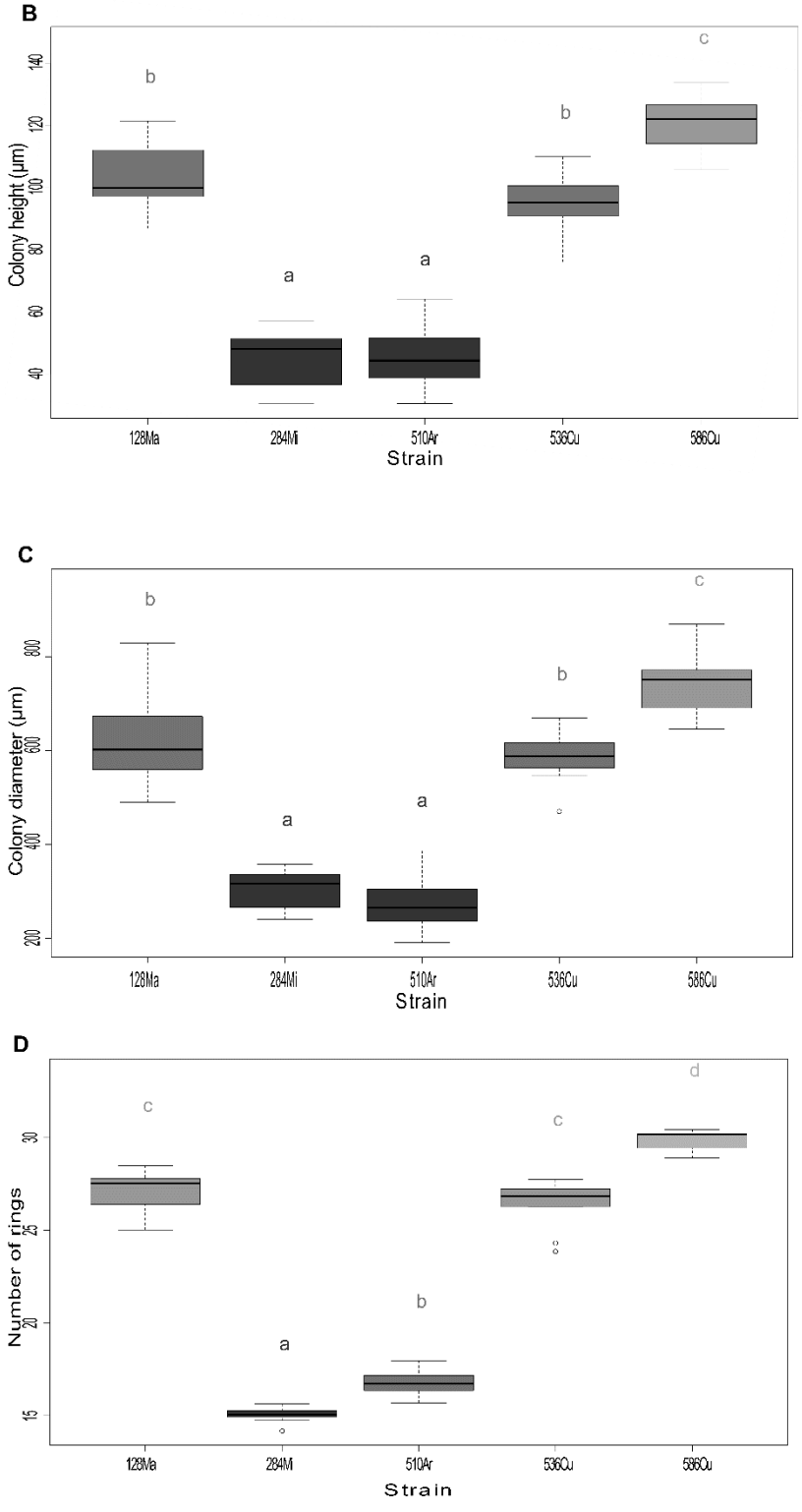
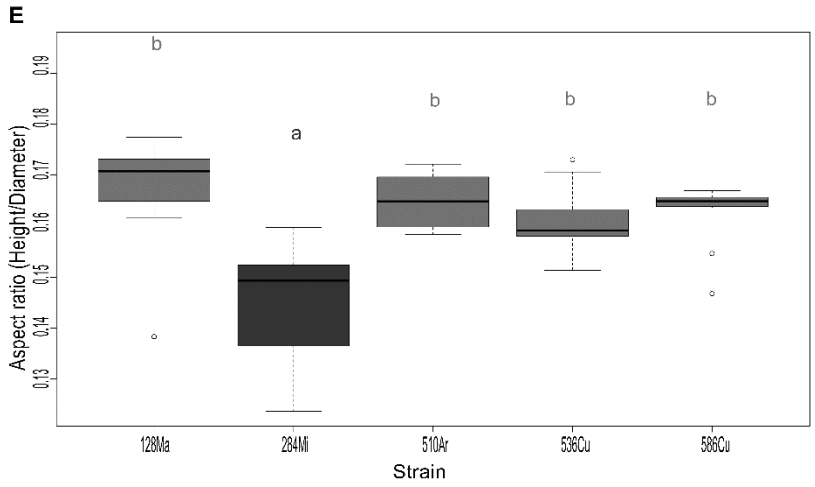


Figure 5.4 continued



5.6 Discussion

The study of native bacterial communities associated with fresh produce has become a promising area which could lead to development of strategies to control the colonization of plants by human pathogens (56, 59, 183). This study presented a comprehensive characterization of the culturable bacterial communities associated with organic and conventional romaine lettuce from grocery stores, since this is the leafy green commonly reported in foodborne disease outbreaks (109).

Previous studies have reported the composition of these communities from a group of different fresh produce items, which included romaine lettuce along with other leafy greens (68), fruits and vegetables (74). Our survey focused only on romaine lettuce heads which are readily available for purchase, besides the bacterial isolation was performed on plate-count agar which is the FDA standard medium for aerobic plate counts (185) with an incubation time of 72 hours. These parameters allowed fast- and slow-growing bacteria to be fully developed and identified, therefore the entire culturable bacterial communities of romaine lettuce growing under our conditions were characterized.

The total culturable bacterial population sizes from romaine lettuce from grocery stores were 4.89 ± 0.21 and 5.70 ± 0.22 log[CFU g^{-1}] (expressed as mean ± SE) for conventional and organic samples, respectively. Our culturable bacteria counts from ready-to-eat romaine lettuce were higher than those from field-collected romaine lettuce in California and Arizona (67); these higher counts were also observed in previous studies where leafy greens were collected from

grocery stores (68, 196, 197). Fresh produce available in supermarkets has gone through a large chain of distribution which not only exposes the product to new sources of bacteria, but also provides less harsh conditions of relative humidity, temperature, UV light, and water availability which affect bacterial growth and composition, when compared with open fields (76, 140, 198).

Comparison of organic and conventional labeled romaine lettuce revealed the farming system leading to divergences in the bacterial ecology of this leafy green. Significant differences were found in this study, where organically grown lettuce showed higher culturable bacterial population counts and higher bacterial richness when compared with conventional production ($p<0.01$). Similar results have been reported in organic-labeled spinach, lettuce and tomatoes purchased from grocery stores in the USA (74), in leafy greens from farmers markets in Brazil (199) and fresh lettuce from farm fields in Spain (200). However, the farming system was not a significant factor in leafy greens collected in farms in the mid-Atlantic region of the United States (159), and in packages of ready-to-eat leaf vegetables (68). The different results can be explained by variations in factors affecting the composition of microbial communities such as plant cultivars, geographical area, growing season, as well as factors related with the design of the surveys such as sample size, specific type or group of leafy greens/vegetables being tested, microbiological growth media, etc.

Regardless of the farming system, *Pseudomonadaceae* was the dominant family and it was present in all our samples of romaine lettuce. This is consistent with the numerous reports of *Pseudomonas* spp. associated with the phyllosphere and endosphere of leafy greens (66–68, 72, 196). On the other hand, some of the previous studies are consistent with our results of *Micrococcaceae* as the second-most-abundant family, while some others described *Enterobacteriaceae* being more predominant. However, all these studies may not be entirely comparable due to variations previously discussed.

The differences in the composition of organic and conventional culturable bacterial communities of romaine lettuce in our study were evident in the less-abundant bacterial genera (<10%). For that reason, it is important to survey a representative number of samples of the same produce type because there are differences in bacterial community diversity and composition across different produce types (74) which may mask the differences due to the farming system. Furthermore, a minimum incubation time of 72 hours is necessary to allow slow-growing bacteria to be included in the survey. According to our work, members of the *Pseudomonadaceae* and

Enterobacteriaceae families are fast growers and may be the only visible bacterial colonies on a plate incubated for only 24-48 hours. The methodological design of our study allowed us to identify a higher number of culturable bacterial genera than previously reported in leafy greens (66, 68). Moreover, eleven culturable representatives of *Massilia* were obtained, despite this genus being identified as a constituent of the phyllosphere but only through culture-independent approaches (67, 68, 140).

Interestingly, a higher taxonomic level, class Alphaproteobacteria, was exclusively isolated from organic samples. This class was represented by isolates of *Brevundimonas, Rhizobium* and *Sphingomonas,* which has been reported as plant-growing promoters (201–203). Therefore, our finding suggests that the application of biological products to replace the conventional fertilization might be driving the observed change of culturable bacterial community diversity and composition between our organic and conventional romaine lettuce. Additionally, this study showed a higher abundance of *Microbacterium* in organic lettuce leaves; this genus was previously found on field-grown and not laboratory-grown lettuce (73), suggesting that a relatively high abundance of *Microbacterium* might be associated with higher bacterial diversity as it was observed in our organic samples. On the other hand, the proportion of *Microbacterium* has also shown to be increased after the inoculation of *E. coli* O157:H7 onto lettuce leaves (73) and strains of this genus have exhibited antagonistic activity against this pathogen (183). Therefore, our findings might lead to a future valuable study about the associations of plants with high bacterial diversity such as organically grown lettuce, and the abundance of *Microbacterium* as a biocontrol agent against *E. coli* O157:H7.

This study is the first application of BARDOT technology to characterize entire culturable bacterial communities from environmental samples based on colony scatter patterns from different bacterial genera. The major challenge in this kind of broad application is the huge bacterial diversity found in the environment, in this case romaine lettuce. Our first question was how to represent all that diversity in our developed libraries. This was addressed by clustering our total culturable community into OTUs (97% similarity) and selecting the 20 most abundant OTUs which accounted for the nine most abundant bacterial genera representing 76.7% of our total population. Our methodology to define the coverage of the developed libraries corresponded with the real coverage observed when new bacterial colonies from new lettuce samples were classified (29.8% unknown colonies).

When dealing with entire bacterial communities, the variation in growth rates is also a challenge. Not all bacterial genera reached the optimal colony size to generate good scatter patterns at the same time, even not all the strains belonging to the same genus did. For that reason, this kind of application requires the development of libraries at different growth times. Additionally, we found that the scatter pattern of every strain changes through the time of incubation, and strains from the same genus can have totally different light scatter patterns under identical conditions of growth (Figure S2 in supplemental material). This effect of growth time on light scatter patterns has been also demonstrated in different serotypes of Shiga-toxin producing Escherichia coli (STEC) (85).

After training the system all classifiers exhibited PPV above 90% in the cross-validation matrices, however the validation tests with pure and mixed cultures allowed us to identify specific classifiers or strains which exhibited lower classification accuracy. Slight variations can be observed in pure cultures grown under the same conditions but on different experimental dates. For that reason, it is imperative to collect scatter patterns collected on different dates to be included in the libraries. On the other hand, some strains developed few colonies with new scatter patterns when they were grown in mixed cultures; it is known that bacteria signal, sense and respond to each other when they are part of a community (204–206). Recently, the effect of bacterial interactions happening in co-cultures was observed by the formation of flower-like patterns in colonies grown from mixtures of motile and non-motile bacterial species on a soft agar surface (207). Therefore, the light-scattering technology could be potentially applied to study microbial interactions modifying the spatiotemporal dynamics of colonies.

Finally, the validation of our libraries with new bacterial colonies isolated from new lettuce samples resulted in lower values of sensitivity and classification accuracy per genus, even when the colonies belonging to bacterial genera not included in the libraries were excluded. This result demonstrated that bacterial genera covered by our libraries can exhibit more variations of the scatter patterns and/or growth times when new strains are evaluated. For that reason, the development of BARDOT libraries to classify total bacterial communities from environmental samples should be conceived as a continuous process, where periodically new scatter patterns from new isolated strains are incorporated. Additionally, the sensitivity and classification accuracy of our BARDOT libraries can be increased by the ongoing improvement of the current built-in image analysis software (192). For instance, the implementation of new features such as calculation of

percent of similarity of the scatter patterns being tested against the patterns that are part of the classifiers. This information will help to set a threshold of similarity which might determine what colonies are unknown for the system (genera not included in the library) (191) or are not sufficiently similar to what is available in the libraries, therefore their classification would be uncertain.

Due to the broad scope of this study, strains from different bacterial genera producing very similar light scatter patterns were identified. This finding represents a limitation for the BARDOT technology as a tool for the characterization of entire culturable bacterial communities at the genus level. Our ICMA results showed that the overlapping strains had significantly different colony diameters and heights which consequently determine different number of rings (208), but their aspect ratios (Height/Diameter) were not significantly different. The aspect ratio of a colony has been demonstrated to determine the size of the light scatter pattern (195, 209), and this was demonstrated to be an important feature to provide differentiation during the classification process in our study. Therefore, our results corroborate the limitation of the single-wavelength laser scatter technology to differentiate bacterial strains or genera developing colonies with similar aspect ratios.

Finally, we suggest possible strategies to mitigate the limitation of overlapping scatter patterns. First, the application of the BARDOT technology as a fully automated system with an incubator, plate-handling robot and laser scanner will allow the scatter patterns to be recorded during the entire growth cycle. In that way, the classification process can be performed with multiple comparison points and with the growth rate as an additional feature. Additionally, the classification based on light scatter patterns might be complemented with phenotypic data, such as colony morphology, color, density and size produced under the same culturable conditions of the BARDOT readings. Last, some of the strains with overlapping scatter patterns exhibit different colony colors when they are fully developed, therefore using a multispectral forward-scatter instrument (210) instead of a single-wavelength laser might maximize their classification efficiency.

5.7 Acknowledgments

We have no conflicts of interest to disclose. This research was supported by the Center for Food Safety Engineering at Purdue University, funded by the U.S. Department of Agriculture, Agricultural Research Service. Any opinions, findings, conclusion, or recommendations expressed

in this publication are those of the author(s) and do not necessarily reflect the view of the U.S. Department of Agriculture.

5.8 Supplemental Material

Table S1. Ten-X cross-validation (CV) matrices of developed libraries. A. Library A (21-22 hours). B. Secondary library A (21-22 hours). C. Library B (32 hours). C. Library C (72 hours). Diagonal bold values show the positive predictive value (PPV) of each classifier.

A.

Ten-fold cross-validation (CV) matrix: Library A (21-22 hours)			
Classifiers	Pa	Er/Ps	Ma/Du/Fl/ Ar
Pa	**99.4**	0.6	0.1
Er/Ps	0.3	**98.3**	1.4
Ma/Du/Fl/Ar	0.4	3.0	**96.6**

B.

Ten-fold cross-validation (CV) matrix: Secondary Library A (21-22h) Erwinia/Pseudomonas		
Classifiers	Ps	Er
Ps	**95.4**	4.6
Er	2.4	**97.6**

Table S1 continued

C.

Ten-fold cross-validation (CV) matrix: Library B (32 hours)							
Classifiers	Ar	Du/Ar	Cu/Mi/ Ar/Ma	Ma	Fl	Ps	Er
Ar	**99.3**	0.3	0.3	0.0	0.0	0.0	0.0
Du/Ar	0.2	**99.8**	0.0	0.0	0.0	0.0	0.0
Cu/Mi/ Ar/Ma	0.0	0.0	**99.4**	0.0	0.1	0.4	0.0
Ma	0.2	1.6	0.0	**96.4**	0.5	1.2	0.0
Fl	0.0	0.0	1.2	0.0	**98.5**	0.3	0.0
Ps	0.0	0.0	1.9	0.3	1.7	**96.1**	0.0
Er	0.0	0.0	0.0	0.0	0.0	1.0	**99.0**

D.

Ten-fold cross-validation (CV) matrix: Library C (72 hours)					
Classifiers	Mi	Fl	Du	Ar	Mi/Ar
Mi	**92.6**	0.3	0.0	0.0	7.0
Fl	0.0	**99.4**	0.0	0.0	0.6
Du	0.8	0.0	**99.2**	0.0	0.0
Ar	0.0	0.0	0.7	**99.3**	0.0
Mi/Ar	8.5	0.3	0.2	0.4	**90.6**

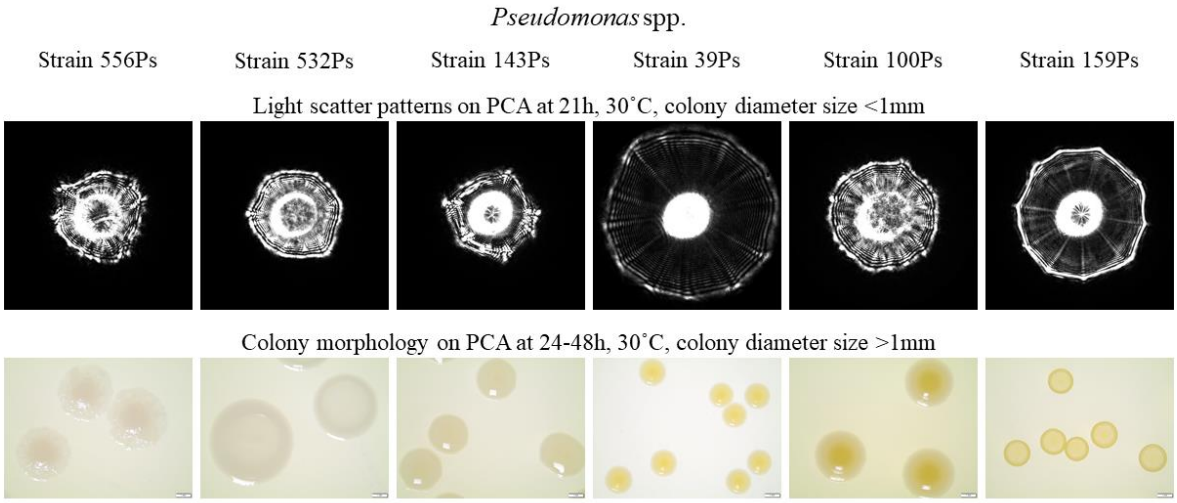


Figure S2. Comparison of light scatter patterns of different strains of Pseudomonas spp. grown under identical conditions: PCA medium, 30°C, 21 hours. Colony morphologies are showing fully developed colonies after their scatter patterns were recorded.

APPENDIX A. R CODE TO ANALYZE MISEQ 16S SEQUENCING DATA

```
### Used R 4.0.2 to perform this analysis
################## PROCESSING OF READS WITH DADA2#########################

# The following packages and their dependencies need to be installed and loaded to run the
analysis
library(dada2); packageVersion ("dada2")
library(reshape2); packageVersion ("reshape2")
library(magrittr); packageVersion ("magrittr")
library(data.table); packageVersion ("data.table")

# Define work directory
path <- "/scratch/snyder/d/dsarriaz/PoolEcoli/Original_Reads"
list.files(path)

# Forward and reverse fastq filenames have format: SAMPLENAME_R1_001.fastq and
#SAMPLENAME_R2_001.fastq
fnFs <- sort(list.files(path, pattern="_R1_001.fastq"))
fnRs <- sort(list.files(path, pattern="_R2_001.fastq"))
## Extract sample names, assuming filenames have format: SAMPLENAME_XXX.fastq
sample.names <- sapply(strsplit(fnFs, "_"), `[`, 1)
sample.names

fnFs <- file.path(path, fnFs)
fnRs <- file.path(path, fnRs)

#Error plots
pdf("/scratch/snyder/d/dsarriaz/PoolBacteria/Qualityplot_Forw1_PoolBacteria.pdf")
plotQualityProfile(fnFs[1:8])
dev.off()

##Filter and Trim
filt_path <- file.path(path, "Filtered")
filtFs <- file.path(filt_path, "Filtered", paste0(sample.names, "_F_filt.fastq"))
filtRs <- file.path(filt_path, "Filtered", paste0(sample.names, "_R_filt.fastq"))

## Filtering parameters: maxN=0 Dada2 requires no Ns, maximum of 2 expected errors per-read,
#Truncate reads at the first instance of a quality score less than or equal to truncQ
# vectors ( , ) for forward and reverse reads
# Truncate reads after truncLen bases. Reads shorter than this are discarded.
# trimleft: nucleotides to remove from the start of each read (primers). If both truncLen and
#trimLeft are provided, filtered reads will have length truncLen-trimLeft
out <- filterAndTrim (fnFs, filtFs, fnRs, filtRs, maxN=0, maxEE=c(2,2), truncQ=2,
                      trimLeft=c(19,17), truncLen=c(220,220), minLen=100, rm.phix=TRUE,
```

```
                compress=TRUE, multithread=TRUE)
write.csv(out, file="/scratch/snyder/d/dsarriaz/PoolEcoli/Summary_FilterAndTrim.csv")

## Learn the Error Rates
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)

pdf("/scratch/snyder/d/dsarriaz/PoolEcoli/plotErrorsF.pdf", width=12.25, height=10.0,
bg="white");
plotErrors(errF,nominalQ=TRUE)
dev.off()

pdf("/scratch/snyder/d/dsarriaz/PoolEcoli/plotErrorsR.pdf", width=12.25, height=10.0,
bg="white");
plotErrors(errR,nominalQ=TRUE)
dev.off()

##Dereplicate the filtered fastq files: combines all identical sequencing reads
## into "unique sequences" with a corresponding "abundance"
derepFs <- derepFastq(filtFs, verbose=TRUE) #unique sequences analysis
derepRs <- derepFastq(filtRs, verbose=TRUE) #unique sequences analysis
# Name the derep-class objects by the sample names
names(derepFs) <- sample.names
names(derepRs) <- sample.names

## Sample Inference: Infer the sequence variants in each sample
dadaFs <- dada(derepFs, err=errF, multithread=TRUE, pool=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE, pool=TRUE)
#Inspecting the returned dada-class object
dadaFs[[1]]

## Merge the denoised forward and reverse reads
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
# Inspect the merger data.frame from the first sample
head(mergers[[1]])
df<-data.frame(mergers = unlist(mergers))
write.csv(df, file= "/scratch/snyder/d/dsarriaz/PoolEcoli/PoolEcoli_merged.csv", row.names =
TRUE)

##construct a "sequence table" or ASV table, a higher-resolution version of the "OTU table"
#Rows are named by the samples, columns named by sequence variants
seqtab <- makeSequenceTable(mergers)
#Print number of samples and number of ASVs
asv <- dim(seqtab)
write.csv(asv,file="/scratch/snyder/d/dsarriaz/PoolEcoli/PoolEcoli_Samples-ASVs.csv",
row.names = FALSE)
```

```
#Length distribution of contigs
lengths <- table(nchar(getSequences(seqtab)))
write.csv(lengths,file="/scratch/snyder/d/dsarriaz/PoolEcoli/PoolEcoli_contiglengths.csv",
row.names = FALSE)

##Remove chimeric sequences
seqtab.nochim <- removeBimeraDenovo(seqtab, method="consensus", multithread=TRUE,
verbose=TRUE)
#Print number of samples and number of ASVs without chimeras
#Interpretation: calculate percent of chimeras/merged SVs
dim <- dim(seqtab.nochim)
write(dim, file = "/scratch/snyder/d/dsarriaz/PoolEcoli/SVs_nochimeras.txt")

##Abundance of ASVs without chimeras
##Interpretation: print percent of chimeric SVs/merged sequenced reads
sum <- sum(seqtab.nochim)/sum(seqtab)
write(sum, file = "/scratch/snyder/d/dsarriaz/PoolEcoli/percent_nochimeras.txt")

##Save ASV table (aka. OTU table)
write.csv(seqtab.nochim, file = "/scratch/snyder/d/dsarriaz/PoolEcoli/ASV-
Table_PoolEcoli.csv", row.names = TRUE, col.names = TRUE)

## As a final check of our progress, we'll look at the number of reads that made it through
#each step in the pipeline
getN <- function(x) sum(getUniques(x))
track <- cbind(out, sapply(dadaFs, getN), sapply(mergers, getN), rowSums(seqtab),
rowSums(seqtab.nochim))
colnames(track) <- c("input", "filtered", "denoisedF", "denoisedR", "merged", "nonchim")
rownames(track) <- sample.names
write.csv(track, file= "/scratch/snyder/d/dsarriaz/PoolEcoli/SummaryReads_PoolEcoli.csv",
row.names = TRUE)

##Assign Taxonomy
set.seed(100)
#Silva V.138
taxa.genus.50.silva <- assignTaxonomy(seqtab.nochim,
refFasta="/scratch/snyder/d/dsarriaz/PoolEcoli/Bacterial_databases/silva_nr99_v138_train_set.fa
.gz", multithread=TRUE, tryRC=TRUE, minBoot=50)
unname(head(taxa.genus.50.silva))
write.csv(taxa.genus.50.silva, file=
"/scratch/snyder/d/dsarriaz/PoolEcoli/PoolEcoli_genus_50_silva138.csv", row.names = TRUE)

taxa.species.50.silva <- addSpecies(taxa.genus.50.silva,
refFasta="/scratch/snyder/d/dsarriaz/PoolEcoli/Bacterial_databases/silva_species_assignment_v1
38.fa.gz", allowMultiple = TRUE)
unname(head(taxa.species.50.silva))
```

```
write.csv(taxa.species.50.silva, file=
"/scratch/snyder/d/dsarriaz/PoolEcoli/PoolEcoli_species_50_silva138.csv", row.names = TRUE)

##Save variables for Phyloseq
saveRDS(seqtab.nochim, "seqtab.nochim.Ecoli.rds")
saveRDS(taxa.genus.50.silva, "taxa.genus.50.silva.Ecoli.rds")
saveRDS(taxa.species.50.silva, "taxa.species.50.silva.Ecoli.rds")

####### DATA VISUALIZATION AND ANALYSIS WITH PHYLOSEQ ###############

# The following packages and their dependencies need to be installed and loaded to run the
analysis

library(reshape2);packageVersion ("reshape2")
library(magrittr);packageVersion ("magrittr")
library(ggplot2); packageVersion ("ggplot2")
library(RColorBrewer);packageVersion ("RColorBrewer")
library(data.table);packageVersion("data.table")
library(DECIPHER);packageVersion("DECIPHER")
library(decontam);packageVersion("decontam")
library(phangorn);packageVersion("phangorn")
library(phyloseq);packageVersion("phyloseq")
library(readr);packageVersion("readr")
library(dplyr);packageVersion("dplyr")
library(vegan);packageVersion("vegan")
library(ape);packageVersion("ape")
library(ggpubr);packageVersion("ggpubr")
library(tidyverse);packageVersion("tidyverse")
library(dada2); packageVersion ("dada2")
library(DESeq2); packageVersion ("DESeq2")
library(Rmisc); packageVersion ("Rmisc")
library(broom); packageVersion ("broom")
library(gplots); packageVersion ("gplots")
library(tidyr); packageVersion ("tidyr")
library(stringr); packageVersion ("stringr")
library(plotrix); packageVersion ("plotrix")
library(lme4); packageVersion ("lme4")
library(nlme); packageVersion ("nlme")
library(lsmeans); packageVersion ("lsmeans")
library(rcompanion); packageVersion ("rcompanion")
library(ggfortify); packageVersion ("ggfortify")
library(car); packageVersion ("car")
library(psych); packageVersion ("psych")
library(plyr); packageVersion ("plyr")

# Define work directory and source functions needed
```

```
setwd("//datadepot.rcac.purdue.edu/depot/pruittr/data/Documents_Vanessa/Thesis_IlluminaAnal
ysis/PoolBacteria/Phyloseq_Analysis")
source("miseqR.r", local = TRUE)

##Bring data from DADA2
seqtab.nochim <- readRDS("seqtab.nochim.Bacteria.rds")

##Taxonomy edited file
edited.taxa.genus <- read.table(file="edited1_PoolBacteria_genus_50_silva138.txt") #saved as
data frame
head(edited.taxa.genus)
#convert to matrix
edited.taxa.genus <- matrix(as(as(edited.taxa.genus, "matrix"), "character"),
                            nrow=nrow(edited.taxa.genus),
                            ncol=ncol(edited.taxa.genus))
row.names(edited.taxa.genus) <- colnames(seqtab.nochim)

##Metadata file: Do not use same word for two columns Ex. SampleID and Sample_Type
metadata.file <- "Metadata_PoolBacteria.csv"
metadata <- read.csv(metadata.file)
#convert dataframe into phyloseq format
metadata <- sample_data(metadata)
#required to merge metadata into phyloseq object: rownames must match sample names
rownames(metadata) <- metadata$SampleID

##Create Phyloseq object
psTotal <- phyloseq(otu_table(seqtab.nochim, taxa_are_rows=FALSE),
                    sample_data(metadata),
                    tax_table(edited.taxa.genus))
psTotal

colnames(tax_table(psTotal)) <- c("Domain", "Phylum", "Class", "Order", "Family", "Genus")
rank_names(psTotal)
sample_variables(psTotal)

#Reorder levels of variables if neede
sample_data(psTotal)$Study <- factor(sample_data(psTotal)$Study, levels = c("Bacteria",
"Control", "Mock", "EcoliStrain"))

############## TAXON FILTERING ###################

## REMOVE NON BACTERIAL SEQS
##Phylum NA or Unclassified, Order Chloroplast, Family Mitochondria
psTotal2 <- subset_taxa(psTotal, Domain == "Bacteria" &
                          !is.na(Phylum) &
                          !Phylum %in% c("", "Unclassified") &
```

```
                    !Order %in% c("Chloroplast") &
                    !Family %in% c("Mitochondria")
                  )
psTotal2
sum(otu_table(psTotal2)) #number of reads remaining
saveRDS(psTotal2, "psTotal2.rds")
saveRDS(psTotal, "psTotal.rds")

###Rename taxa names from entire sequence to only >SV1, >SV2, etc
new.names <- paste0("ASV", seq(ntaxa(psTotal2))) # Define new names ASV1, ASV2, ...
seqs <- taxa_names(psTotal2) #### Store sequences
names(seqs) <- new.names # Make map from ASV1 to full sequence
taxa_names(psTotal2) <- new.names # Now taxa names of Phyloseq object are are ASV1, ..
# Coerce to data.frame
seqs_df = as.data.frame(seqs)
write.csv(seqs_df, “ASV_sequences_PoolBacteria.csv")
saveRDS(seqs_df, "seqs_df.rds")


##### SAVE ASV TABLES ONLY BACTERIA BEFORE CONTAMINANT REMOVAL ##

# Extract abundance matrix from the phyloseq object
filtered_otutable = as(otu_table(psTotal2), 'matrix')
# transpose if necessary
if(taxa_are_rows(psTotal2)){filtered_otutable <- t(filtered_otutable)}
# Coerce to data.frame
filtered_otutable_df = as.data.frame(filtered_otutable)
# OTU table with entire sequence of ASVs as columns and samples as rows, showing abundance
write.csv(filtered_otutable_df, “ASV_abundances.csv")

# Extract taxonomy matrix from the phyloseq object
filtered_taxtable = as(tax_table(psTotal2), 'matrix')
# Taxonomy table with taxonomy levels as columns and entire sequence of ASVs as rows
write.csv(filtered_taxtable, "ASV_taxonomy.csv")


########### CONTAMINANT REMOVAL FROM CONTROLS ##################

#Remove contaminant seqs with decontam, using PCR negative controls.
#Code adapted from https://benjjneb.github.io/decontam/vignettes/decontam_intro.html

#Library size as a function of whether its a negative control or not
df <- as.data.frame(sample_data(psTotal2)) # Put sample_data into a ggplot-friendly data.frame
df$LibrarySize <- sample_sums(psTotal2)
df <- df[order(df$LibrarySize),]
df$Index <- seq(nrow(df))
```

```
ggplot(data=df, aes(x=Index, y=LibrarySize, color=Study)) + geom_point()

##Prevalence method to identify contaminants: presence/absence in samples against controls
#must remove sample with 0 reads, then code will run.
psTotal2.rm0 <- prune_samples(sample_sums(psTotal2)>=1, psTotal2)
sample_data(psTotal2.rm0)$is.neg <- sample_data(psTotal2.rm0)$Study == "Control"
#Default threshold for statistical test is 0.1
contamdf.prev <- isContaminant(psTotal2.rm0, method="prevalence", neg="is.neg",
threshold=0.1)
table(contamdf.prev$contaminant)
head(which(contamdf.prev$contaminant)) ##Print the rank of abundance of the SVs identified as
contaminants
write.csv(contamdf.prev, "contaminants.csv")

#plot contaminants based on number of control samples they appear in
# Make phyloseq object of presence-absence in negative controls
ps.control <- prune_samples(sample_data(psTotal2.rm0)$Study == "Control", psTotal2.rm0)
ps.control.presence <- transform_sample_counts(ps.control, function(abund) 1*(abund>0))
# Make phyloseq object of presence-absence in true positive samples
ps.sample <- prune_samples(sample_data(psTotal2.rm0)$Study != "Control", psTotal2.rm0)
ps.sample.presence <- transform_sample_counts(ps.sample, function(abund) 1*(abund>0))
# Make data.frame of prevalence in positive and negative samples
df.pres <- data.frame(prevalence.sample=taxa_sums(ps.sample.presence),
prevalence.control=taxa_sums(ps.control.presence),
                      contam.prev=contamdf.prev$contaminant)
write.csv(df.pres, "contaminants_presence.csv")
## Check if identified contaminants (TRUE) are more prevalent in samples or controls)
ggplot(data=df.pres, aes(x=prevalence.control, y=prevalence.sample, color=contam.prev)) +
geom_point()

#What is in controls?
ps.control = prune_taxa(taxa_sums(ps.control) > 0, ps.control)
Top.controls = names(sort(taxa_sums(ps.control), TRUE)[1:20])
ps.control.Top = prune_taxa(Top.controls, ps.control)
plot_bar(ps.control.Top, x="Type", fill = "Genus")

# What is in samples?
#ps.sample <- prune_samples(sample_data(psTotal2.rm0)$Type == "Plant", psTotal2.rm0) To
check abundances in plant samples
ps.sample = prune_taxa(taxa_sums(ps.sample) > 0, ps.sample)
Top.samples = names(sort(taxa_sums(ps.sample), TRUE)[1:20])
ps.sample.Top = prune_taxa(Top.samples, ps.sample)
plot_bar(ps.sample.Top, x="Distribution", fill = "Genus", facet_grid = ~RH)

#Filter identified contaminants with prevalence method out of dataset
keep <- !contamdf.prev$contaminant
```

```
psTotal2.free <- prune_taxa(keep, psTotal2.rm0)

#Other method: Filter out the most abundant identified contaminant ASVs
ps.control = prune_taxa(taxa_sums(ps.control) > 0, ps.control)
Top.controls = names(sort(taxa_sums(ps.control), TRUE)[1:8]) ##Only remove Top8 of controls
taxa.no.top.controls <- setdiff(taxa_names(psTotal2.rm0), Top.controls)  ##Keeping taxa no
present in Top control
psTotal2.free = prune_taxa(taxa.no.top.controls, psTotal2.rm0) ##New phyloseq object with NO
#top controls
psTotal2.free
sum(otu_table(psTotal2.free))


#how many reads removed?
sum(otu_table(psTotal2.rm0)) #before filtering
sum(otu_table(psTotal2.free)) #after filtering

#Remove control samples
psTotal2.free <- subset_samples(psTotal2.free, Study != "Control")
psTotal2.free <- prune_taxa(taxa_sums(psTotal2.free) > 0, psTotal2.free)
psTotal2.free
sum(otu_table(psTotal2.free))

#Remove E.coli strain and Mock controls
pBacteria <- subset_samples(psTotal2.free, Type != "EcoliStrain") #125 samples
pBacteria <- subset_samples(pBacteria, Study != "Mock") # 123 samples
pBacteria <- prune_taxa(taxa_sums(pBacteria) > 0, pBacteria)
pBacteria
sum(otu_table(pBacteria))
get_taxa_unique(pBacteria, "Genus")


############## CHECK MOCK CONTROLS ######################################

pMock <- subset_samples(psTotal2.free, Study == "Mock")
pMock <- prune_taxa(taxa_sums(pMock) > 0, pMock)

pMock_genus <- pMock %>%
  tax_glom(taxrank = "Genus") %>%                    # agglomerate at phylum level
  transform_sample_counts(function(x) {x/sum(x)} ) %>% # Transform to rel. abundance
  psmelt() %>%                                       # Melt to long format
  filter(Abundance > 0.01) %>%                       # Filter out low abundance taxa
  arrange(Genus)                                     # Sort data frame alphabetically by phylum

# Plot
ggplot(pMock_genus, aes(x = Type, y = Abundance*100, fill = Genus)) +
```

```
  geom_bar(stat = "identity") +
  scale_fill_brewer(palette="Dark2") +
  geom_text(aes(label=format(Abundance*100, digits = 2)), position = position_stack(vjust =
0.5)) +
  # Remove x axis title
  theme(axis.title.x = element_blank()) +
  #
  guides(fill = guide_legend(reverse = FALSE, keywidth = 1, keyheight = 1)) +
  ylab("Relative Abundance (Genera > 1%) \n") +
  ggtitle("Genera Composition of Mock Controls")

### FILTER LOW ABUNDANCE TAXA AND LOW SAMPLE SIZES ############

#Remove low abundance taxa: not seen more than 2 times in at least 10% of the samples
pBacteria.f = filter_taxa(pBacteria, function(x) sum(x > 2) > (0.1*length(x)), TRUE)
pBacteria.f  ## 1516 ASV, 123 samples, 417 genera
get_taxa_unique(pBacteria.f, "Genus")
sum(otu_table(pBacteria.f)) #1,814,947

#Plot number of reads
sdt2 = data.table(as(sample_data(pBacteria.f), "data.frame"),
                  TotalReads = sample_sums(pBacteria.f), keep.rownames = TRUE)
write.csv(sdt2, "TotalReads_pBacteria.f.csv")

pSeqDepth = ggplot(sdt2, aes(x = Condition, y = TotalReads, fill=Cultivar)) +
  ggtitle("Library sizes") + xlab("Experimental Conditions") + ylab("Number of Reads") +
  scale_x_discrete(guide = guide_axis(angle = 90)) + labs(fill="Cultivar") +
  theme(plot.title = element_text(hjust = 0.5), plot.subtitle=element_text(hjust=0.5,
face="italic")) + geom_boxplot() + scale_y_log10()
pSeqDepth
pSeqDepth + facet_grid(~RH, scales="free")

#Remove samples with fewer than 1000 reads
pBacteria.r <- prune_samples(sample_sums(pBacteria.f)>=1000, pBacteria.f)
sum(otu_table(pBacteria.r)) # Total of seqs: 1,813,464, 1516 ASV, 121 samples
mean(rowSums(otu_table(pBacteria.r))) # average of seqs:14,987

#Which samples were removed?
removed.samples <- setdiff(sample_names(pBacteria), sample_names(pBacteria.r))
removed.samples # "TH3RH65Upr1" "TH3RH65Upr2"

saveRDS(pBacteria.r, "pBacteria-r-before_tree.rds")
```

```
########## PHYLOGENETIC TREE ############################

##The DADA2 sequence inference method is reference-free, so we must construct the
##phylogenetic tree relating the inferred sequence variants de novo.
##We begin by performing a multiple-alignment using the DECIPHER R package

#Get sequences back from seqs_df, only the ASVs present in pBacteria.r
seqs.pBacteria.r <- data.frame(seqs_df[row.names(seqs_df) %in% taxa_names(pBacteria.r),],
row.names = taxa_names(pBacteria.r))
write.csv(seqs.pBacteria.r, "ASV_sequences_pBacteria-r.csv")

final.seqs <- seqs.pBacteria.r[,1]
taxa_names(pBacteria.r) <- final.seqs  #rename ASV again with sequences to add the
#phylogenetic tree
names(final.seqs) <- final.seqs # This propagates to the tip labels of the tree
alignment <- AlignSeqs(DNAStringSet(final.seqs), anchor=NA)

##first construct a neighbor-joining tree, and then fit a
#GTR+G+I (Generalized time-reversible with Gamma rate variation) maximum likelihood tree
using the neighbor-joining tree as a starting point.
phang.align <- phyDat(as(alignment, "matrix"), type="DNA")
dm <- dist.ml(phang.align)
treeNJ <- NJ(dm) # Note, tip order != sequence order
fit = pml(treeNJ, data=phang.align) ## negative edges length changed to 0!
fitGTR <- update(fit, k=4, inv=0.2)
fitGTR <- optim.pml(fitGTR, model="GTR",  optInv=TRUE, optGamma=TRUE,rearrangement
= "stochastic", control = pml.control(trace = 0,epsilon= 1e-08 ,maxit=10))
saveRDS(fitGTR, "fitGTR.rds")
#fitGTR <- readRDS("fitGTR.rds")

##Add tree to phyloseq object
phy_tree(pBacteria.r) <- phy_tree(fitGTR$tree)  ##Add tree to phyloseq object
saveRDS(pBacteria.r, "pBacteria-r.rds")
tree <- phy_tree(fitGTR$tree)
write.tree(tree, "pBacteria-tree.newick")

############## SAVE ASV TABLES OF FINAL PROCESSED DATA #########

#Change name of ASVs
pBacteria.asv <- pBacteria.r
taxa_names(pBacteria.asv) <- paste0("ASV", seq(ntaxa(pBacteria.asv)))
### New ASV numbers NOT related with previous

# ASV table raw counts: ASVs as columns and samples as rows, showing abundance
write.csv(t(otu_table(pBacteria.asv)), "ASV_Table_pBacteria-asv.csv")
# Sample Data: Metadata
```

```
write.csv(sample_data(pBacteria.asv), "Metadata_Table_pBacteria-asv.csv")
# Taxonomy table: taxonomy levels as columns and ASVs as rows
write.csv(tax_table(pBacteria.asv), "Taxonomy_Table_pBacteria-asv.csv")
#Save Tree with ASV as taxa names
tree.asv <- phy_tree(pBacteria.asv)
write.tree(tree.asv, "pBacteria-tree-asv.newick")

saveRDS(pBacteria.asv, "pBacteria.asv-final.rds")

### FIG. STACKED BARPLOT OF RELATIVE ABUNDANCES AT CLASS LEVEL ####
## Stats
get_taxa_unique(pBacteria.asv, "Phylum")
get_taxa_unique(pBacteria.asv, "Class")
get_taxa_unique(pBacteria.asv, "Order")
get_taxa_unique(pBacteria.asv, "Family")
get_taxa_unique(pBacteria.asv, "Genus")

# Agglomerate at Class level
Ec.phyl <- tax_glom(pBacteria.asv, taxrank="Class")
# Transform to rel. abundance
Ec.phyl.ra <- transform_sample_counts(Ec.phyl, function(x) {x/sum(x)} )

map <- sample_data(Ec.phyl.ra)
phyla <- as.vector(tax_table(Ec.phyl)[,3]) #change column number if evaluating other tax rank
#Create Phyla relative abundance table for all samples
PhylOTU <- as.data.frame(otu_table(Ec.phyl.ra))
colnames(PhylOTU) <- phyla

##sort by sample ID (so that they are in consecutive order)
#PhylOTU <- PhylOTU[order(rownames(PhylOTU)),]

#combine phyla that contribute less than 1% of rel ab
below01=PhylOTU[,colMeans(PhylOTU)<0.01]
# Sum rel ab of all phyla below 1%
below01.cs=rowSums(below01)

#remove those below01 phyla from the table
PhylOTU=PhylOTU[,colMeans(PhylOTU)>0.01]
#add the sum from all phyla below 1%
PhylOTU=cbind(PhylOTU,below01.cs)

#rename the last row
colnames(PhylOTU)[ncol(PhylOTU)]="Below_1%"

#Merge with map file
phylOTU.merge <- merge(PhylOTU, map, by="row.names")
```

```
#Average each by variable of interest, in this case, Condition
phyl.avg <- aggregate(phylOTU.merge[,2:11], by=list(phylOTU.merge$CultivarCondition),
mean) #change number of columns including the taxa ranks

rownames(phyl.avg) <- phyl.avg$Group.1
colnames(phyl.avg)[[1]] <- "CultivarCondition"

phyl.avg.melt <- phyl.avg %>%
  melt()

colnames(phyl.avg.melt)[[2]] <- "Class"

colourCount = length(unique(phyl.avg.melt$Class))
getPalette = colorRampPalette(brewer.pal(colourCount, "Dark2"))

pdf("Class_stackedbarplot_ByCultivarCondition.pdf",width=7,height=4)
ggplot(phyl.avg.melt) +
  geom_bar(stat = "identity", aes(x = CultivarCondition, y = value, fill = Class)) +
  scale_fill_manual(values = getPalette(colourCount)) +
  theme(axis.title.x = element_blank()) +
  guides(fill = guide_legend(reverse = FALSE, keywidth = 1, keyheight = 1)) +
  ylab("Relative Abundance \n") +
  theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5), axis.text = element_text(size
=10), legend.text = element_text(size =10))
dev.off()

phyl.com.summary <- phyl.avg.melt %>%
  dplyr::group_by(CultivarCondition, Class) %>%
  dplyr::summarize(mean = mean(value))
write.csv(phyl.com.summary, "Relative_abundances_Class.csv")

### DIFFERENTIAL ABUNDANCE BETWEEN TREATMENTS AT FAMILY LEVEL ###

# Remove samples no wanted in analysis
Ec.growth <- subset_samples(pBacteria.asv, Cultivar == "TH")

#agglomerate to tax rank
Ec.fam<- tax_glom(pBacteria.asv, "Family")
Ec.gen<- tax_glom(Ec.growth, "Genus")
#converts phyloseq-format data into a `DESeqDataSet` with dispersions estimated, using the
#experimental design formula (~ Variable)
diagdds = phyloseq_to_deseq2(Ec.fam, ~ RH)
#calculate geometric means prior to estimate size factors
gm_mean = function(x, na.rm=TRUE){
  exp(sum(log(x[x > 0]), na.rm=na.rm) / length(x))
```

```
}
geoMeans = apply(counts(diagdds), 1, gm_mean)
diagdds = estimateSizeFactors(diagdds, geoMeans = geoMeans)
diagdds.fam = DESeq(diagdds, fitType="local") #fit dispersion by local data
#Note: The default multiple-inference correction is Benjamini-Hochberg, and occurs within the
#`DESeq` function.

# Create table of results with function results
alpha = 0.001
log2 = 4
RtoE <- results(diagdds.fam, contrast=c("RH", "B", "C")) #first level will be used for plotting
#the fold changes
RtoE = RtoE[order(RtoE$padj, na.last=NA), ]
sigtabRtoE = RtoE[(RtoE$padj < alpha), ]
sigtabRtoE = sigtabRtoE[(abs(sigtabRtoE$log2FoldChange) > log2), ]
sigtabRtoE = cbind(as(sigtabRtoE, "data.frame"),
as(tax_table(Ec.fam)[rownames(sigtabRtoE), ], "matrix"))
Comparison <- rep("RHBtoRHC", length(rownames(sigtabRtoE)))
sigtabRtoE <- cbind(sigtabRtoE, Comparison)
write.csv(sigtabRtoE, "DeSeq_Family_ALL-RHBvsRHC_p0.001_log4.csv")

####Visualize differential abundance
# Class order
x = tapply(sigtabRtoE$log2FoldChange, sigtabRtoE$Class, function(x) max(x))
x = sort(x, TRUE)
sigtabRtoE$Class = factor(as.character(sigtabRtoE$Class), levels=names(x))

# Family order
x = tapply(sigtabRtoE$log2FoldChange, sigtabRtoE$Family, function(x) max(x))
x = sort(x, TRUE)
sigtabRtoE$Family = factor(as.character(sigtabRtoE$Family), levels=names(x))

# Genus order
x = tapply(sigtabRtoE$log2FoldChange, sigtabRtoE$Genus, function(x) max(x))
x = sort(x, TRUE)
sigtabRtoE$Genus = factor(as.character(sigtabRtoE$Genus), levels=names(x))

colourCount = length(unique(sigtabRtoE$Class))
getPalette = colorRampPalette(brewer.pal(8, "Spectral"))

pdf("DeSeq_Family_ALL-RHBvsRHC_p0.001_log4.pdf",width=7,height=3.5)
ggplot(sigtabRtoE, aes(x=Family, y=log2FoldChange, fill = Class)) +
  geom_point(size=3, colour= "black", pch = 21, position = position_jitterdodge()) +
    scale_fill_manual(values = getPalette(colourCount)) +
  theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5, size = 10, face="italic"),
axis.text.y=element_text(size = 10))
```

```
dev.off()

######## ALPHA DIVERSITY STATISTICAL ANALYSIS ############################

#alpha diversity calculations adapted from Denef lab tutorial.
#http://deneflab.github.io/MicrobeMiseq/
min_lib <- min(sample_sums(pBacteria.asv))

# Initialize matrices to store richness and evenness estimates
nsamp = nsamples(pBacteria.asv)
trials = 100

richness <- matrix(nrow = nsamp, ncol = trials)
row.names(richness) <- sample_names(pBacteria.asv)

shannon <- matrix(nrow = nsamp, ncol = trials)
row.names(shannon) <- sample_names(pBacteria.asv)

for (i in 1:100) {
  # Subsample
  r <- rarefy_even_depth(pBacteria.asv, sample.size = min_lib, verbose = FALSE, replace = T)

  # Calculate richness
  rich <- as.numeric(as.matrix(estimate_richness(r, measures = "Observed")))
  richness[ ,i] <- rich

  # calculate shannon diversity
  shan <- as.numeric(as.matrix(estimate_richness(r, measures = "Shannon")))
  shannon[ ,i] <- shan
}

# Create a new dataframe to hold the means and standard deviations of richness estimates
SampleID <- row.names(richness)
mean <- apply(richness, 1, mean)
sd <- apply(richness, 1, sd)
measure <- rep("Richness", nsamp)
rich_stats <- data.frame(SampleID, mean, sd, measure)

# Create a new dataframe to hold the means and standard deviations of shannon diversity
estimates
SampleID <- row.names(shannon)
mean <- apply(shannon, 1, mean)
sd <- apply(shannon, 1, sd)
measure <- rep("Shannon Diversity", nsamp)
shan_stats <- data.frame(SampleID, mean, sd, measure)
```

```
#combine richness and evenness into one dataframe
alpha <- rbind(rich_stats, shan_stats)

#adds sample data
s <- data.frame(sample_data(pBacteria.asv))
alphadiv <- merge(alpha, s, by = "SampleID")
write.csv(alphadiv, "Alpha_diversity_pBacteria-asv.csv")

alphadiv.shan <- subset(alphadiv, measure == "Shannon Diversity")
alphadiv.shan <- alphadiv.shan[order(alphadiv.shan$mean), ]
alphadiv.rich <- subset(alphadiv, measure == "Richness")
alphadiv.rich <- alphadiv.rich[order(alphadiv.rich$mean), ]

colourCount = 3 #number of elements to be colored (ex. 3 RH)
getPalette = colorRampPalette(brewer.pal(3, "Dark2"))

#plot alpha diversity
pdf("Alpha diversity Richness.pdf",width=6.5,height=4)
#ggplot(alphadiv.rich, aes(x = CultivarCondition, y = mean)) + geom_boxplot() +
ylab("Richness") + xlab("E.coli growth level")
ggplot(alphadiv.rich, aes(x = CultivarCondition, y = mean, color = Distribution)) +
  geom_boxplot() +
  facet_grid(. ~ RH, scales = "free", space = "free_x") +
  #geom_point(aes(fill = Distribution), size = 1, shape = 21, position = position_jitterdodge()) +
  scale_color_manual(values = getPalette(colourCount)) +
  scale_fill_manual(values = getPalette(colourCount)) +
  theme(axis.text.x = element_text(angle = 90, size = 10), legend.position ="bottom") +
  ylab("Richness") + xlab("Experimental conditions")
dev.off()

pdf("Alpha diversity Shannon.pdf",width=6.5,height=4)
ggplot(alphadiv.shan, aes(x = CultivarCondition, y = mean, color = Distribution)) +
  geom_boxplot() +
  facet_grid(. ~ RH, scales = "free", space = "free_x") +
  #geom_point(aes(fill = EcoliGrowth), size = 1, shape = 21, position = position_jitterdodge()) +
  scale_color_manual(values = getPalette(colourCount)) +
  scale_fill_manual(values = getPalette(colourCount)) +
  theme(axis.text.x = element_text(angle = 90, size = 10), legend.position ="bottom") +
  ylab("Shannon Diversity") + xlab("Experimental conditions")
dev.off()

################### BETA DIVERSITY ################################
set.seed(3)
########## Subset samples if needed
GT <- subset_samples(pBacteria.asv, Cultivar == "GT")
RHA <- subset_samples(pBacteria.asv, RH == "A")
```

```
#####Unconstrained Ordination######
# Scale reads to even depth
pBacteria_scale <- RHC %>%
  scale_reads(round = "round")
#Smallest sample size in dataset (it is the default in scale_reads if n is not given)
min(sample_sums(pBacteria.asv))

# Ordinate
pBacteria_scale_pcoa <- ordinate(
  physeq = pBacteria_scale,
  method = "PCoA",
  distance = "bray"
)

# Plot for all RH together
plot_ordination(
  physeq = pBacteria_scale,
  ordination = pBacteria_scale_pcoa,
  color = "RH",
  shape = "Distribution",
  #label = "EcoliGrowth",
  title = "PCoA-Bray of Romaine lettuce bacterial communities"
) + geom_point(aes(color = RH), alpha = 0.7, size = 4) +
  geom_point(colour = "grey90", size = 1.5)  +
  geom_text(aes(label= Cultivar), size = 3, vjust = 1.5) +
  labs(shape= "Distribution")


# Plot for each RH separate
plot_ordination(
  physeq = pBacteria_scale,
  ordination = pBacteria_scale_pcoa,
  color = "Distribution",
  shape = "Cultivar",
  #label = "Plant",
  title = "PCoA-Bray of RH-C",
) + geom_point(aes(color = Distribution), alpha = 0.7, size = 4) +
  geom_point(colour = "grey90", size = 1.5) +
  geom_text(aes(label= Plant), size = 3, vjust = 1.5)

# Ordinate
pBacteria_scale_pcoa <- ordinate(
  physeq = pBacteria_scale,
  method = "PCoA",
  distance = "unifrac"  ##Needs phylogenetic tree!
)
```

```
# Plot for all RH
plot_ordination(
  physeq = pBacteria_scale,
  ordination = pBacteria_scale_pcoa,
  color = "RH",
  shape = "Distribution",
  #label = "SampleID",
  title = "PCoA-Unifrac of all"
) + geom_point(aes(color = RH), alpha = 0.7, size = 4) +
  geom_point(colour = "grey90", size = 1.5)  +
  #geom_text(aes(label= EcoliGrowth), size = 3, vjust = 1) +
  #geom_text(aes(label= Cultivar), size = 3, vjust = 1.5) +
  labs(shape= "Distribution")

# Plot for RH separate
plot_ordination(
  physeq = pBacteria_scale,
  ordination = pBacteria_scale_pcoa,
  color = "Distribution",
  shape = "Cultivar",
  #label = "SampleID",
  title = "PCoA-Unifrac of RHC"
) + geom_point(aes(color = Distribution), alpha = 0.7, size = 4) +
  geom_point(colour = "grey90", size = 1.5)  +
  #geom_text(aes(label= Plant), size = 3, vjust = 1.5)
  #labs(color= "Experimental conditions", shape = "Ecoli growth")

#Permanova: use scaled data
# Calculate unifrac distance matrix
pBacteria.nc <- subset_samples(pBacteria_scale, Distribution != "Plant")
unif <- phyloseq::distance(pBacteria.nc, method = "unifrac")
bray <- phyloseq::distance(pBacteria.nc, method = "bray")
wunif <- phyloseq::distance(pBacteria.nc, method = "wunifrac")

# make a data frame from the sample_data
sampledf <- data.frame(sample_data(pBacteria.nc))

# Adonis test
set.seed(3)
adonis(unif ~ Cultivar*Distribution, data = sampledf)
adonis(bray ~ Cultivar*Distribution, data = sampledf)
adonis(wunif ~ RH*Distribution, data = sampledf)
```

REFERENCES

1. Turner TR, James EK, Poole PS. 2013. The plant microbiome. Genome Biol 14:209.
2. Bulgarelli D, Schlaeppi K, Spaepen S, Ver Loren van Themaat E, Schulze-Lefert P. 2013. Structure and functions of the bacterial microbiota of plants. Annu Rev Plant Biol 64:807–838.
3. Vorholt JA. 2012. Microbial life in the phyllosphere. Nat Rev Microbiol 10:828–840.
4. Scallan E, Griffin PM, Angulo FJ, Tauxe RV, Hoekstra RM. 2011. Foodborne illness acquired in the United States--unspecified agents. Emerg Infect Dis 17:16–22.
5. Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson M-A, Roy SL, Jones JL, Griffin PM. 2011. Foodborne illness acquired in the United States--major pathogens. Emerg Infect Dis 17:7–15.
6. Dewey-Mattia D, Manikonda K, Hall AJ, Wise ME, Crowe SJ. 2018. Surveillance for foodborne disease outbreaks — United States, 2009–2015. Morb Mortal Wkly Rep Surveill Summ 67:1–11.
7. 2013. Centers for Disease Control and Prevention, p. 368–368. *In* Gellman, MD, Rick Turner, J (eds.), Encyclopedia of Behavioral Medicine. Springer New York.
8. Painter JA, Hoekstra RM, Ayers T, Tauxe RV, Braden CR, Angulo FJ, Griffin PM. 2013. Attribution of foodborne illnesses, hospitalizations, and deaths to food commodities by using outbreak data, United States, 1998-2008. Emerg Infect Dis 19:407–415.
9. Callejón RM, Rodríguez-Naranjo MI, Ubeda C, Hornedo-Ortega R, Garcia-Parrilla MC, Troncoso AM. 2015. Reported foodborne outbreaks due to fresh produce in the United States and European Union: trends and causes. Foodborne Pathog Dis 12:32–38.
10. Lynch MF, Tauxe RV, Hedberg CW. 2009. The Growing Burden of Foodborne Outbreaks Due to Contaminated Fresh Produce: Risks and Opportunities. Epidemiol Infect 137:307–315.
11. Brandl MT, Sundin GW. 2013. Focus on food safety: human pathogens on plants. Phytopathology 103:304–305.
12. Usda N. 2018. Vegetables 2017 summary. USDA, National Agricultural Statistics Service Washington, DC.

13. Barak JD, Schroeder BK. 2012. Interrelationships of food safety and plant pathology: the life cycle of human pathogens on plants. Annu Rev Phytopathol 50:241–266.
14. Brandl MT. 2006. Fitness of human enteric pathogens on plants and implications for food safety. Annu Rev Phytopathol 44:367–392.
15. Olaimat AN, Holley RA. 2012. Factors influencing the microbial safety of fresh produce: a review. Food Microbiol 32:1–19.
16. Lim JY, Yoon J, Hovde CJ. 2010. A brief overview of Escherichia coli O157:H7 and its plasmid O157. J Microbiol Biotechnol 20:5–14.
17. Riley LW, Remis RS, Helgerson SD, McGee HB, Wells JG, Davis BR, Hebert RJ, Olcott ES, Johnson LM, Hargrett NT, Blake PA, Cohen ML. 1983. Hemorrhagic colitis associated with a rare Escherichia coli serotype. N Engl J Med 308:681–685.
18. Centers for Disease Control and Prevention (CDC). 1993. Update: multistate outbreak of Escherichia coli O157:H7 infections from hamburgers--western United States, 1992-1993. MMWR Morb Mortal Wkly Rep 42:258–263.
19. Hilborn ED, Mermin JH, Mshar PA, Hadler JL, Voetsch A, Wojtkunski C, Swartz M, Mshar R, Lambert-Fair MA, Farrar JA, Glynn MK, Slutsker L. 1999. A multistate outbreak of Escherichia coli O157:H7 infections associated with consumption of mesclun lettuce. Arch Intern Med 159:1758–1764.
20. Sivapalasingam S, Friedman CR, Cohen L, Tauxe RV. 2004. Fresh produce: a growing cause of outbreaks of foodborne illness in the United States, 1973 through 1997. J Food Prot 67:2342–2353.
21. Rangel JM, Sparling PH, Crowe C, Griffin PM, Swerdlow DL. 2005. Epidemiology of Escherichia coli O157:H7 outbreaks, United States, 1982-2002. Emerg Infect Dis 11:603–609.
22. Delaquis P, Bach S, Dinu L-D. 2007. Behavior of Escherichia coli O157:H7 in leafy vegetables. J Food Prot 70:1966–1974.
23. Fletcher J, Leach JE, Eversole K, Tauxe R. 2014. Human Pathogens on Plants: Designing a Multidisciplinary Strategy for Research. Phytopathology PHYTO09120236RVWtest.

24. Erickson MC, Liao J, Payton AS, Riley DG, Webb CC, Davey LE, Kimbrel S, Ma L, Zhang G, Flitcroft I, Doyle MP, Beuchat LR. 2010. Preharvest internalization of Escherichia coli O157:H7 into lettuce leaves, as affected by insect and physical damage. J Food Prot 73:1809–1816.
25. Aruscavage D, Lee K, Miller S, LeJeune JT. 2006. Interactions Affecting the Proliferation and Control of Human Pathogens on Edible Plants. J Food Sci 71:R89–R99.
26. Solomon EB, Yaron S, Matthews KR. 2002. Transmission of Escherichia coli O157:H7 from contaminated manure and irrigation water to lettuce plant tissue and its subsequent internalization. Appl Environ Microbiol 68:397–400.
27. Martínez-Vaz BM, Fink RC, Diez-Gonzalez F, Sadowsky MJ. 2014. Enteric pathogen-plant interactions: molecular connections leading to colonization and growth and implications for food safety. Microbes Environ 29:123–135.
28. Himathongkham S, Bahari S, Riemann H, Cliver D. 1999. Survival of Escherichia coli O157:H7 and Salmonella typhimurium in cow manure and cow manure slurry. FEMS Microbiol Lett 178:251–257.
29. Lim J-A, Lee DH, Heu S. 2014. The interaction of human enteric pathogens with plants. Plant Pathol J 30:109–116.
30. Cools D, Merckx R, Vlassak K, Verhaegen J. 2001. Survival of E. coli and Enterococcus spp. derived from pig slurry in soils of different texture. Appl Soil Ecol 17:53–62.
31. Guan TTY, Blank G, Holley RA. 2005. Survival of pathogenic bacteria in pesticide solutions and on treated tomato plants. J Food Prot 68:296–304.
32. Saldaña Z, Sánchez E, Xicohtencatl-Cortes J, Puente JL, Girón JA. 2011. Surface structures involved in plant stomata and leaf colonization by shiga-toxigenic Escherichia coli o157:h7. Front Microbiol 2:119.
33. Deering AJ, Mauer LJ, Pruitt RE. 2012. Internalization of E. coli O157:H7 and Salmonella spp. in plants: A review. Food Res Int 45:567–575.
34. Erickson MC. 2012. Internalization of fresh produce by foodborne pathogens. Annu Rev Food Sci Technol 3:283–310.
35. Jablasone J, Warriner K, Griffiths M. 2005. Interactions of Escherichia coli O157:H7, Salmonella typhimurium and Listeria monocytogenes plants cultivated in a gnotobiotic system. Int J Food Microbiol 99:7–18.

36. Franz E, Visser AA, Van Diepeningen AD, Klerks MM, Termorshuizen AJ, van Bruggen AHC. 2007. Quantification of contamination of lettuce by GFP-expressing Escherichia coli O157:H7 and Salmonella enterica serovar Typhimurium. Food Microbiol 24:106–112.
37. Fink RC, Black EP, Hou Z, Sugawara M, Sadowsky MJ, Diez-Gonzalez F. 2012. Transcriptional responses of Escherichia coli K-12 and O157:H7 associated with lettuce leaves. Appl Environ Microbiol 78:1752–1764.
38. Hou Z, Fink RC, Sugawara M, Diez-Gonzalez F, Sadowsky MJ. 2013. Transcriptional and functional responses of Escherichia coli O157:H7 growing in the lettuce rhizoplane. Food Microbiol 35:136–142.
39. Klerks MM, Franz E, van Gent-Pelzer M, Zijlstra C, van Bruggen AHC. 2007. Differential interaction of Salmonella enterica serovars with lettuce cultivars and plant-microbe factors influencing the colonization efficiency. ISME J 1:620–631.
40. Barak JD, Liang A, Narm K-E. 2008. Differential Attachment to and Subsequent Contamination of Agricultural Crops by Salmonella enterica. Appl Environ Microbiol 74:5568–5570.
41. Xicohtencatl-Cortes J, Sánchez Chacón E, Saldaña Z, Freer E, Girón JA. 2009. Interaction of Escherichia coli O157:H7 with leafy green produce. J Food Prot 72:1531–1537.
42. Shaw RK, Berger CN, Feys B, Knutton S, Pallen MJ, Frankel G. 2008. Enterohemorrhagic Escherichia coli exploits EspA filaments for attachment to salad leaves. Appl Environ Microbiol 74:2908–2914.
43. Deng K, Wang S, Rui X, Zhang W, Tortorello ML. 2011. Functional analysis of ycfR and ycfQ in Escherichia coli O157:H7 linked to outbreaks of illness associated with fresh produce. Appl Environ Microbiol 77:3952–3959.
44. Salazar JK, Deng K, Tortorello ML, Brandl MT, Wang H, Zhang W. 2013. Genes ycfR, sirA and yigG contribute to the surface attachment of Salmonella enterica Typhimurium and Saintpaul to fresh produce. PLoS One 8:e57272.
45. Roy D, Panchal S, Rosa BA, Melotto M. 2013. Escherichia coli O157:H7 induces stronger plant immunity than Salmonella enterica Typhimurium SL1344. Phytopathology 103:326–332.

46. Schikora A, Carreri A, Charpentier E, Hirt H. 2008. The Dark Side of the Salad: Salmonella typhimurium Overcomes the Innate Immune Response of Arabidopsis thaliana and Shows an Endopathogenic Lifestyle. PLoS One 3:e2279.
47. Melotto M, Underwood W, Koczan J, Nomura K, He SY. 2006. Plant stomata function in innate immunity against bacterial invasion. Cell 126:969–980.
48. Schikora A, Garcia AV, Hirt H. 2012. Plants as alternative hosts for Salmonella. Trends Plant Sci 17:245–249.
49. Schikora A, Virlogeux-Payant I, Bueso E, Garcia AV, Nilau T, Charrier A, Pelletier S, Menanteau P, Baccarini M, Velge P, Hirt H. 2011. Conservation of Salmonella infection mechanisms in plants and animals. PLoS One 6:e24112.
50. Shirron N, Yaron S. 2011. Active suppression of early immune response in tobacco by the human pathogen Salmonella Typhimurium. PLoS One 6:e18855.
51. Quilliam RS, Williams AP, Jones DL. 2012. Lettuce cultivar mediates both phyllosphere and rhizosphere activity of Escherichia coli O157:H7. PLoS One 7:e33842.
52. Mitra R, Cuesta-Alonso E, Wayadande A, Talley J, Gilliland S, Fletcher J. 2009. Effect of route of introduction and host cultivar on the colonization, internalization, and movement of the human pathogen Escherichia coli O157:H7 in spinach. J Food Prot 72:1521–1530.
53. Hunter PJ, Hand P, Pink D, Whipps JM, Bending GD. 2010. Both leaf properties and microbe-microbe interactions influence within-species variation in bacterial population diversity and structure in the lettuce (Lactuca Species) phyllosphere. Appl Environ Microbiol 76:8117–8125.
54. Brandl MT, Amundson R. 2008. Leaf age as a risk factor in contamination of lettuce with Escherichia coli O157:H7 and Salmonella enterica. Appl Environ Microbiol 74:2298–2306.
55. Critzer FJ, Doyle MP. 2010. Microbial ecology of foodborne pathogens associated with produce. Curr Opin Biotechnol 21:125–130.
56. Cooley MB, Chao D, Mandrell RE. 2006. Escherichia coli O157:H7 survival and growth on lettuce is altered by the presence of epiphytic bacteria. J Food Prot 69:2329–2335.
57. Cooley MB, Miller WG, Mandrell RE. 2003. Colonization of Arabidopsis thaliana with Salmonella enterica and enterohemorrhagic Escherichia coli O157:H7 and competition by Enterobacter asburiae. Appl Environ Microbiol 69:4915–4926.

58. Poza-Carrion C, Suslow T, Lindow S. 2013. Resident bacteria on leaves enhance survival of immigrant cells of Salmonella enterica. Phytopathology 103:341–351.
59. Teplitski M, Warriner K, Bartz J, Schneider KR. 2011. Untangling metabolic and communication networks: interactions of enterics with phytobacteria and their implications in produce safety. Trends Microbiol 19:121–127.
60. Rastogi G, Coaker GL, Leveau JHJ. 2013. New insights into the structure and function of phyllosphere microbiota through high-throughput molecular approaches. FEMS Microbiol Lett 348:1–10.
61. Upreti R, Thomas P. 2015. Root-associated bacterial endophytes from Ralstonia solanacearum resistant and susceptible tomato cultivars and their pathogen antagonistic effects. Front Microbiol 6:255.
62. Berg G, Hallmann J. 2006. Control of plant pathogenic fungi with bacterial endophytes. Microbial root endophytes.
63. Cao L, Qiu Z, You J, Tan H, Zhou S. 2005. Isolation and characterization of endophytic streptomycete antagonists of fusarium wilt pathogen from surface-sterilized banana roots. FEMS Microbiol Lett.
64. Lindow SE, Brandl MT. 2003. Microbiology of the phyllosphere. Appl Environ Microbiol 69:1875–1883.
65. Lund BM. 1992. Ecosystems in vegetable foods. Soc Appl Bacteriol Symp Ser 21:115S–26S.
66. Dees MW, Lysøe E, Nordskog B, Brurberg MB. 2015. Bacterial communities associated with surfaces of leafy greens: shift in composition and decrease in richness over time. Appl Environ Microbiol 81:1530–1539.
67. Rastogi G, Sbodio A, Tech JJ, Suslow TV, Coaker GL, Leveau JHJ. 2012. Leaf microbiota in an agroecosystem: spatiotemporal variation in bacterial community composition on field-grown lettuce. ISME J 6:1812–1822.
68. Jackson CR, Randolph KC, Osborn SL, Tyler HL. 2013. Culture dependent and independent analysis of bacterial communities associated with commercial salad leaf vegetables. BMC Microbiol 13:274.
69. Mercier J, Lindow SE. 2000. Role of leaf surface sugars in colonization of plants by bacterial epiphytes. Appl Environ Microbiol 66:369–374.

70. Bunster L, Fokkema NJ, Schippers B. 1989. Effect of Surface-Active Pseudomonas spp. on Leaf Wettability. Appl Environ Microbiol 55:1340–1345.
71. Leben C. 1988. Relative humidity and the survival of epiphytic bacteria with buds and leaves of cucumber plants. Phytopathology.
72. Williams TR, Moyne A-L, Harris LJ, Marco ML. 2013. Season, irrigation, leaf age, and Escherichia coli inoculation influence the bacterial diversity in the lettuce phyllosphere. PLoS One 8:e68642.
73. Williams TR, Marco ML. 2014. Phyllosphere microbiota composition and microbial community transplantation on lettuce plants grown indoors. MBio 5.
74. Leff JW, Fierer N. 2013. Bacterial communities associated with the surfaces of fresh fruits and vegetables. PLoS One 8:e59310.
75. Kalivas A, Ganopoulos I, Psomopoulos F, Grigoriadis I, Xanthopoulou A, Hatzigiannakis E, Osathanunkul M, Tsaftaris A, Madesis P. 2017. Comparative metagenomics reveals alterations in the soil bacterial community driven by N-fertilizer and Amino 16® application in lettuce. Genom Data 14:14–17.
76. Di Carli M, De Rossi P, Paganin P, Del Fiore A, Lecce F, Capodicasa C, Bianco L, Perrotta G, Mengoni A, Bacci G, Daroda L, Dalmastri C, Donini M, Bevivino A. 2016. Bacterial community and proteome analysis of fresh-cut lettuce as affected by packaging. FEMS Microbiol Lett 363:fnv209.
77. Bodenhausen N, Horton MW, Bergelson J. 2013. Bacterial Communities Associated with the Leaves and the Roots of Arabidopsis thaliana. PLoS One 8:e56329.
78. Banada PP, Huff K, Bae E, Rajwa B, Aroonnual A, Bayraktar B, Adil A, Robinson JP, Hirleman ED, Bhunia AK. 2009. Label-free detection of multiple bacterial pathogens using light-scattering sensor. Biosens Bioelectron 24:1685–1692.
79. Singhal N, Kumar M, Kanaujia PK, Virdi JS. 2015. MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. Front Microbiol 6:791.
80. Robinson JP, Rajwa BP, Bae E, Patsekin V, Roumani AM, Bhunia AK, Dietz JE, Davisson VJ, Dundar MM, Thomas J, Hirleman ED. 2011. Using scattering to identify bacterial pathogens. Opt Photonics News 22:20.

81. He Y, Reed S, Bhunia AK, Gehring A, Nguyen L-H, Irwin PL. 2015. Rapid identification and classification of Campylobacter spp. using laser optical scattering technology. Food Microbiol 47:28–35.
82. Banada PP, Guo S, Bayraktar B, Bae E, Rajwa B, Robinson JP, Hirleman ED, Bhunia AK. 2007. Optical forward-scattering for detection of Listeria monocytogenes and other Listeria species. Biosens Bioelectron 22:1664–1671.
83. Huff K, Aroonnual A, Littlejohn AEF, Rajwa B, Bae E, Banada PP, Patsekin V, Hirleman ED, Robinson JP, Richards GP, Bhunia AK. 2012. Light-scattering sensor for real-time identification of Vibrio parahaemolyticus, Vibrio vulnificus and Vibrio cholerae colonies on solid agar plate. Microb Biotechnol 5:607–620.
84. Singh AK, Bettasso AM, Bae E, Rajwa B, Dundar MM, Forster MD, Liu L, Barrett B, Lovchik J, Robinson JP, Hirleman ED, Bhunia AK. 2014. Laser optical sensor, a label-free on-plate Salmonella enterica colony detection tool. MBio 5:e01019-13.
85. Tang Y, Kim H, Singh AK, Aroonnual A, Bae E, Rajwa B, Fratamico PM, Bhunia AK. 2014. Light scattering sensor for direct identification of colonies of Escherichia coli serogroups O26, O45, O103, O111, O121, O145 and O157. PLoS One 9:e105272.
86. Singh AK, Leprun L, Drolia R, Bai X, Kim H, Aroonnual A, Bae E, Mishra KK, Bhunia AK. 2016. Virulence Gene-Associated Mutant Bacterial Colonies Generate Differentiating Two-Dimensional Laser Scatter Fingerprints. Appl Environ Microbiol 82:3256–3268.
87. Uhlich GA, Reichenberger ER, Cottrell BJ, Fratamico P, Andreozzi E. 2017. Whole-genome sequence of Escherichia coli serotype O157:H7 strain B6914-ARS. Genome Announc 5.
88. Fratamico PM, Deng MY, Strobaugh TP, Palumbo SA. 1997. Construction and characterization of Escherichia coli O157:H7 strains expressing firefly Luciferase and green fluorescent protein and their use in survival studies. J Food Prot 60:1167–1173.
89. Vialette M, Jandos-Rudnik A-M, Guyard C, Legeay O, Pinon A, Lange M. 2004. Validating the use of green fluorescent-marked Escherichia coli O157:H7 for assessing the organism behaviour in foods. J Appl Microbiol 96:1097–1104.
90. Kudva IT, Blanch K, Hovde CJ. 1998. Analysis of Escherichia coli O157:H7 survival in ovine or bovine manure and manure slurry. Appl Environ Microbiol 64:3166–3174.
91. Chelius MK, Triplett EW. 2001. The Diversity of Archaea and Bacteria in Association with the Roots of Zea mays L. Microb Ecol 41:252–263.

92. Beckers B, Op De Beeck M, Thijs S, Truyens S, Weyens N, Boerjan W, Vangronsveld J. 2016. Performance of 16s rDNA Primer Pairs in the Study of Rhizosphere and Endosphere Bacterial Microbiomes in Metabarcoding Studies. Front Microbiol 7:650.
93. Huse SM, Dethlefsen L, Huber JA, Mark Welch D, Welch DM, Relman DA, Sogin ML. 2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. PLoS Genet 4:e1000255.
94. Callahan BJ, Sankaran K, Fukuyama JA, McMurdie PJ, Holmes SP. 2016. Bioconductor workflow for microbiome data analysis: from raw reads to community analyses. F1000Res 5.
95. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583.
96. McMurdie PJ, Holmes S. 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One 8:e61217.
97. Koukkidis G, Haigh R, Allcock N, Jordan S, Freestone P. 2017. Salad leaf juices enhance Salmonella growth, colonization of fresh produce, and virulence. Appl Environ Microbiol 83:AEM.02416-16.
98. Moyne A-L, Harris LJ, Marco ML. 2013. Assessments of total and viable Escherichia coli O157:H7 on field and laboratory grown lettuce. PLoS One 8:e70643.
99. Kim S-R, Yoon Y, Kim W-I, Park K-H, Yun H-J, Chung DH, Yun JC, Ryu KY. 2012. Comparison of sample preparation methods for the recovery of foodborne pathogens from fresh produce. J Food Prot 75:1213–1218.
100. Brandl MT. 2008. Plant Lesions Promote the Rapid Multiplication of Escherichia coli O157:H7 on Postharvest Lettuce. Appl Environ Microbiol 74:5285–5289.
101. Weller DL, Kovac J, Roof S, Kent DJ, Tokman JI, Kowalcyk B, Oryang D, Ivanek R, Aceituno A, Sroka C, Wiedmann M. 2017. Survival of Escherichia coli on lettuce under field conditions encountered in the northeastern United States. J Food Prot 80:1214–1221.
102. Moyne A-L, Sudarshana MR, Blessington T, Koike ST, Cahn MD, Harris LJ. 2011. Fate of Escherichia coli O157:H7 in field-inoculated lettuce. Food Microbiol 28:1417–1425.
103. Erickson MC, Webb CC, Diaz-Perez JC, Phatak SC, Silvoy JJ, Davey L, Payton AS, Liao J, Ma L, Doyle MP. 2010. Surface and internalized Escherichia coli O157:H7 on field-grown spinach and lettuce treated with spray-contaminated irrigation water. J Food Prot 73:1023–1029.

104. Erickson MC, Webb CC, Davey LE, Payton AS, Flitcroft ID, Doyle MP. 2014. Internalization and fate of Escherichia coli O157:H7 in leafy green phyllosphere tissue using various spray conditions. J Food Prot 77:713–721.

105. Gay AP, Hurd RG. 1975. The influence of light on stomatal density in the tomato. New Phytol 75:37–46.

106. Peel JR, Mandujano Sanchez MC, Lopez Portillo J, Golubov J. 2017. Stomatal density, leaf area and plant size variation of Rhizophora mangle (Malpighiales: Rhizophoraceae) along a salinity gradient in the Mexican Caribbean. Rev Biol Trop 65.

107. Salisbury EJ. 1928. On the causes and ecological significance of stomatal frequency, with special reference to the woodland flora. Philos Trans R Soc Lond B Biol Sci 216:1–65.

108. Moyne A-L, Blessington T, Williams TR, Koike ST, Cahn MD, Marco ML, Harris LJ. 2020. Conditions at the time of inoculation influence survival of attenuated Escherichia coli O157:H7 on field-inoculated lettuce. Food Microbiol 85:103274.

109. Centers for Disease Control and Prevention (CDC). 2017. Surveillance for Foodborne Disease Outbreaks, United States, 2015, Annual Report. CDC.

110. Berg G, Köberl M, Rybakova D, Müller H, Grosch R, Smalla K. 2017. Plant microbial diversity is suggested as the key to future biocontrol and health trends. FEMS Microbiol Ecol 93.

111. Center for Food Safety, Applied Nutrition. 2020. BAM chapter 3: Aerobic plate count.

112. Korhonen JT, Huhtamäki T, Ikkala O, Ras RHA. 2013. Reliable measurement of the receding contact angle. Langmuir 29:3858–3863.

113. Lu L, Ku K-M, Palma-Salgado SP, Storm AP, Feng H, Juvik JA, Nguyen TH. 2015. Influence of epicuticular physicochemical properties on porcine Rotavirus adsorption to 24 leafy green vegetables and tomatoes. PLoS One 10:e0132841.

114. Fernandez-Moreno J-P, Malitsky S, Lashbrooke J, Biswal AK, Racovita RC, Mellerowicz EJ, Jetter R, Orzaez D, Aharoni A, Granell A. 2016. An efficient method for medium throughput screening of cuticular wax composition in different plant species. Metabolomics 12.

115. McElwain JC, Yiotis C, Lawson T. 2016. Using modern plant trait relationships between observed and theoretical maximum stomatal conductance and vein density to examine patterns of plant macroevolution. New Phytol 209:94–103.

116. Savvides A, Fanourakis D, van Ieperen W. 2012. Co-ordination of hydraulic and stomatal conductances across light qualities in cucumber leaves. J Exp Bot 63:1135–1143.
117. Hanshew AS, Mason CJ, Raffa KF, Currie CR. 2013. Minimization of chloroplast contamination in 16S rRNA gene pyrosequencing of insect herbivore bacterial communities. J Microbiol Methods 95:149–155.
118. Hussing C, Kampmann M-L, Mogensen HS, Børsting C, Morling N. 2018. Quantification of massively parallel sequencing libraries - a comparative study of eight methods. Sci Rep 8:1110.
119. Callahan BJ, McMurdie PJ, Holmes SP. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J 11:2639–2643.
120. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590-6.
121. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550.
122. Moore, Salisbury FB, Moore R. 2004. Introduction to Plant Physiology. Thomson Brooks/Cole.
123. Chang F-M, Hong S-J, Sheng Y-J, Tsao H-K. 2009. High contact angle hysteresis of superhydrophobic surfaces: Hydrophobic defects. Appl Phys Lett 95:064102.
124. Burton Z, Bhushan B. 2006. Surface characterization and adhesion and friction properties of hydrophobic leaf surfaces. Ultramicroscopy 106:709–719.
125. Brewer CA, Nuñez CI. 2007. Patterns of leaf wettability along an extreme moisture gradient in western Patagonia, Argentina. Int J Plant Sci 168:555–562.
126. Wang H, Shi H, Wang Y. 2015. The wetting of leaf surfaces and its ecological significancesWetting and Wettability. InTech.
127. Holroyd GH, Hetherington AM, Gray JE. 2002. A role for the cuticular waxes in the environmental control of stomatal development. New Phytol 153:433–439.
128. Karabourniotis G. 2001. Epicuticular phenolics over guard cells: Exploitation for in situ stomatal counting by fluorescence microscopy and combined image analysis. Ann Bot 87:631–639.

129. Kroupitski Y, Golberg D, Belausov E, Pinto R, Swartzberg D, Granot D, Sela S. 2009. Internalization of Salmonella enterica in leaves is induced by light and involves chemotaxis and penetration through open stomata. Appl Environ Microbiol 75:6076–6086.
130. Hanba YT, Moriya A, Kimura K. 2004. Effect of leaf surface wetness and wettability on photosynthesis in bean and pea. Plant Cell Environ 27:413–421.
131. Beattie GA. 2011. Water relations in the interaction of foliar bacterial pathogens with plants. Annu Rev Phytopathol 49:533–555.
132. Koch K, Hartmann KD, Schreiber L, Barthlott W, Neinhuis C. 2006. Influences of air humidity during the cultivation of plants on wax chemical composition, morphology and leaf surface wettability. Environ Exp Bot 56:1–9.
133. Ku K-M, Chiu Y-C, Shen C, Jenks M. 2020. Leaf cuticular waxes of lettuce are associated with reduced attachment of the foodborne pathogen Salmonella spp. at harvest and after postharvest storage. Lebenson Wiss Technol 117:108657.
134. Ormeño E, Ruffault J, Gutigny C, Madrigal J, Guijarro M, Hernando C, Ballini C. 2020. Increasing cuticular wax concentrations in a drier climate promote litter flammability. For Ecol Manage 473:118242.
135. Beattie GA, Lindow SE. 1999. Bacterial colonization of leaves: a spectrum of strategies. Phytopathology 89:353–359.
136. Leveau JH, Lindow SE. 2001. Appetite of an epiphyte: quantitative monitoring of bacterial sugar consumption in the phyllosphere. Proc Natl Acad Sci U S A 98:3446–3453.
137. VAN der Linden I, Eriksson M, Uyttendaele M, Devlieghere F. 2016. Is there a relation between the microscopic leaf morphology and the association of Salmonella and Escherichia coli O157:H7 with iceberg lettuce leaves? J Food Prot 79:1784–1788.
138. Koch K, Ensikat H-J. 2008. The hydrophobic coatings of plant surfaces: epicuticular wax crystals and their morphologies, crystallinity and molecular self-assembly. Micron 39:759–772.
139. Broadley MR, Escobar-Gutiérrez AJ, Burns A, Burns IG. 2001. Nitrogen-limited growth of lettuce is associated with lower stomatal conductance. New Phytol 152:97–106.

140. Lopez-Velasco G, Welbaum GE, Boyer RR, Mane SP, Ponder MA. 2011. Changes in spinach phylloepiphytic bacteria communities following minimal processing and refrigerated storage described using pyrosequencing of 16S rRNA amplicons. J Appl Microbiol 110:1203–1214.

141. Kroupitski Y, Pinto R, Belausov E, Sela S. 2011. Distribution of Salmonella typhimurium in romaine lettuce leaves. Food Microbiol 28:990–997.

142. Kim M, Singh D, Lai-Hoe A, Go R, Abdul Rahim R, Ainuddin AN, Chun J, Adams JM. 2012. Distinctive phyllosphere bacterial communities in tropical trees. Microb Ecol 63:674–681.

143. Ottesen AR, González Peña A, White JR, Pettengill JB, Li C, Allard S, Rideout S, Allard M, Hill T, Evans P, Strain E, Musser S, Knight R, Brown E. 2013. Baseline survey of the anatomical microbial ecology of an important food plant: Solanum lycopersicum (tomato). BMC Microbiol 13:114.

144. Fahlgren C, Hagström A, Nilsson D, Zweifel UL. 2010. Annual variations in the diversity, viability, and origin of airborne bacteria. Appl Environ Microbiol 76:3015–3025.

145. Lax S, Cardona C, Zhao D, Winton VJ, Goodney G, Gao P, Gottel N, Hartmann EM, Henry C, Thomas PM, Kelley ST, Stephens B, Gilbert JA. 2019. Microbial and metabolic succession on common building materials under high humidity conditions. Nat Commun 10:1767.

146. Krimm U, Abanda-Nkpwatt D, Schwab W, Schreiber L. 2005. Epiphytic microorganisms on strawberry plants (Fragaria ananassa cv. Elsanta): identification of bacterial isolates and analysis of their interaction with leaf surfaces. FEMS Microbiol Ecol 53:483–492.

147. Sivakumar N, Sathishkumar R, Selvakumar G, Shyamkumar R, Arjunekumar K. 2020. Phyllospheric microbiomes: Diversity, ecological significance, and biotechnological applications, p. 113–172. *In* Sustainable Development and Biodiversity. Springer International Publishing, Cham.

148. Joseph CM, Phillips DA. 2003. Metabolites from soil bacteria affect plant water relations. Plant Physiol Biochem 41:189–192.

149. Kolton M, Sela N, Elad Y, Cytryn E. 2013. Comparative genomic analysis indicates that niche adaptation of terrestrial Flavobacteria is strongly linked to plant glycan metabolism. PLoS One 8:e76704.

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

151. Reisberg EE, Hildebrandt U, Riederer M, Hentschel U. 2013. Distinct phyllosphere bacterial communities on Arabidopsis wax mutant leaves. PLoS One 8:e78613.

152. Dassanayake M, Larkin JC. 2017. Making plants break a sweat: The structure, function, and evolution of plant salt glands. Front Plant Sci 8:406.

153. Gómez Brandón M, Aira M, Kolbe AR, de Andrade N, Pérez-Losada M, Domínguez J. 2019. Rapid bacterial community changes during vermicomposting of grape Marc derived from red winemaking. Microorganisms 7:473.

154. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. 2011. Metagenomic biomarker discovery and explanation. Genome Biol 12:R60.

155. Hughes AR, Byrnes JE, Kimbro DL, Stachowicz JJ. 2007. Reciprocal relationships and potential feedbacks between biodiversity and disturbance. Ecol Lett 10:849–864.

156. Reese AT, Dunn RR. 2018. Drivers of microbiome biodiversity: A review of general rules, feces, and ignorance. MBio 9.

157. Frenzen PD, Drake A, Angulo FJ, Emerging Infections Program FoodNet Working Group. 2005. Economic cost of illness due to Escherichia coli O157 infections in the United States. J Food Prot 68:2623–2630.

158. Oliveira M, Viñas I, Usall J, Anguera M, Abadias M. 2012. Presence and survival of Escherichia coli O157:H7 on lettuce leaves and in soil treated with contaminated compost and irrigation water. Int J Food Microbiol 156:133–140.

159. Marine SC, Pagadala S, Wang F, Pahl DM, Melendez MV, Kline WL, Oni RA, Walsh CS, Everts KL, Buchanan RL, Micallef SA. 2015. The growing season, but not the farming system, is a food safety risk determinant for leafy greens in the mid-Atlantic region of the United States. Appl Environ Microbiol 81:2395–2407.

160. Stine SW, Song I, Choi CY, Gerba CP. 2005. Effect of relative humidity on preharvest survival of bacterial and viral pathogens on the surface of cantaloupe, lettuce, and bell peppers. J Food Prot 68:1352–1358.

161. Gibson AM, Roberts TA. 1986. The effect of pH, water activity, sodium nitrate and storage temperature on the growth of enteropathogenic Escherichia coli and salmonellae in a laboratory medium. Int J Food Microbiol 3:183–194.

162. Griffin PM, Ostroff SM, Tauxe RV, Greene KD, Wells JG, Lewis JH, Blake PA. 1988. Illnesses associated with Escherichia coli O157:H7 infections. A broad clinical spectrum. Ann Intern Med 109:705–712.

163. McKellar RC, Peréz-Rodríguez F, Harris LJ, Moyne A-L, Blais B, Topp E, Bezanson G, Bach S, Delaquis P. 2014. Evaluation of different approaches for modeling Escherichia coli O157:H7 survival on field lettuce. Int J Food Microbiol 184:74–85.

164. Bezanson G, Delaquis P, Bach S, McKellar R, Topp E, Gill A, Blais B, Gilmour M. 2012. Comparative examination of Escherichia coli O157:H7 survival on romaine lettuce and in soil at two independent experimental sites. J Food Prot 75:480–487.

165. Fonseca JM, Fallon SD, Sanchez CA, Nolte KD. 2011. Escherichia coli survival in lettuce fields following its introduction through different irrigation systems. J Appl Microbiol 110:893–902.

166. Xiong P, Chen Z, Jia Z, Wang Z, Palta JA, Xu B. 2018. Variability in leaf wettability and surface water retention of main species in semiarid Loess Plateau of China. Ecohydrology 11:e2021.

167. Aryal B, Neuner G. 2010. Leaf wettability decreases along an extreme altitudinal gradient. Oecologia 162:1–9.

168. Holder CD. 2007. Leaf water repellency of species in Guatemala and Colorado (USA) and its significance to forest hydrology studies. J Hydrol (Amst) 336:147–154.

169. Holder CD. 2012. The relationship between leaf hydrophobicity, water droplet retention, and leaf angle of common species in a semi-arid region of the western United States. Agric For Meteorol 152:11–16.

170. 2010. Leaf surface wettability of major plant species for urban greening in Xi'an and related affecting factors. Chinese Journal of Ecology 29:630–636.

171. Kumar N, Pandey S, Bhattacharya A, Ahuja PS. 2004. Do leaf surface characteristics affect Agrobacterium infection in tea [Camellia sinensis (L.) O Kuntze]? J Biosci 29:309–317.

172. Mariano RL, McCarter SM. 1993. Epiphytic survival of Pseudomonas viridiflava on tomato and selected weed species. Microb Ecol 26:47–58.

173. Papierowska E, Szporak-Wasilewska S, Szewińska J, Szatyłowicz J, Debaene G, Utratna M. 2018. Contact angle measurements and water drop behavior on leaf surface for several deciduous shrub and tree species from a temperate zone. Trees (Berl West) 32:1253–1266.

174. Godfrey BES. 1976. Leachates from aerial parts of plants and their relation to plant surface microbial populations, p. 433–439. *In* Microbiology of Aerial Plant Surfaces. Elsevier.

175. Khodadad CLM, Hummerick ME, Spencer LE, Dixit AR, Richards JT, Romeyn MW, Smith TM, Wheeler RM, Massa GD. 2020. Microbiological and nutritional analysis of lettuce crops grown on the International Space Station. Front Plant Sci 11:199.

176. Aydogan EL, Moser G, Müller C, Kämpfer P, Glaeser SP. 2018. Long-term warming shifts the composition of bacterial communities in the phyllosphere of Galium album in a permanent grassland field-experiment. Front Microbiol 9.

177. Schlechter RO, Miebach M, Remus-Emsermann MNP. 2019. Driving factors of epiphytic bacterial communities: A review. J Adv Res 19:57–65.

178. Ho A, Lonardo DPD, Bodelier PLE. 2017. Revisiting life strategy concepts in environmental microbial ecology. FEMS Microbiol Ecol fix006.

179. Leff JW, Jones SE, Prober SM, Barberán A, Borer ET, Firn JL, Harpole WS, Hobbie SE, Hofmockel KS, Knops JMH, McCulley RL, La Pierre K, Risch AC, Seabloom EW, Schütz M, Steenbock C, Stevens CJ, Fierer N. 2015. Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. Proc Natl Acad Sci U S A 112:10967–10972.

180. Goldfarb KC, Karaoz U, Hanson CA, Santee CA, Bradford MA, Treseder KK, Wallenstein MD, Brodie EL. 2011. Differential growth responses of soil bacterial taxa to carbon substrates of varying chemical recalcitrance. Front Microbiol 2:94.

181. Seo S, Matthews KR. 2012. Influence of the Plant Defense Response to Escherichia coli O157:H7 Cell Surface Structures on Survival of That Enteric Pathogen on Plant Surfaces. Appl Environ Microbiol 78:5882–5889.

182. Yadav RKP, Karamanoli K, Vokou D. 2005. Bacterial colonization of the phyllosphere of mediterranean perennial species as influenced by leaf structural and chemical features. Microb Ecol 50:185–196.

183. Lopez-Velasco G, Tydings HA, Boyer RR, Falkinham JO 3rd, Ponder MA. 2012. Characterization of interactions between Escherichia coli O157:H7 with epiphytic bacteria in vitro and on spinach leaf surfaces. Int J Food Microbiol 153:351–357.

184. Robinson JP, Rajwa BP, Bae E, Patsekin V, Roumani AM, Bhunia AK, Dietz JE, Davisson VJ, Dundar MM, Thomas J, Hirleman ED. 2011. Using Scattering to Identify Bacterial Pathogens. Optics & Photonics News, OPN 22:20–27.

185. Maturin L, Peeler JT, Others. 2001. Bacteriological analytical manual. Chapter 3. Aerobic plate count. FDA Med Bull.

186. Yang B, Wang Y, Qian P-Y. 2016. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. BMC Bioinformatics 17:135.

187. Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. J Comput Biol 7:203–214.

188. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73:5261–5267.

189. Gregory Caporaso J, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Publishing Group 7:335–336.

190. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461.

191. Rajwa B, Dundar MM, Akova F, Bettasso A, Patsekin V, Dan Hirleman E, Bhunia AK, Robinson JP. 2010. Discovering the unknown: detection of emerging pathogens using a label-free light-scattering system. Cytometry A 77:1103–1112.

192. Ahmed WM, Bayraktar B, Bhunia A, Hirleman ED, Robinson JP, Rajwa B. 2013. Classification of bacterial contamination using image processing and distributed computing. IEEE J Biomed Health Inform 17:232–239.

193. Alsulami TS, Zhu X, Abdelhaseib MU, Singh AK, Bhunia AK. 2018. Rapid detection and differentiation of Staphylococcus colonies using an optical scattering technology. Anal Bioanal Chem 410:5445–5454.

194. Baldi P, Brunak S, Chauvin Y, Andersen CA, Nielsen H. 2000. Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics 16:412–424.

195. Kim H, Bai N, Bhunia AK, King GB, Hirleman ED, Bae E. 2013. Development of an integrated optical analyzer for characterization of growth dynamics of bacterial colonies. J Biophotonics 6:929–937.

196. Fröhling A, Rademacher A, Rumpold B, Klocke M, Schlüter O. 2018. Screening of microbial communities associated with endive lettuce during postharvest processing on industrial scale. Heliyon 4:e00671.

197. Abadias M, Usall J, Anguera M, Solsona C, Viñas I. 2008. Microbiological quality of fresh, minimally-processed fruit and vegetables, and sprouts from retail establishments. Int J Food Microbiol 123:121–129.

198. Rudi K, Flateland SL, Hanssen JF, Bengtsson G, Nissen H. 2002. Development and evaluation of a 16S ribosomal DNA array-based approach for describing complex microbial communities in ready-to-eat vegetable salads packed in a modified atmosphere. Appl Environ Microbiol 68:1146–1156.

199. Maffei DF, Silveira NF de A, Catanozi M da PLM. 2013. Microbiological quality of organic and conventional vegetables sold in Brazil. Food Control 29:226–230.

200. Oliveira M, Usall J, Viñas I, Anguera M, Gatius F, Abadias M. 2010. Microbiological quality of fresh lettuce from organic and conventional production. Food Microbiol 27:679–684.

201. Kumar V, Gera R. 2014. Isolation of a multi-trait plant growth promoting Brevundimonas sp. and its effect on the growth of Bt-cotton. 3 Biotech 4:97–101.

202. Korir H, Mungai NW, Thuita M, Hamba Y, Masso C. 2017. Co-inoculation Effect of Rhizobia and Plant Growth Promoting Rhizobacteria on Common Bean Growth in a Low Phosphorus Soil. Front Plant Sci 8:141.

203. Yang S, Zhang X, Cao Z, Zhao K, Wang S, Chen M, Hu X. 2014. Growth-promoting Sphingomonas paucimobilis ZJSH1 associated with Dendrobium officinale through phytohormone production and nitrogen fixation. Microb Biotechnol 7:611–620.

204. Stubbendieck RM, Vargas-Bautista C, Straight PD. 2016. Bacterial Communities: Interactions to Scale. Front Microbiol 7:1234.

205. Traxler MF, Watrous JD, Alexandrov T, Dorrestein PC, Kolter R. 2013. Interspecies interactions stimulate diversification of the Streptomyces coelicolor secreted metabolome. MBio 4.

206. Jeanson S, Floury J, Gagnaire V, Lortal S, Thierry A. 2015. Bacterial Colonies in Solid Media and Foods: A Review on Their Growth and Interactions with the Micro-Environment. Front Microbiol 6:1284.

207. Xiong L, Cao Y, Cooper R, Rappel W-J, Hasty J, Tsimring L. 2020. Flower-like patterns in multi-species bacterial colonies. Elife 9.

208. Doh I-J, Sturgis J, Sarria Zuniga DV, Pruitt RE, Robinson JP, Bae E. 2019. Generalized spectral light scatter models of diverse bacterial colony morphologies. J Biophotonics e201900149.

209. Bae E, Bai N, Aroonnual A, Robinson JP, Bhunia AK, Hirleman ED. 2010. Modeling light propagation through bacterial colonies and its correlation with forward scattering patterns. J Biomed Opt 15:045001.

210. Kim H, Rajwa B, Bhunia AK, Robinson JP, Bae E. 2017. Development of a multispectral light-scatter sensor for bacterial colonies. J Biophotonics 10:634–644.

PUBLICATIONS

- Sarria-Zuniga, D.V., Bae, E., Robinson, J.P., Pruitt, R.E. 2020. Characterization of Culturable Bacterial Communities of Romaine Lettuce Leaves Using Elastic Light Scattering Technology (BARDOT). Manuscript under revision.

- Doh, l-J., Sturgis, J., Sarria-Zuniga, D.V., Pruitt, R.E., Robinson, J.P., Bae, E. 2019. Generalized spectral modeling of diverse bacterial colony morphologies. J Biophotonics e201900149.