Title

Understanding the native microbial populations for better produce safety and quality

Author

Liao, Chao

Publication Date

2022

Peer reviewed|Thesis/dissertation

Understanding the native microbial populations for better produce safety and quality

By

CHAO LIAO

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Food Science

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Luxin Wang, Chair

Gerald Quon

Linda J. Harris

Maria L. Marco

Committee in Charge

2023

Acknowledgments

I would like to sincerely express my gratitude to my advisor, Dr. Luxin Wang. It is hard to believe that I have already worked with Dr. Wang for eight years. During the past years, I received research training in her laboratory and was deeply influenced by her self-discipline, diligence, and wisdom in teaching and research. She always encouraged me to dive deeply into research ideas and gave me enough time and freedom to explore interesting research topics.

I also sincerely thank Dr. Gerald Quon for instructing me to carry out a machine learning-based fresh produce microbiota study in the past two years. I am grateful for his patience in teaching me in this new area, which really opened a new door for me to get involved in the development of artificial intelligence in food systems. I really appreciate Dr. Linda J. Harris and Dr. David A. Mills proposing insightful questions about my research projects and giving valuable suggestions in my dissertation committee meetings. Those questions and suggestions markedly helped improve my dissertation to be rigorous and integrative.

I deeply appreciate all my qualifying exam and dissertation committee members, Dr. Juliana Bell, Dr. Ameer Taha, Dr. Timothy J. Hackmann, Dr. Linda J. Harris, Dr. David A. Mills, Dr. Gerald Quon, Dr. Luxin Wang, and Dr. Maria L. Marco for their patient guidance and insightful suggestions to my research projects and dissertation. I believe the systematic research training and the inspiration from them will be a solid foundation to allow me to pursue my career goals after graduation.

In addition, I would like to thank my lab mates, Dr. Qifan Zeng, Dr. Lina Sheng, Dr. Hongye Wang, Dr. Dong Han, Zhuosheng Liu, Xiran Li, Jing Yuan for their help with my experiments, presentation practices, and dissertation editing. I also want to thank my friends, Nueon Non, Xiao Liu, Jiewen Zhang, Zhitao Gong, Ping Huang, Hairui Liu, Yi Zeng, Luyining Gan, and Luoliang Xu in the United States and China, for making lots of fun and good memories together.

Last but most importantly, I would like to deeply thank my parents, grandmother, aunt, uncle, and cousins. In the past years, they witnessed my ups and downs. They always stood by me and encouraged me emotionally. Without them, I could not reach this far.

Table of Contents

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Abstract

Foodborne outbreaks have been continuously reported during the past decade in the United States. Overall, 129 outbreaks of pathogenic *Escherichia coli* O157:H7, *Listeria monocytogenes*, and *Salmonella enterica* infections linked to fresh produce occurred from 2010 to 2021 in the U.S., causing a total of 8,889 illnesses with 2,412 hospitalizations and 90 deaths. Among the different kinds of fresh produce, Romaine lettuce (RL) is one of the most frequently reported variety of fresh produce associated with the outbreaks. Pathogen contamination can occur at any point during the preharvest, e.g., soil, irrigation water, and insect; and during the postharvest, e.g., produce processing, distribution, and preparation. Intensive studies have shown that foodborne pathogens can persist in RL for a period of time. Unfortunately, current washing strategies cannot eliminate the potential pathogens. The fact is that foodborne pathogens present in fresh produce are typically at low levels, making the detection of pathogens even more challenging. The development of next-generation sequencing techniques provides us with a better understanding of the native microbiota present in RL. The RL microbiota data allow us to not only understand the connection between native microbiota and pathogen contamination but also potentially can be used as a tool to predict produce safety and shelf-life.

In this dissertation, I first investigated the impact of storage time, season, and brands on bacterial populations in commercially washed, chopped, and bagged RL by using culture-dependent and culture-independent methods. The results showed that RL had almost no differences in aerobic plate count (APC) and anaerobic plate count (AnPC) between "Use By" date (UBD) and 5 days after the UBD (UBD5) at 4 °C, but differences were seen between RL from two seasons as well as among three brands of RL. At the same time, the results based on 16S rRNA gene amplicon DNA sequencing showed no difference in bacterial diversity and composition between UBD and UBD5 at 4 °C, while the results showed that the season when the RL was harvested generated a significant impact on the diversity and composition of bacterial communities. The brand is the factor that significantly shaped the population as well, indicating that harvest locations, processing strategies, and distribution conditions may impact bacterial populations (Chapter 2).

Through my artificial inoculation study (Chapter 3), I found the native microbiota in RL impacted the behaviors of *Listeria monocytogenes* (LM) at 4 °C. LM decreased in abundance from UBD to UBD5 in the late-season RL but remained at the same levels in the early-season RL. Results of 16S rRNA gene amplicon DNA sequencing with differential abundance analysis showed that the identified *Leuconostoc*, *Lactococcus*, and *Erwinia* in late-season RL may be the potential reason for the LM decrease. In addition, *Pseudomonas*, *Chryseobacterium*, *Duganella*, and *Listeria* were identified as indicators for RL inoculated with a high level of LM, and *Serratia*, *Pedobacter*, *Janthinobacterium*, *Flavobacterium*, and *Weissella* were identified as indicators for RL inoculated with a low level of LM. Through culture-based methods, 32 antagonistic bacteria were isolated and confirmed by using the double-layer and spot-on-lawn methods. Three species were identified by using Sanger sequencing and whole genome sequencing, with these species being *Carnobacterium maltaromaticum*, *Lactococcus lactis*, and *Bacillus filamentous*. Quantitative PCR was used to further confirm and quantify the three antagonistic species in RL. The results showed no difference in abundances of *C. maltaromaticum* and *Lc. lactis* between RL from two seasons, while *B. filamentosus* presented higher abundances in late-season RL than in early-season RL. Additional research is needed to better confirm the antagonist function of *B. filamentosus* by using the co-culturing method and better identify the anti-LM mechanism of *B. filamentosus*.

During sequencing data processing, up to more than 50% of sequence loss occurred during alignment and denoising, which may cause important bacteria species information to be missing or lost. To test the hypothesis, I explored an alignment-free strategy, k-mer hashes, by establishing classifiers to predict fresh produce contamination and quality reduction. The k-mer hash approach was compared against the amplicon sequence variant (ASV) approach with a typical denoising step. The results showed that random forests (RF)-based classifiers for fresh produce safety and quality that were trained on 7-mer hash-preprocessed publicly available datasets had significantly higher classification accuracy than those using the ASV datasets, supporting our hypothesis that data preprocessing strategies without sequence loss (k-mer hash) retain more important information about bacteria species for produce safety and quality classification than the approach with sequence loss (ASV). In addition, the integration of multiple

produce microbiota datasets led RF-based classifiers to have a higher classification accuracy than the classifiers trained with individual datasets. The RF-based classifiers based on integrated datasets identified more consistent and generalizable indicators associated with fresh produce safety and quality (Chapter 4).

In summary, this dissertation systematically profiled the structures of bacterial populations in RL and investigated how to better utilize culture-dependent methods, culture-independent methods, and publicly available datasets to identify bacteria associated with produce safety and quality and to isolate competitive exclusion bacteria that can potentially be used as biocontrol agents to reduce foodborne pathogen contamination of produce.

Chapter 1. Literature review

Part 1. Major foodborne outbreaks linked to fresh produce from 2010-2021

Introduction

Fresh fruits and vegetables (fresh produce) have become increasingly popular as part of a healthy diet in the past two decades (1, 2). In 2004, the World Health Organization (WHO) recommended that the minimum daily intake of fruits and vegetables is 400 g per day (3). The Economic Research Service (ERS) Food Availability Data System (FADS) shows that the consumption of fresh produce per capita in the U.S. increased from 144.7 kg in 2010 to 149.9 kg in 2018 (newest available data) (4). Unfortunately, increased consumption of fresh produce has been accompanied by an increasing number of foodborne outbreaks, leading to increased public concern about the safety of fresh produce. Fresh produce can be contaminated by pathogens at different points across the food supply chain. Contamination can come from wild animals, birds, feces, dust, and irrigation water during the pre-harvest phase; from an improperly cleaned or sanitized conveyor belt during food processing; or by hands and unclean knives and cutting boards during household processing (5–7).

Major foodborne outbreaks associated with fresh produce

The National Outbreak Reporting System (NORS) was launched by the Centers for Disease Control and Prevention (CDC) in 2009 with the purpose of tracking foodborne disease outbreaks from public health departments in the U.S. (8). The dataset was downloaded from the NORS dashboard (9). The outbreak datasets include information about the outbreak year, number of outbreaks, serotype, number of illnesses, hospitalizations, deaths, and food vehicles. Data with any missing information about the number of illnesses, hospitalizations, and deaths were filtered out. The main pathogens associated with fresh produce outbreaks from 2010 to 2021 were Shiga-toxin-producing *Escherichia coli* (STEC), *Listeria monocytogenes*, *Salmonella enterica*, *Cyclospora cayetanensis*, and Hepatitis A (9). Cases caused by pathogenic *E. coli*, *L. monocytogenes*, and *S. enterica* from 2010 to 2021 were reviewed and

summarized. Overall, 129 reported outbreaks of pathogenic bacteria infections linked to fresh produce occurred from 2010 to 2021 in the U.S., causing a total of 8,889 illnesses with 2,412 hospitalizations and 90 deaths.

A total of 39 outbreaks associated with fresh produce were attributed to STEC infections (**Table 1**). These STEC outbreaks caused 1,237 illnesses with a 38.2% hospitalization rate and a 0.6% death rate. The implicated food vehicles included spinach, lettuce, Romaine lettuce, iceberg lettuce, packaged leafy greens, cabbage, pre-packaged salad, spring salad, clover sprouts, and alfalfa sprouts. The serotypes of STEC causing outbreaks were O157:H7, O157:NM, O145, O103, O111, O121, O26, and O26:H11. Among them, the serotype O157:H7 accounted for 27 outbreaks with 1,007 illnesses, 417 hospitalizations (41.4% hospitalization rate), and eight deaths (0.8% death rate), which indicates that *E. coli* O157:H7 is the most important foodborne pathogen to prevent or eliminate for improving the safety of fresh produce.

L. monocytogenes caused a total of 15 outbreaks (**Table 2**). These outbreaks involved 352 illnesses with 326 hospitalizations and 65 deaths. Compared to STEC infections, the hospitalization rate (92.6%) and death rate (18.5%) were higher. Implicated as the food vehicle in *L. monocytogenes* infection outbreaks were fresh vegetables including sprouts, mung bean sprouts, prepackaged leafy greens, lettuce, and enoki mushrooms. Also implicated were fresh fruits, including cantaloupes, peaches, nectarines, caramel apples, and stone fruits. The fresh vegetables and fresh fruits accounted for 60% and 40% of outbreaks respectively. Unfortunately, no *L. monocytogenes* serotype information was provided in the NORS database.

S. enterica caused a total of 75 outbreaks with7,300 illnesses, including 1,614 hospitalizations (22.1% hospitalization rate) and 17 deaths (0.2% death rate) (**Table 3**). A total of 36 different food vehicles were related to the outbreaks, including fresh vegetables and fresh fruits accounting for 50.6% and 35.6% of outbreaks, respectively. Thirty-seven serotypes of *S. enterica* were reported to be the etiologies. Among these *S. enterica* serotypes, the top three causing the most illnesses are Newport, Oranienburg, and Poona. Specifically, *S.* Newport caused 11 outbreaks linked to fresh produce, with 1,929 illnesses, 286 hospitalizations, and three deaths. *S.* Oranienburg caused one outbreak with 1,040

illnesses, 260 hospitalizations, and no deaths. *S.* Poona accounted for one outbreak with 907 illnesses, 204 hospitalizations, and six deaths. *S.* Enteritidis and *S.* Typhimurium were not the main cause of *Salmonella* outbreaks linked to fresh produce from 2010 to 2021, but they have been reported to be the main causes of outbreaks linked to animal-based foods during 1998-2008 in the U.S. (10).

Common foodborne pathogens

Pathogenic *E. coli*

E. coli is a Gram-negative, rod-shaped, facultative anaerobic and nonsporulating coliform bacterium that can commonly be found in the gut microbiota of animals and humans (11, 12). More than 700 *E. coli* serotypes have been identified, which are classified based on a combination of somatic (O), flagellar (H), and capsular (K) surface antigens (11, 13). *E. coli* are commonly utilized as indicators of fecal contamination in freshwater because they can be released into the environment from fecal matter (14). The optimum growth temperature for *E. coli* is 37 °C. Some serotypes, such as *E. coli* O157:H7, can multiply at temperature as low as 7 °C (15) or up to 49 °C (16). *E. coli* can survive in the environment with pH > 5.4, 6% sodium chloride, or aw > 0.95 (17). *E. coli* strains are commonly found living in the intestine as commensal bacteria (18). They can prevent the colonization of pathogenic bacteria in the host intestine and produce vitamin B2 (riboflavin) and K2 (menaquinone) to benefit the host (19).

Unfortunately, some *E. coli* strains are pathogenic to humans after consumption of contaminated foods. Pathogenic *E. coli* are grouped into six main pathotypes, including Enterohemorrhagic *E. coli*/Shiga-toxin-producing *E. coli* (EHEC/STEC, also known as Verocytotoxin-producing *E. coli* [VTEC]), Enteropathogenic *E. coli* (EPEC), Enteroinvasive *E. coli* (EIEC), Enteroaggregative *E. coli* (EAEC), Enterotoxigenic *E. coli* (ETEC), and diffusely invasive *E. coli* (DAEC). Amongst these, EHEC/STEC can cause severe human diseases with a low infection dose of less than 100 cells (20). STEC can be transmitted to humans by the consumption of contaminated foods such as raw meats, milk, or fresh produce. More than 100 serotypes of STEC have been identified as associated with human diseases (21). The initial symptoms of STEC infections are vomiting, abdominal cramps, and diarrhea

(often bloody). Severe symptoms can progress to hemolytic uremic syndrome (HUS), resulting in kidney failure and even death (12). More than 112,000 non-O157 STEC infections are reported annually in the U.S., with a hospitalization rate of 12.8% (22), accounting for $23.9 million of economic loss (23). The most frequently reported serotype of STEC implicated in foodborne outbreaks is *E. coli* O157:H7. Some 63,153 cases of *E. coli* O157:H7 infection are reported annually to the CDC with a hospitalization rate of 28.8% and an annual economic cost of $254.8 million (23).

Mechanisms of STEC infection and its virulence genes

Most foodborne outbreaks occurring in developed countries involve infections by STEC or EHEC (24). In the United States, STEC serotype O157:H7 is a predominant cause of foodborne outbreaks. Almost all STEC O157:H7 contain pO157, a virulence plasmid of 92 kb, which can encode several virulence factors (25). Nevertheless, the most important virulence factor of STEC is Shiga toxin (Stx)/verocytotoxin encoded by a lambdoid bacteriophage (26). Stx contains two subgroups, Stx1 and Stx2. They can be generally found in various STEC isolates. Stx2 is more prevalent than Stx1 in STEC infection, causing HUS and hemorrhagic colitis (24).

To infect hosts, STEC can induce attaching and effacing lesions on epithelial cells of the intestine. Once attached, the pathogens can efface the microvilli and overturn the actin of the host cell to form pedestals on the cell (27). This action is encoded by the locus of enterocyte effacement (LEE) genes on pathogenicity islands (PAIs) located on plasmids of STEC. The LEE highly encodes the Type III secretion system (T3SS) to secrete effectors and translocate the effectors into the cytoplasm of the host cell. These effectors can impact activities of exchanger Cl– –OH– and Na+ –H+, prevent sodium-d-glucose cotransporter 1, and mis-localize aquaporins (24). STEC attachment to the host is mediated through the connection between intimin and the translocated intimin receptor (Tir). Interaction of intimin with nucleolin can enhance the intimate attachment. The nucleolin is an intimin receptor on the cell surface whose expression can increase because of the presence of Stx2. Tir is linked to the Tir cytoskeleton-coupling protein (TccP) through insulin receptor tyrosine kinase substrate (IRTKS) of host

protein. Then, the actin-related protein 2/3 (ARP2/3) complex and the Neural Wiskott–Aldrich syndrome protein (N-WASP) are activated by interaction with TccP to mediate actin rearrangements, which are necessary to activate pedestal formation. In addition, STEC applies the *E. coli* hemorrhagic coli pilus (HCP) and the common pilus (ECP) to attach to the large bowel cells. After attaching to host cells, STEC can control the cell processes through injecting many effectors into the cells. STEC lacks a mechanism to secrete Stx, but it applies phage-mediated cell lysis to release Stx. Thus, applying antibiotics to damage STEC may lead the release of Stx in the host. After release, the Stx can be taken up by micropinocytosis in human intestinal cells or bind to the Stx receptors, globotriaosylceeramides (Gb3s), which can be found on the surface of Paneth cells and kidney epithelial cells in the human intestinal mucosa. The internalized Stx can be trafficked into the Golgi through endosomes, which can lead to cell necrosis or death (24).

Virulence factors in STEC include Stx, intimin, and enterohaemolysin (Cointe et al., 2018). STEC is defined by the presence of the virulence genes *stx*1 and *stx*2, borne by a prophage that can encode Shiga toxins (Stx1 and Stx2) (28, 29). In addition to the Stx-encoded virulence genes, most STEC strains contain genes coding intimin and T3SS related to a locus of enterocyte effacement (LEE) on the chromosome, which are *eae* and T3SS-encoding genes (30). STEC strains also harbor the plasmid-borne virulence genes *ehx*A/*hly*A encoding enterohaemolysin (31). Thus, STEC strains acquire the major modes of virulence gene acquisition through prophage, PAI, and plasmid.

L. monocytogenes

L. monocytogenes is a Gram-positive, facultative anaerobic, and rod-shaped bacterium that is ubiquitous in the environment (32). *L. monocytogenes* can grow at temperatures ranging from 1 to 45°C, with the optimum temperature ranging from 30 to 37 °C (33). It can grow and survive within a wide range of pH, from 4 to 9 (33), and in high salt conditions with up to 10% NaCl (34). In addition, *L. monocytogenes* can form biofilms on the surface of various materials such as polystyrene, glass, and stainless steel (35), posing a challenge to cleaning and sanitizing programs (36). *L. monocytogenes* contains 13 serotypes distributed in four lineages (37). Lineage I consists of 1/2b, 3b, 4b, 4d, and 4e,

which are frequently involved in most human listeriosis outbreaks. Lineage II includes 1/2a, 1/2c, 3a, 3c, and 4a, accounting for animal listeriosis cases and sporadic human clinic cases (37). Lineages III and IV are rare in nature (37). Through the consumption of contaminated foods, *L. monocytogenes* can infect humans. Infection by *L. monocytogenes* is called listeriosis, which is a severe invasive illness with a high mortality rate (20-30%) (38). Listeriosis typically causes septicemia, meningitis, and severe outcomes such as stillbirths and death. Certain populations, including children, the elderly, pregnant females and their fetuses, and immunocompromised individuals, are at increased risk of infection by *L. monocytogenes* (38). In the U.S., *L. monocytogenes* is implicated in approximately 1,600 illnesses of listeriosis annually, with the highest hospitalization rate (94%) and a high death rate (15.9%) compared to other foodborne pathogen illnesses (22). Listeriosis is responsible for $2.6 billion of economic loss per year (23).

Mechanisms of *L. monocytogenes* infection and its virulence genes

After the consumption of contaminated foods, *L. monocytogenes* can attach to liver and spleen tissues by passing through the intestinal barrier and spreading through the lymph node network and the bloodstream (39). In healthy individuals, *L. monocytogenes* usually causes mild symptoms. In immunocompromised humans as well as infants, pregnant women, or the elderly, *L. monocytogenes* can cause potentially deadly meningitis, sepsis, or even abortion by traversing the blood-brain barrier or the fetoplacental barrier (40). For non-phagocytic cells, e.g. epithelial cells, *L. monocytogenes* enters cells by binding Internalin A (InlA) and InlB to the host cell membrane receptor (receptor-mediated endocytosis) and is internalized into the vacuole (41). In most cases, *L. monocytogenes* can escape from the vacuole by physically breaking the vacuole membrane. In some macrophages, *L. monocytogenes* is internalized into spacious *Listeria*-containing phagosomes (SLAPs) and reproduces in the SLAPs. The escape of *L. monocytogenes* from vacuoles is mediated by multiple potent virulent factors, including phospholipase A (PlcA), PlcB, and listeriolysin O (LLO) (42). After its escape from vacuoles, *L. monocytogenes* polymerizes the actin through actin assembly-inducing protein, ActA, which can make it available to

spread from cell to cell (40). Extracellular LLO produced by *L. monocytogenes* can cause pore formation on the cells. Entry into the cells by those LLO can lead to mitochondrial morphology changes and fission, histone modification, desumolylation, lysosomal permeabilization, and endoplasmic reticulum (ER) stress. In the cytosol of host cells, *L. monocytogenes* can produce *Listeria* nuclear targeted protein A (LntA), which can depress interferon-stimulated genes (ISGs) by interacting with the Bromo adjacent homology domain-containing 1 protein (BAHD1) (40). LntA can change chromatin packing, leading to alteration of downstream gene expression and DNA damage in the nucleus. To combat infection, host cells increasingly produce antimicrobial effectors, e.g., ISG15. This effector can modulate cytokine secretion, such as interleukin-6 (IL-6) and IL-8, by modifying covalent bonds of Golgi proteins and ER (40).

The virulence genes of *L. monocytogenes* play important roles at various stages of the infection process (43). In the early stages, genes of the internalins, such as *inl*A, *inl*B, *inl*F, and *inl*J, highly associate with attachment and invasion. Genes located in the major *prf*A-regulated virulence gene cluster (pVGC), also referred to as *Listeria* pathogenicity island 1 (LIPI-1) play critical roles in intracellular pathogenesis (44). The pVGC comprises the *prf*A, *hly*, *plc*A, *plc*B, *act*A, and *mpl* genes. The *prf*A gene encodes PrfA protein that isis needed for the transcription of *prf*A itself and pVGC. The *hly* gene encodes the LLO that can induce pore-formation on the cell membranes and lyse phagocytic vacuoles containing bacteria to release the bacteria into the cytoplasm of host cells. The *plc*A and *plc*B genes encode the phosphatidylinositol-specific phospholipase C (PI-PLC) and zinc-dependent broad-spectrum phospholipase C (PC-PLC), respectively. Similarly, the PI-PLC and PC-PLC can mediate the escape of *L. monocytogenes* from the vacuoles. After escape from the vacuoles, the actin polymerization of surface protein actin A (ActA) can modulate the cell-to-cell dissemination and intracellular motility. Besides, the ActA has other functions,including colonization, invasion, persistence, and aggregation in the human intestinal lumen, (45). The *mpl* encodes a zinc metalloproteinase to activate PC-PLC that can start a new cycle of infection (46).

Salmonella

Salmonella is a Gram-negative, flagellate, facultative anaerobic, and rod-shaped, microorganism (47). *S. enterica* can grow at temperatures ranging from 8 to 45 °C, pH ranging 4 to 9.5, and low water activity as low as 9.4 (48). *S. enterica* is considered to have some strains that are able to survive at cold temperature (2 °C) or high temperatures (54 °C) (49). *Salmonella* can survive at low pH (4-5) and thus may tolerate the harsh acidity conditions in the stomach (49). In addition, *S. enterica* is not able to grow at a low water activity ($a_w < 0.94$) but can survive for a long period in low-a_w foods ($aw \leq 0.85$) (50, 51). The *Salmonella* genus contains two species (*S. enterica* and *S. bongori*) six subspecies (*enterica*, *arizonae*, *salamae*, *diarizonae*, *indica*, and *houtenae*), and more than 2,579 serovars (49). *S. enterica* contain most of the serotypes that have been implicated in *Salmonella* outbreaks (52). *S. enterica* serovars can be found in the intestinal tracts of birds, insects, reptiles, farm animals, and humans (49). As *S. enterica* do not originate from water, the presence of *S. enterica* in water indicates fecal contamination (53). *S. enterica* are transmitted to infect humans through the consumption of contaminated foods or water. The illness, called salmonellosis, can cause typical gastroenteritis. More invasive infections can occur in at-risk individuals, including children, the elderly, and immunocompromised individuals. In the U.S., more than one million non-typhoidal *S. enterica* illnesses are reported each year, with the hospitalization rate estimated at 27.2% (23). The illness accounts for $3.3 billion of economic loss, costing more than any other major foodborne pathogen (23).

Mechanisms of *S. enterica* infection and its virulence genes

S. enterica with serovars that can cause human disease are classified into typhoidal and non-typhoidal groups. The typhoidal *S. enterica* serovars consist of serovar Typhi and serovars Paratyphi A, B and C, which cause a systemic disease called enteric fever in humans. The most common non-typhoidal *S. enterica* (NTS) serovars contain Enteritidis and Typhimurium. The NTS has a wide vertebrate host range and cause an invasive and frequently fatal disease in immunocompromised individuals (54). After ingestion of *S. enterica*- contaminated food, *S. enterica* subsp. *enterica* serovar Typhimurium comes in

contact with the human intestinal epithelium, which then activates the T3SS on the *Salmonella* pathogenicity island 1 (T3SS1) (55). *S. enterica* can invade intestinal microfold (M) cells and then enter the intestinal submucosa. Subsequently, *S. enterica* is up-taken by dendritic cells and macrophages to form *Salmonella*-containing vacuoles (SCVs). The intracellular environment triggers *S*. Typhimurium to activate the T3SS on its pathogenicity island 2 (T3SS2) (56). The activation of T3SS2 allows the pathogen to avoid innate immune defense reactions and enables *S. enterica* to reproduce in the SCVs. *S. enterica* that can survive in SCVs can migrate to the mesenteric lymph nodes. Both intracellular and extracellular *S. enterica* can reproduce in the cytosol of infected cells. The invasive infection can spread over the host cells through the bloodstream and lymph network. Finally, *S. enterica* can reproduce in the reticuloendothelial system (54).

S. enterica infection in the human intestine is driven by prophage-associated genes encoded on the *S. enterica* pathogenicity islands (SPIs) (57).When *S. enterica* colonizes the intestinal lumen, inflammation is activated by *sop*E genes that are encoded on the prophage, effectors that are secreted by SPI-1, and a type 6 secretion system (T6SS) that is encoded on the SPI-6, which favor outcompeting commensal microbes (57). The two component systems, PhoPQ and PmrAB, can modulate resistance to antibacterial compounds generated by the human intestinal lumen (58). Attachment of *S. enterica* to epithelial cells is driven by surface proteins that are encoded on the SPI-3 and SPI-4 (59). The internalization of bacteria into epithelial cells is activated by genes encoded on SPI-1 and SPI-5 (59). *S. enterica* survival in SCVs is mediated by the Gifsy-1 prophage and the SPI-1. Resistance to immune responses such as oxidative stress in the immune cells and the SCV, such as macrophages and neutrophils, is driven by the Fels-1 and Gifsy-2 prophages. In addition, to avoid immune responses, SPI-11, SPI-12, and SPI-16 can encode genes to reconstruct the outer membranes of *S. entercia*. The survival and reproduction of *S. enterica* in the macrophages are regulated by genes encoded on the SPI-2, SPI-5, and SPI-13 (57). Taken together, the virulent genes summarized here play critical roles in colonization and infection by *S. enterica.*

Part 2. Survival of *Escherichia coli* O157:H7, *Listeria monocytogenes*, and *Salmonella enterica* in fresh produce

Introduction

Outbreaks of foodborne pathogen infections linked to fresh produce are caused by the consumption of pathogen-contaminated produce products. Contamination of foodborne pathogens in fresh produce can occur at any point from farm to fork. Lynch et al. (2009) summarized three points where fresh produce is most probably to be contaminated: the production field, the processing step, and the household processing stage (60). Reported sources of foodborne pathogen contamination during preharvest include soil, irrigation water, inadequately composted manure, human handling, wild and domestic animals, reconstituted fungicides and insecticides, dust, insects, and seasons. Reported postharvest sources of contamination include harvesting equipment, water used for washing and rinsing, human handling, transport vehicles, conveyor belts, processing equipment, poor hygiene in plant workers, improper packaging and storage, cross-contamination, and consumer handling (6, 61). Foodborne pathogens can be transmitted from the sources to fresh produce during preharvest or postharvest and survive or even grow in fresh produce when conditions permit, then potentially causing foodborne infections or even outbreaks if the contamination is not eliminated or well controlled before consumption. The most frequently reported foodborne pathogenic bacteria related to fresh produce are *E. coli* O157:H7, *L. monocytogenes*, and *S. enterica*. Therefore, understanding the survival of these pathogenic bacteria in fresh produce and the factors significantly impacting the fitness of foodborne pathogens is critical to develop effective strategies to control the microbial safety of fresh produce.

Factors impacting survival of pathogens in fresh produce

The incidence, survival, persistence, and even growth of foodborne pathogens in fresh produce are determined by various factors across the fresh produce supply chain, including growing, harvesting, processing, distribution, storage, and food handling. Generally, foodborne pathogens usually survive but do not grow on the intact surface of fresh produce (6, 62, 63), whereas the population of foodborne

pathogens may grow on fresh produce when nutrients are released because of bruised produce, fungal rot, or physical damage (64). Previous literature reviews have summarized major factors that play critical roles in influencing the contamination and cross-contamination potentials, incidence, and survival of human pathogens in fresh produce during preharvest, harvest and processing, distribution and storage, and preparation stages (6, 61, 62, 65–73).

During the preharvest stage, the incidence and fitness of foodborne pathogens can be influenced by complex factors, such as produce type, soil type, irrigation water, farming location, animals and insects, production systems, and climatic conditions. Specifically, produce types include cultivation methods, produce features, and plant immunity (65). Soil types involve pH, texture, basic cation saturation ratio, nutrient levels, and application of soil amendments (62, 71, 73). Survival of pathogens in irrigation water is related to water source, the irrigation system, and use of water treatment methods. Farming location refers to distances from surrounding animal farming or production areas (71). Animals and insects consist of animal or insect types, activity frequency, and animals and insects carrying pathogens (62). Production systems include conventional, organic, or hydroponics, and the application of pesticides and insecticides. In addition, climatic conditions are associated with temperature, humidity, sunlight/UV, and rainfall (66).

During the harvest and processing stage, factors include the primary preparation, such as cleaning, trimming, coring of raw materials (66); processing steps, such as sanitizing, washing, drying, mixing, waxing, packaging (70, 74); the washing process, such as washing time, washing agents, and washing equipment (66, 75); the packaging process, such as packaging materials, packaging types, and atmosphere compositions (76, 77); and worker hygiene, such as good manufacturing practices, handwashing, and personal protective equipment (78). These factors related to fresh produce can not only impact the populations and diversity of fresh produce microbiota but also impact the incidence and survival of foodborne pathogens (75, 79).

During the distribution and storage stage, temperature has been considered as one of the most important factors that influence the survival of foodborne pathogens in fresh produce (65). Refrigerator

temperature (at or below 4 °C) is commonly used to control microbial multiplication. The low temperature can reduce the cell activity of large numbers of bacteria, including some foodborne pathogens such as *E. coli* O157 and *S. enterica*. However, it is known that *L. monocytogenes* is a psychrotrophic bacteria that can grow at a temperature as low as 1 °C (80), thus low temperature may not prevent the growth of *L. monocytogenes* in fresh produce during distribution and storage (65). Other intervention approaches may be needed in combination with low temperature to control or mitigate the pathogen contamination, multiplication, and infection risk. Temperature abuse during product distribution and storage stage can lead to remarkable increases of total bacterial populations, including the levels of foodborne pathogens. In addition, the storage time of fresh produce before consumption also influences the survival of pathogens in fresh produce (66).

During the preparation stage, the hygiene of household or foodservice establishment personnel and the fresh produce handling environment are important factors impacting the survival of pathogens. Fresh produce handling, such as cutting, sorting, washing, mixing, and cooking, also determine the contamination or survival of foodborne pathogen in fresh produce (66). A couple of key intrinsic factors that can play major roles in determining the fitness and behavior of pathogenic bacteria include produce types and the presence of native microbiota. The produce types determine its topographic structure, preparation and distribution method, consumption method as well as the native microbiota associated with them. The involvement of native microbial populations in the microbial safety control of fresh produce spans from the preharvest to postharvest stages. Native microbes can compete with external pathogens for attachment space and nutrients, and/or generate antimicrobial substances that can also impact the incidence or survival of pathogens in fresh produce (77, 81).

Survival of pathogens in fresh produce

The following section summarizes papers published in the recent five years (2018-2022) related to the fitness of three common foodborne pathogens, *E. coli* O157:H7, *L. monocytogenes*, and *S. enterica*,

in fresh produce. Key factors impacting pathogen survival also were summarized. The Web of Science was used as the database for paper search. The following key words/topics were used during the search process: Topic 1: '*Escherichia coli* O157 OR STEC' AND 'survival OR persistence' AND 'fresh produce OR vegetables OR fruits'; Topic 2: '*Listeria monocytogenes*' AND 'survival OR persistence' AND 'fresh produce OR vegetables OR fruits'; Topic 3: '*Salmonella*' AND 'survival OR persistence' AND 'fresh produce OR vegetables OR fruits'. Year: last five years. A total of 75 papers were identified from the database, including 15 associated with *E. coli* O157/non-O157 STEC growth and survival, 22 associated with *L. monocytogenes* growth/survival, and 38 associated with *Salmonella* growth/survival.

Survival of *E. coli* O157:H7 in fresh produce

Fifteen studies from 2018 to 2022 investigating *E. coli* O157 survival on fresh produce were selected and summarized for this review (**Table 4**). Among these studies, five investigated the survival of *E. coli* O157:H7 in various kinds of fresh produce during the preharvest stage, and 10 of them studied the survival of *E. coli* O157 in fresh produce during the postharvest phase. One study addressed the survival of *E. coli* O157:H7 in fresh produce during both preharvest and postharvest stages. These studies covered various types of fresh produce, including fresh vegetables (such as Romaine lettuce, cabbage, spinach, butter lettuce, basil, cilantro, fresh herbs, chili habanero, cucumber, radish, beet, onion, tomato, and spring mix), and fresh fruits (including kiwifruits and strawberry).

The survival of *E. coli* O157:H7 in Romaine lettuce, spinach, and cabbage was studied under field conditions or growth chamber conditions (82–86). Results showed that 3-6 Log CFU/plant of *E. coli* O157 decreased to the level below the limit of detection (0.5 Log CFU/plant) in about 10 days post-inoculation in all types of vegetables grown in the field (**Table 4**). One critical new finding among these studies was the impact of pathogens' surface polysaccharides on the plant defense system. Jang et al. (2018) found that the polysaccharides on the *E. coli* O157 cell surface can induce the plant defense response which then lead to enhanced inactivation of more *E. coli* O157:H7 cells (87). Erickson et al. (2019a) illustrated that while the cultivar of lettuce had no effect on the fate of *E. coli* O157 post-

inoculation, the maturity stage of cabbage could impact the survival of *E. coli* O157 (82). Gutiérrez-Rodríguez et al. (2019) investigated the impact of initial inoculation levels (2 and 4 Log CFU/ml) on the survival of *E. coli* O157 in spinach cultivated under field conditions (84). No obvious difference in the survival levels of *E. coli* O157 in spinach on day 14 after inoculation was observed.

For fresh produce during postharvest, 10 studies investigated *E. coli* O157 survival or growth in a variety of fresh produce at a temperature from 3 to 35 °C (**Table 4**). These studies showed that produce types, temperatures, storage time, structures of the fruit surface, and fruit cultivars play critical roles in influencing the survival of *E. coli* O157 in fresh produce during storage. In general, storage at low temperatures (3-7 °C) lead to the gradual reduction of *E. coli* O157 by 0.5 to 4 Log CFU/g during up to 21-day of storage (81, 88–92).

Survival of *L. monocytogenes* in fresh produce

A total of 22 studies evaluated the behavior of LM in fresh produce in recent 5 years and they are summarized in **Table 5**. Among these studies, four studied the survival of *L. monocytogenes* in fresh produce during the preharvest stage while the rest of the studies addressed the survival of *L. monocytogenes* during the postharvest stage, including six related to fresh vegetables and 12 associated with fresh fruits.

The four studies related to the preharvest investigated the survival of *L. monocytogenes* in lettuce leaves, alfalfa sprouts, and herb plants under field conditions (ca. 20 °C) or in the growth chamber for up to 10 weeks (93–96). Kyere et al. (2021) found that the *L. monocytogenes* levels increased in lettuce on Day 4 post-inoculation at 20 °C in the field (95), while Honjoh et al. (2018) showed that the *L. monocytogenes* level fell below the limit of detection (2 Log CFU/g) in lettuce leaves after 10 weeks of cultivation under the same condition (94). The strains of *L. monocytogenes* and inoculation levels had no significant impact on the survival of *L. monocytogenes* in fresh produce during preharvest (94). Leafy damage could extend the persistence of *L. monocytogenes* in fresh produce (94). Produce types

dramatically impacted the *L. monocytogenes* survival. For example, Bardsley et al. (2019) showed that *L. monocytogenes* populations reduced to ca. 1 Log CFU/g 7 days after the inoculation (93), while Rossi and Lathrop (2019) reported that *L. monocytogenes* remarkably increased from 0.7 and 2.7 Log CFU/g to 6.5 to 7.0 Log CFU/g in alfalfa after 5 days post-inoculation (96).

The survival of LM during postharvest at postharvest handling temperatures from -20 to 30 °C has been evaluated. Studies showed that low storage temperatures (4-8 °C) had limited impact on the levels inoculated LM (97–105). It is known that *L. monocytogenes* is a psychrotrophic bacteria that can grow at temperatures as low as 1 °C. The growth potential of *L. monocytogenes* is significantly impacted by pH. Olaimat et al. (2021) found that LM gradually reduce when the pH of produce was below 4, for example the pH of salad mixed with salad sauce (Toum), or acidic fruits such as fresh-cut kiwifruit, pineapple, white peach, yellow peach, and yellow nectarine (102).

Survival of *Salmonella enterica* in fresh produce

There were a total of 36 studies identified from the recent five years that investigated the survival of *Salmonella* in fresh produce during preharvest and postharvest. Among them, five were related to *Salmonella* survival in fresh produce during preharvest, and the others were associated with fresh produce during postharvest (**Table 6**).

The four preharvest studies showed the survival of *Salmonella* in fresh produce cultivated under field conditions or in growth chambers. These studies studied the impact of cabbage cultivars, lettuce cultivars, and lettuce sizes on the fitness of *Salmonella*. In addition, the impact of inoculation levels was also investigated. Studies conducted by Erickson et al. (2019a) (82), Erickson et al. (2019b) (83), Oblessuc and Melotto (2020) (106), and Xylia et al. (2022) (107) all indicated that the inoculation levels had a remarkable impact on the survival of *Salmonella* in fresh produce during cultivation. In these studies, the amount of *Salmonella* reduced by 1-6 Log CFU/plant in cabbage and lettuce based on the

fresh produce type and cultivar. Conversely, *Salmonella* in alfalfa sprouts grown in the incubator at 20 ℃ showed a dramatic increase in concentration after 5 days of cultivation (96).

Studies related to fresh produce during the postharvest investigated the survival of *Salmonella* in fresh produce stored at different temperatures from 3-37 ℃ (**Table 6**). These results showed that different factors influence the survival of *Salmonella* in fresh produce during the storage. These factors include produce type, temperatures, storage time, and inoculation levels. In addition, Tokarskyy et al. (2018) illustrated that the ripeness stages of tomato significantly impacted the survival of *Salmonella* (108). Specifically, *Salmonella* had higher levels in red tomatoes than green and pink (ripeness stages) tomatoes at the end of seven days of storage at 10 and 20 °C. Chen and Ming (2021) showed that desiccation stress also played a significant role in determining the survival of *Salmonella* (109). The desiccation-stressed *Salmonella* had enhanced ability to survive for seven days in baby spinach under cold conditions or temperature abuse conditions. Singh et al. (2021) showed the effect of relative humidity (55% and 90% RH) and fruit ripeness on the level of reduction of *Salmonella* in papaya stored at 4 ℃ for 14 days. Papaya at 0% and 25% ripeness stored at 55% RH had greater reduction of the *Salmonella* population after 14 days of storage at 4 °C compared to papaya stored at 90% RH (110).

In summary, *E. coli* O157, *L. monocytogenes*, and *Salmonella* can contaminate fresh produce at any point during preharvest and postharvest. Once contaminated, pathogens in fresh produce can survive for an extended period, posing potential threats to public health. Understanding factors impacting the fitness of pathogens in fresh produce at different production stages is the foundation for developing efficient intervention strategies for protecting and improving the microbial food safety of fresh produce.

Part 3. Biological pathogen mitigation strategies for fresh produce

Introduction

Overall, 129 outbreaks of pathogenic bacteria infections linked to fresh produce occurred from 2010 to 2021 in the U.S., causing a total of 8,889 illnesses with 2,412 hospitalizations and 90 deaths (CDC, 2022). These data suggest that current fresh produce safety control strategies may be limited in controlling foodborne pathogens. Due to being cultivated in open environments, fresh produce can be contaminated by microbial pathogens at any point as the contamination sources during the preharvest are rich, such as soil, irrigation water, inadequately composted manure, human handling, wild and domestic animals, reconstituted fungicides and insecticides, dust, insects, and seasons (6, 61). Conventional chemical and physical control methods have been commonly used to control microbes in the food industry, such as chlorine, acidified sodium chloride, peracetic acid, electrolyzed water, ozone, hydrostatic pressure, ultraviolet, ultrasound, pulsed electric fields, pulsed light, ionizing radiation, and modified atmosphere packaging (111–115). However, these methods contain some drawbacks, including possible adverse effects on human health and the environment as well as on the sensory qualities and nutrients of food and the possible inducement of antibiotic resistance genes, all of which can limit their use as safety control approaches in the food industry (116). In addition, chemical and physical control methods are usually carried out for one-time treatment for fresh produce during produce processing (117). Although microbial flora in produce can be disinfected in different degrees after the chemical and physical treatments, the surviving microbes can re-grow under proper conditions during the following transportation and storage phases (75, 117). To solve these issues, natural, environmentally friendly, and cost-effective alternative approaches are needed. This section reviews the key approaches developed, with an emphasis on recent developments in competitive exclusion cultures for produce safety.

Biological methods have received increasing attention in recent years due to several advantages associated with them: first of all, these natural food safety control methods are considered health-friendly and may have a lower negative impact on the sensory and nutritional properties of fresh produce

compared to physical and chemical decontamination approaches; secondly, the cost associated with food processing generated by biological control can be less than chemical and physical control since no advanced equipment is required; and thirdly, the use of biocontrol has less impact on the formation of antibiotic resistance than chemical control (e.g. antibiotics) in foodborne pathogens. In addition, biological control methods can be consistently effective in antagonism against foodborne pathogens, continuously protecting fresh produce safety after treatment (111). The biocontrol agents include antagonistic microorganisms (e.g. competitive exclusion bacteria), antimicrobial substances (e.g. bacteriocin, endolysins, and peptides) generated by these microorganisms, and plant-derived natural compounds (111). The important criteria for developing biocontrol agents include: first, the agents should have high antagonistic activity against foodborne pathogens; second, the agents should be safe for human health; and third the agents should not alter the sensory properties of fresh produce (118). Among the biological agents, bacterial antagonism, also known as "bacterial interference" or "competitive exclusion," is a method to apply non-pathogenic bacteria to reduce colonization and inhibit pathogenic bacteria in an environment of animal intestines, plants, or food products (119). Nurmi et al. (1992) were the first ones to report the application of antagonist to control *Salmonella* contamination in poultry after outbreaks of *S.* Infantis infections occurred in 1971 (120). Over the years of development, antagonistic cultures have been used as additives in the feeds of domestic animals to inhibit pathogens in animal guts (121). Commercial antagonistic products have been applied to treat domestic animal and plant diseases (122, 123). In the past decade, the application of antagonistic cultures has spread from the domestic animal field to the food industry (124, 125). The aim of this section is to discuss major antagonist isolation methods, review dominant strains that have been isolated from food sources with antimicrobial effects over common foodborne pathogens, and describe their antimicrobial mechanisms and antagonistic activities against foodborne pathogens in fresh produce.

Antagonist isolation and antimicrobial activity test

Antagonistic bacteria inhabit various environments, including soil, water, plants, animals, vegetables, fruits, fermented food, food processing plants, or even animal wastes (126–130). Culture-dependent methods are the essential techniques for isolation of Antagonist and testing their antimicrobial activity. The commonly used techniques include agar disk diffusion, agar well diffusion, and spot-on-lawn methods, with the spot-on-lawn approach being the most common for isolating antagonistic cultures and determining their antagonistic activities. The agar-disk-diffusion method is a well-known procedure for determining routine antimicrobial susceptibility in many clinical microbiology laboratories. First, the agar plates are inoculated with the test microbes, e.g., foodborne pathogens. Second, paper discs (ca. 6 mm in diameter) of test antimicrobial compounds or antibiotics in certain concentrations are placed on the surface of the agar plates. After incubation, the antimicrobial-inhibiting zones around the discs form. The combination of ability of the antimicrobial to diffuse into the agar and to inhibit the growth of the organism into the agar is evaluated based on sizes of the antimicrobial inhibiting zones. However, since the inhibition does not represent microbial death, this approach cannot reveal the difference between bactericidal and bacteriostatic effects (131). The agar-well-diffusion method is also widely used to test the antimicrobial ability of extracts from plants or microbes. Agar well diffusion also requires a volume of tested microbes to be spread on the agar plates first. Then, a hole with a diameter of about 6 mm is made by using a sterile cork borer or a tip. About 20-100 µL of the extract solution will be filled into the hole. The agar plates are incubated at a certain temperature for a time based on the microbial inoculum. After incubation, the extract solution diffuses in the agar, and an antimicrobial inhibition zone is generated. The antimicrobial ability can be evaluated based on measuring the diameter of the inhibiting zone (131). The spot-on-lawn method is widely used to assess the antimicrobial effect of microorganisms. First, the overnight fresh bacterial culture is prepared and added to about 5 ml of 0.75% soft agar. The soft agar mixed with bacterial culture is pulled onto an agar plate. Once the soft agar is solidified, 10 µl of fresh overnight antagonistic culture or 2-5 µl of supernatant are spotted on the double-layer agar plate. The agar plate is incubated at a certain temperature and time. After incubation, the antimicrobial inhibiting zone

can be observed. The antimicrobial effect can be evaluated by measuring the diameter of the inhibiting zone (132, 133).

Based on the culture-dependent methods of antagonist isolation and antagonistic activity determination, an increasing number of antagonists have been isolated and identified from fresh produce and fermented vegetable sources that demonstrate antimicrobial activities against foodborne pathogens. Lopez-Velasco et al. (2012) first plated fresh spinach homogenates onto R2A media and randomly picked 1512 isolated colonies after incubating at 25 °C for 48 h (128). Afterwards, the spot-on-lawn method was employed to determine the inhibitory abilities of potential bacterial isolates against *E. coli* O157:H7. Briefly, 100 µl of *E. coli* O157:H7 at 6 Log CFU/ml were plated on the R2A plates. The isolated fresh spinach colonies were transferred onto the R2A plates with *E. coli* O157:H7 lawns using sterile toothpicks. After incubation at 25 °C for 48 h, the isolates generating inhibition zones with diameters greater than 5 mm were considered as antagonist with active antagonism against *E. coli* O157:H7. The antagonist isolates were then identified based on 16S rRNA gene sequence using amplified fragments with 16S rRNA gene primers (27f and 1392r). In total, 23 species were identified, including *Stenotrophomonas maltophilia*, *Rhodococcus* spp., *Pseudomonas* spp., *P. rhodesiae*, *P. koreensis*, *P. fluorescens*, *Pantoea agglomerans*, *Paenibacillus polymyxa*, *Microbacterium phyllospherae*, *M. oleivorans*, *Kaistia* spp., *Frigobacterium* spp., *Flavobaterium* spp., *Erwinia rhapontici*, *E. persicina*, *Curtobacterium herbarum*, *Brevundimonas vesicularis*, *Bacillus* spp., *B. pumilus*, *B. cereus*, *Arthrobacter* spp., *Aeromonas encheleia,* and *Acinetobacter calcoaceticus*.

Tinrat and Sedtananun (2022) isolated lactic acid bacteria (LAB) from pickled cabbage, canned pickled cabbage, kimchi, and pickled bamboo shoots (134). The homogenized samples were inoculated on deMan, Rogosa and Sharpe (MRS) agar using the pour-plating method. After incubation at 37 °C for 48 h, 40 colonies with different morphologies were picked and streaked on MRS to purify cultures. Subsequently, the agar-well-diffusion method was applied to test the antimicrobial activity of the isolates against *E. coli* ATCC 25922. Briefly, the overnight *E. coli* culture was adjusted to obtain a cell

concentration at approximately 8 Log CFU/ml in 0.75% soft tryptic soy agar (TSA). The soft agar was then poured onto 1.5% TSA agar plates. Supernatants of overnight isolates in MRS broth (80 µl) were added into the 5 mm in diameter agar well made by a cork borer. After incubation at 37 °C for 24-48 h, the diameters (mm) of inhibition zones around the wells were measured. The supernatants of isolates having inhibition zones with diameters greater than 1 mm were considered as lactic acid bacteria (LAB) with positive antagonistic activity against the *E. coli* strain. The LAB isolates were identified using 16S rRNA gene sequencing based on the PCR amplicons using the universal oligonucleotide primers (UFUL and URUL) (134). After sequencing data analysis, *Rummeliibacilus pycnus* SCTB112 and *R. stabekisii* B1-37c-21 were identified.

Lactic acid bacteria

LAB are genetically diverse group in the Firmicutes phylum that produce lactic acid as a result of fermentative energy conservation metabolism. The five core genera are *Lactobacilli*, *Lactococcus*, *Leuconostoc*, *Pediococcus*, and *Streptococcus*. LAB also include *Weissella*, *Carnobacterium*, *Enterococcus*, *Oenococcus*, *Sporolactobacillus*, *Tetragenococcus*, *Aerococcus*, and *Vagococcus* (135). Certain *Lactobacillus* and *Lactococcus* species are considered generally recognized as safe (GRAS), which means they can potentially be applied as biological control agents in the food industry (136). LAB are major components of antagonistic bacteria, with their major characteristic being the generation of antimicrobial compounds, including organic acids, hydrogen peroxide, carbon dioxide, bacteriocins, antifungal peptides, and redox potential (137, 138). Based on these properties, homofermentative LAB can produce lactic acid as the major metabolite to lower the pH of the environment for outcompeting other bacteria, including foodborne pathogens, while heterofermentative LAB can produce multiple metabolites, including lactic acid, ethanol, acetic acid, and carbon dioxide (139). The production of carbon dioxide can create an anaerobic environment to inhibit aerobic microbes, and generated ethanol potentially acts to denature proteins of the microbial cells until inactivating the microbes (138).

Several reviews have summarized LAB isolated from food products that negatively impact the survival or the growth of common foodborne pathogens (124, 138, 140), including *E. coli* O157:H7, *Listeria monocytogenes*, and *S. enterica*. For *E. coli* O157:H7, Ibrahim et al. (2021) summarized that *Weisella halotolerans*, *Pediococcus pentosaceus*, *Pediococcus* spp., *Acidilactici* spp., *Leuconostoc mensenteroides*, *Limosilactobacillus reuteri*, *Lactobacillus helveticus* KLDS 1.8701, *Lactobacillus paraplantarum*, *Lacticaseibacillus paracasei*, *Latobacillus coryniformis*, *Levilactobacillus brevis*, *Lactobacillus sakei*, *Lactiplantibacillus plantarum*, and *Lactobacillus curvatus* reduced the level of *E. coli* O157:H7 (140–144). For *L. monocytogenes*, Ibrahim et al. (2021) presented *Weisella halotolerans*, *Pediococcus pentosaceus* KTU05-8, KTU05-9, and KTU05-10, *Leuconostoc mesenteroides*, *Limosilactobacillus reuteri*, *Lactobacillus helveticus* KLDS 1.8701, *Lactiplantibacillus plantarum* CC10, *Lactobacillus paraplantarum*, *Lacticaseibacillus*, *Lactobacillus coryniformis*, *Levilactobacillus brevis*, *Lactobacillus curvatus*, *Lacticaseibacillus paracasei* IMC 502, *Lacticaseibacillus rhamnosus* IMC 501, *Lactobacillus sakei* KTU05-6, and *Enterococcus faecium* QPII presented inhibitory activities (140, 141, 145, 146). In addition, Webb et al. (2022) (138) summarized that *Lactococcus lactic* spp. *lactis* (147), *Lactococcus lactis* subsp. *cremoris* FT27 (148), *Enterococcus faecalis* (149), *Lactococcus psicium* EU2241 (150), and *Carnobacterium piscicola* CS526 (147) could reduce the level of *L. monocytogenes* on different food matrices, including dairy and cheese, meat products, and fish products. For *S. enterica*, Ibrahim et al. (2021) (140) reported that *Limosilactobacillus reuteri* had an antimicrobial effect on *S.* Enteritidis; and *Lactobacillus helveticus* KLDS 1.8701, *Levilactobacillus brevis* CM22, and *Pediococcus pentosaceus* CM16 showed antagonism against *S.* Typhimurium (141, 143, 151, 152).

A review by Khubber et al. (2022) (124) addressed LAB roles in fermentation or safety control in fruit and vegetable products. Firstly, *Weissella koreensis*, *Weissella cibaria*, *Leuconostoc mesenteroides*, and *Latilactobacillus* were found in fermented kimchi, which generated lactic acid, succinic acid, and phenyl lactic acid to inhibit multiple foodborne pathogens, including *E. coli* O157:H7, *S.* Enteritidis, *S.* Typhimurium, and *Staphylococcus aureus* (153). Secondly, *Apilactobacillus kunkeei*, *Leuconostoc mesenteroides*, *Pediococcus pentosaceus*, *Lactobacillus brevis*, and *Lactiplantibacillus plantarum* were

isolated from common purslane, which could inhibit *E. coli* and *Bacillus megaterium* (154). Thirdly, *Lactiplantibacillus plantarum* was isolated from mulberry juice, apple juice, and pomegranate juice (155–157). In addition, *Lactobacillus acidophilus* and *Lacticaseibacillus paracasei* were isolated from mulberry juice; *Lactobacillus helveticus*, *Lacticaseibacillus*. *casei*, *Lacticaseibacillus paracasei*, and *Lactobacillus acidophilus* were isolated from apple juice (155, 156). However, the antimicrobial abilities of these LAB against foodborne pathogens were not evaluated. In a study from Siroli et al. (2015), 39 LAB strains were isolated from minimally processed apples and lamb's lettuce, and the inhibitory activity against *E. coli*, *S.* Enteritidis, and *L. monocytogenes* were evaluated (158). *Lactobacillus plantarum* CIT3 and *Lactobacillus paracasei* M3B6 and were selected as biological control methods for the sliced apples; the *Lacticaseibacillus casei* V4B4 and *Lactiplantibacillus plantarum* V7B3 were chosen as biological control methods for the minimally processed lamb's lettuce. The apples and lettuce were treated by dipping into the solution with 7 Log CFU/ml of each LAB and 3 Log CFU/ml of each pathogen (*E. coli* and *L. monocytogenes*) and stored at 6 °C for up to 16 days. The results showed that the sliced apples treated with CIT3 reduced by approximately 0.8 Log CFU/g of *E. coli* and *L. monocytogenes* after 7 days and 16 days compared to the pathogen controls, respectively. For the lettuce, only V7B3 inhibited about 1 Log CFU/g of *L. monocytogenes* after 9 days compared to the *L. monocytogenes* control. The results indicate that the antimicrobial activity of LAB against foodborne pathogens in produce is affected by the produce types and LAB strains.

Non-LAB antagonists in food products

In addition to LAB, an increasing number of studies have focused on investigation of non-LAB conferring antagonistic activities against foodborne pathogens, their advantages of improving food quality, and their health benefits to hosts. Among the non-LAB, spore formers, e.g., *Bacillus* and *Clostridium*, are the main genera that contain members generating antimicrobial substances to inhibit foodborne pathogens. Since the spores have the properties of resistance to heat, low water activity, or high acidity, these bacteria are relatively more persistent in harsh environments (159). *Bacillus* as a gut

commensal genus contains some probiotic strains that have been used for humans and animals. The human-use *Bacillus* members include *B. cereus*, *B. subtilis*, *B. pumilus*, *B. indicus*, *B. mesentericus*, *B. coagulans*, *B. clausii*, *B. licheniformis*, and *B. polyfermenticus*, and the animal-use *Bacillus* members include *B. cereus*, *B. subtilis*, *B. licheniformis*, and *B. coagulans* (160, 161). Although some *B. cereus* strains have been used as probiotic feed additives to reduce pathogens in animal guts, some strains can cause foodborne illness in humans due to generation of enterotoxins (160). *Clostridium butyricum* is also a common human and animal gut commensal species. However, some strains are pathogenic as they can produce infant botulism and necrotizing enterocolitis (162). *C. butyricum* CM588 and UBCB-70 have been used as two commercial probiotic products in the food industry. The main safety concern of applying *Bacillus* and *Clostridium* is the potential for toxin generation. Although these probiotic strains have been tested without direct toxin production, the toxin genes may be transferred through horizontal gene transfer into other bacteria (162, 163). Several studies have shown that some strains of *Bacillus* spp. isolated from fresh produce had antimicrobial activities against foodborne pathogens (**Table 7**), reflecting their potential application as biological control agents in the food industry. Lopez-Velasco et al. (2012) isolated *Bacillus pumilus* from fresh spinach. After 24 h at 25 °C, *E. coli* O157:H7 populations declined by 1.15 and 0.88 Log CFU/ml compared to the control when co-cultured with *Bacillus pumilus* in TSB broth and on detached spinach leaves, respectively (128). *Bacillus subtitlis* LCA1 isolated from lettuce and mung bean seeds (117) reduced populations of *S. enterica* and enterohaemorrhagic *E. coli* by 1.4 and 2.0 Log CFU/g, respectively, compared to the control in mung bean seeds after 5 days at 22 °C.

In addition to *Bacillus* and *Clostridium*, *E. coli* and *Propionibacteria* have been reported to be applied as CE bacteria to inhibit foodborne pathogens (164, 165). *E. coli* Nissle 1917 (EcN) was reported to inhibit the biofilm formation of several pathogens, including enterohemorrhagic *E. coli* (EHEC), *Pseudomonas aeruginosa*, *Staphylococcus aureus*, and *S. epidermidis*. DegP, a bifunctional (chaperone and protease) periplasmic protein, was identified as the key factor to inhibit EHEC biofilm formation (164). *Propionibacteria freudenreichii* originating from dairy products is the only commercial probiotic

species, including *P. freudenreichii* W200 and *P. freudenreichii* ET-3. Yuksekdag et al. (2014) reported that *P. freudenreichii* have strong antagonistic activity against *E. coli* ATCC 11229 (165). Lopez-Velasco et al. (2012) isolated other non-LAB from fresh spinach, including *Stenotrophomonas maltophilia*, *Rhodococcus* spp., *Pseudomonas* spp., *P. rhodesiae*, *P. koreensis*, *P. fluorescens*, *Pantoea agglomerans*, *Paenibacillus polymyxa*, *Microbacterium phyllospherae*, *M. oleivorans*, *Kaistia* spp., *Frigobacterium* spp., *Flavobaterium* spp., *Erwinia rhapontici*, *E. persicina*, *Curtobacterium herbarum*, *Brevundimonas vesicularis*, *Arthrobacter* spp., *Aeromonas encheleia*, and *Acinetobacter calcoaceticus* (128). All the antagonists presented certain antagonistic activities against *E. coli* O157:H7 when co-cultured in TSB at 25 °C after 48 h. *Pseudomonas graminis* CPA-7, isolated from apple surface, exhibited strong antagonism against *E. coli* O145:H7, *L. innocua*, and *S. enterica* in both fresh-cut apples and peaches after 2 days at 20 °C (170) (**Table 7**).

Reviews of isolation of antagonistic bacteria in fresh produce

Although previous reviews have summarized numerous antagonistic bacteria including LAB and non-LAB isolated from different sources and applied to various food products (111, 118, 124, 125, 139, 140, 166), only a few antagonists were reported to be isolated from fresh produce sources, and their antimicrobial effects against foodborne pathogens were evaluated (158, 167). This review summarizes the studies published during the past decade (2010-2022), mainly focusing on investigating the antagonistic activities of antagonistic strains against foodborne pathogens in fresh produce. The paper search was carried out on the Web of Science based on the following criteria: "(Competitive exclusion OR bacterial antagonism OR bacterial interference OR antagonistic) AND (*Salmonella* OR *Listeria monocytogenes* OR *Escherichia coli* O157:H7 OR STEC) AND (Fresh produce OR vegetables OR fruits)." The years were specified as from 2010 to 2022. In total, 80 papers were searched. After reading through all contents, 15 studies fit the topic and were reviewed. **Table 7** shows that various antagonistic cultures, including LAB and non-LAB, that were isolated from various food sources such as spinach, apples,

peaches, pineapples, sourdough, fermented vegetables, fresh-cut vegetables, sprouts, mung bean seeds, lettuce stems, curd, cow dung, *Ginkgo biloba* L., vegetable seeds, and Amazonian acai fruits. To investigate the antagonistic effects of these isolated antagonistic bacteria on foodborne pathogens, the isolated antagonists were co-inoculated with different foodborne pathogens (*E. coli*, *E. coli* O157:H7, *S. enterica*, *L. innocua*, *L. monocytogenes*, *S.* Enteritidis, *S. aureus*, *B. cereus*, *P. aeruginosa*, *Proteus vulgaris*, *S.* Typhi, *Serritia* spp., *Xanthomonas campestris*, *S.* Typhimurium, and *S.* Paratyphi) on different types of fresh produce.

Among the isolated antagonistic bacteria, some LAB have been reported, including *L. plantarum*, *L. fermentum*, *L. brevis*, *L. rhamnosus*, and *Pediococcus pentosaceus*. Most of these LAB had been widely summarized in previous reviews, except for *L. fermentum*, *Rummeliibacillus pycnus* STR 0103f, and *Rummeliibacillus stabekisii* STR 0404f. Russo et al. (2014) presented that *L. fermentum* had stronger antagonistic effects against *L. monocytogenes* than *L. plantarum*, reducing by 2 versus 1.1 Log CFU/g of *L. monocytogenes* in pineapples stored at 5 °C after 7 days, but *L. fermentum* had no inhibitory activities on *E. coli* O157:H7 (129). A study (168) conducted by Russo et al. (2015) compared the antagonistic effects of 8 Log CFU/g of *L. plantarum* B2 and *L. fermentum* PBCC11.5 on inhibiting against 4 Log CFU/g of inoculated *L. monocytogenes* in fresh-cut cantaloupes stored at 4 °C for 9 days. *L. monocytogenes* populations declined by 2 and 3 Log CFU/g, respectively, when co-cultured with *L. plantarum* B2 and *L. fermentum* PBCC11.5. Addition of LAB to fresh vegetables or fruits reduced pathogens by1-3 Log at the end of storage under different conditions (**Table 7**). Tinrat and Sedtananun (2022) (134) isolated *Rummeliibacillus pycnus* STR 0103f, and *Rummeliibacillus stabekisii* STR 0404f from fermented vegetable products, which exhibited the broad inhibitory spectrum against both Gram-negative pathogens, including *E. coli* ATCC 25922, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* ATCC 27853, *S.* Typhimurium ATCC 13311, *S.* Paratyphi A, and Gram-positive pathogens, including *Bacillus cereus* DMST 5040, *B. subtilis*, and *Staphylococcus aureus* ATCC 25923 (134).

In addition to LAB antagonists, **Table 7** illustrates various non-LAB antagonists isolated from various food sources that could inhibit different foodborne pathogens. *Bacillus cereus*, *B. pumilus*, and *B.*

subtilis with antagonistic activities against pathogens have been reviewed by Yeni et al. (2021) (125). The rest of non-LAB antagonists consist of diverse species. Especially in a study conducted by Lopez-Velasco (2012), 23 species were reported to be isolated from fresh spinach, which reduced *E. coli* O157:H7 on spinach leaves by 0.26-1.29 Log CFU/g after 24 h at 25 °C compared to the control (128). Among these non-LAB antagonistic species, the *Pseudomonas* genus had the most species, including *Pseudomonas fluorescens*, *P. koreensis*, *P. rhodesiae*, and *Pseudomonas* spp., followed by *Bacillus* with three species. Species from the *Pseudomonas* genus were also reported in other studies, including *P. graminis* CPA-7 (126, 169, 170), *Pseudomonas* sp. strain M309 (70), *P. flavescens* and *P. orientalis* (171). In comparison with the antagonistic activities of LAB antagonists against pathogens, non-LAB antagonists generally showed stronger inhibitory effects on different pathogens. For example, Alegre et al. (2013) showed that *P. graminis* CPA-7 could reduce the levels of *E. coli* O157:H7, *S. enterica*, and *L. innocua* by 4.5, 4.7, and 5.9 Log units, respectively, after 2 days at 20 °C (126). Based on these studies, the antagonistic effects of antagonists against pathogens were impacted by antagonistic species, storage temperatures, storage time, fresh produce types, initial inoculation levels, and pathogen species.

Antagonist mechanisms

Antagonistic bacteria elicit multiple mechanisms for antagonistic activity against foodborne pathogens in fresh produce. These mechanisms include: competing for attachment sites; competing for limiting nutrients; and producing antimicrobial substances (e.g., bacteriocin) (**Figure 1**). Antagonistic bacteria use individual or combinations of the modes to interact with foodborne pathogens on/in fresh produce. The wise exploitation of these mechanisms of microbial interference may be beneficial to produce safety and quality, the food industry, human and animal health, and the economy.

Competition for limiting nutrients

Antagonists can inhibit the pathogens in fresh produce by competing for nutrients, including carbohydrates, vitamins, and minerals, to support their survival or growth. Normally, the nutrients on

fresh produce surfaces are relatively limited. The competition between antagonists and pathogens depends on fresh produce variety, antagonistic species, and pathogen species. For example, Lima et al. (1999) reported that the antimicrobial activity of some antagonistic microbes against foodborne pathogens was more effective on citrus than on strawberries (172). Janisiewicz et al. (2000) developed a non-destructive method using a plate filled with tissue culture including antagonists with cylinders inserted in the wells (173). A cylinder contains a membrane at the bottom to allow tissue culture movement but excludes the antagonist. Poppe et al. (2003) applied this method to investigate the impact of tissue culture with low and high nutrients on the inhibitory activity of *Pantoea agglomerans* CPA-2 against *Penicillium* spp. (174). The results showed that CPA-2 could inhibit the conidia germination in tissue culture at low nutrient but not higher nutrient concentrations, which suggests that CPA-2 consumed the nutrients from the tissue culture, probably including nitrogen, oxygen, amino acids, or vitamins, to prevent the survival or growth of foodborne pathogens. Otherwise, the tissue culture with high nutrient concentration would support sufficient nutrients for both CPA02 and conidia. In addition, Yu and Lee (2015) applied *Pseudomonas putida* JBC17 (JBC17) to treat wounded mandarin fruits to suppress the incidence of *Penicillium digitatum* through nutrient competition and identified the involved genes (175). The results of the nutrient competition assay showed the inhibition of conidial germination by JBC17 was caused by nutrient starvation. The construction of the transposon library of JBC17 identified *ppx* and *pepP* genes can encode exopolyphosphatase and Xaa-Pro aminopeptidase, respectively, which were related to the inhibition of conidial germination.

Competition for attachment sites

For the competition between antagonistic bacteria and foodborne pathogens for attachment sites, antagonistic bacteria can displace the foodborne pathogens previously attaching on the sites of fresh produce surfaces or they can co-attach on the same sites (111, 166). Collado et al. (2007) have confirmed that *Lactobacillus rhamnosus* GG, *L. rhamnosus* LC705, *Bifidobacterium breve* 99 and *Propionibacterium freudenreichii* ssp. *shermanii* JS could displace pathogens colonizing on the mucosal

surface of the human intestine (176). The motility of antagonistic bacteria is a factor impacting competition with foodborne pathogens for the binding sites and nutrients from fresh produce. The antagonist with highly motile cells can reach the attachment sites and nutrients more efficiently (177). Thus, the property of high motility of bacteria should be taken into consideration during the antagonist selection. During the competition for binding sites, antagonistic bacteria can alter the competing environments, e.g., lactic acid bacteria can lower pH level, making a harsh environment for foodborne pathogen growth in foods (178). In addition, Valeriano et al. (2016) reported that *Lactobacillus mucosae* LM1 and *Lactobacillus johnsonii* PF01 shared carbohydrate-binding sites with several enteric pathogens, such as *E. coli* K88 and *S. Typhimurium* KCCM 40253 (179). Thus, the LAB can compete with specific pathogens for the host binding receptor sites. Kawai and Akira (2009) reported that microbe-associated molecular patterns may be expressed by LAB, which matches with the trans-membrane host pattern recognition receptors, e.g. Toll-like receptors, so as to exclude competitors (180). Generally, antagonistic bacteria can apply steric hindrance to occupy pathogen receptor sites as a way to inhibit the adherence of foodborne pathogens (181). Additionally, antagonistic bacteria can produce certain proteins to degrade carbohydrate receptors or form biofilm or biosurfactants to inhibit the attachment of pathogens to host cells (182).

Production of antibacterial substances

Antimicrobial proteins, including antibiotics, bacteriocins, and colicins, are generated by some antagonistic bacteria, which can inhibit foodborne pathogens in fresh produce (Yang et al., 2014). Antagonistic bacteria also can generate a variety of antimicrobial substances that inhibit Gram-negative and Gram-positive bacteria. The antimicrobial substances include organic acids, hydrogen peroxide, and bacteriocin. The organic acids generated by antagonistic bacteria mainly include lactic acids and acetic acids, which can establish a low pH environment to inhibit the growth of foodborne pathogens, especially Gram-negative pathogens sensitive to low pH (183). Pathogen cells can accumulate the organic acids, which leads to gradual pH reduction in the cytoplasm and cell death at certain points (183). The

bacteriocins generated by antagonistic bacteria are antimicrobial proteins or peptides, which can be classified into low-molecular-weight bacteriocin (< 1,000 Da) and high-molecular-weight bacteriocins (> 1,000 Da) (166). Many bacteriocins have narrow antimicrobial activity. Most bacteriocins only have antimicrobial effects on Gram-positive bacteria (e.g., nisin active on *L. monocytogenes*), but some LAB bacteriocins can inhibit both Gram-negative and Gram-positive bacteria. For instance, Cleusix et al. (2008) reported that *Lactobacillus reuteri* ATCC 55730 generated reuterin (3-hydroxypropionaldehyde) with a wide antimicrobial spectrum against Gram-negative and Gram-positive bacteria (184). In addition, bacteriocins produced by non-LAB bacteria have been reported and investigated, including variacin, cerein 8A, microcins, and colicins (111). Microcins are small antimicrobial peptides that can inhibit Gram-negative bacteria. Pomares et al. (2009) reported that Microcin J25 had an inhibitory effect on *E. coli* O157:H7, *Shigella* spp., and *S. enterica*. A chymotrypsin-susceptible Microcin J25 mutant was developed and could be used as a bio-preservative agent to mitigate Gram-negative pathogens in foods (185).

In summary, due to the limitations of physical and chemical antimicrobial approaches, demand has been increasing in the past decade for natural biocontrol methods to inactivate foodborne pathogens in fresh produce. Antagonistic bacteria can be isolated from numerous food sources, including fresh vegetables, fruits, or plants. The isolated antagonistic bacteria can be classified into the LAB group and the non-LAB group. The LAB antagonists have been widely investigated, and their mode of action has been addressed in a relatively clear fashion. Throughout the summary of studies on investigating the antagonistic effects of antagonistic bacteria against foodborne pathogens in fresh produce, the LAB antagonists presented a limited inhibitory effect on pathogens, while the non-LAB antagonists showed high diversity in food sources and stronger antagonistic activities against foodborne pathogens. Thus, future studies may focus on deeply investigating the mode of action of the non-LAB antagonists, evaluating their safety after application in fresh produce, and improving their pathogen-inactivating strength by combining with other physical, chemical, or biological strategies. The antimicrobial ability of a antagonist cocktail containing multiple strains against foodborne pathogens could also be explored. In

addition, the potential for antimicrobial resistance of foodborne pathogens induced by antagonistic bacteria should be considered.

Part 4. Bacterial communities in fresh vegetables during the postharvest

Introduction

Fresh produce harbors various microbial communities on the phyllosphere as epiphytes and within the leaf as endophytes (186). The microbial communities play essential roles in affecting fresh produce safety and quality as they may contain foodborne pathogens, spoilage microbes, and/or antagonistic microorganisms that can inhibit external foodborne pathogens (187). The antagonists potentially can be applied as biological control measures to reduce the risk of fresh produce contamination during the produce supply chain. Although many studies have isolated antagonists from different kinds of fresh produce by using culture-dependent approaches (127, 134, 188–190), a large number of potential antagonists may not be identified and isolated currently, as only approximately 1% of microbes are culturable (191). Thus, understanding the composition and diversity of fresh produce microbiota and the interactions between microbes and/or microbes and hosts is critical knowledge to facilitate the exploration of effective biocontrol strategies for improving fresh produce safety and quality. The composition and diversity of fresh produce microbiota can be impacted by various factors, such as soil types, irrigation water, farming practices, environmental temperatures, relative humidity, produce types, cultivars, handling, wash processing, sanitizer types, waxing, storage temperatures, storage time, and/or pathogen contamination (81, 186, 192–197).

In the past decade, with the development of DNA sequencing technologies and bioinformatic tools for producing, processing, and analyzing microbiota data, research on profiling microbial communities in various kinds of fresh produce increasingly has been reported. Several previous reviews have summarized profiles of microbial communities in fresh fruits and vegetables and dynamic changes in microbial composition and diversity during the preharvest and postharvest stages (187, 198–200). Gorni et al. (2015) reviewed studies related to microbial communities in fresh-cut fruits and vegetables published from 2012-2015 (198). This review summarized the main technologies applied to profile the microbial communities in different types of fresh-cut produce. However, only seven studies related to

microbial communities of fresh-cut fruits and vegetables were reviewed (192, 193, 201–205). All these studies applied the Roche 454 Pyrosequencing platform to sequence amplicons based on different primers of the 16S rRNA gene. Kusstatscher et al. (2020) reviewed research published from 2015-2020 investigating microbial communities in fresh fruits and vegetables during postharvest and their importance in fresh produce quality and safety (187). He summarized the key information for developing biological control against foodborne pathogens in fresh produce and storability of fresh produce. Among the 23 studies evaluated, 20 were related to investigating fruit microbiota during postharvest, and three were associated with vegetable microbiota. Droby and Wisniewski (2018) and Zhang et al. (2021) summarized the research related to fruit microbiotas and stated that the fruit microbiota contains beneficial microorganisms that could be used for biological control of postharvest diseases (199, 200). Fruit microbiota research can enhance our understanding of the postharvest microbial communities playing an important role in the health and physiology of postharvest fruit. Advances in DNA and RNA sequencing techniques, e.g., meta-omics, can facilitate deeper investigation of gene expression, proteins, and metabolites of microbial communities present in fresh produce during the preharvest and postharvest.

High-throughput Sequencing

Traditionally, classic methods include selective culture media, colony morphology, Gram staining, and biochemical tests applied for identifying specific bacterial species (206). Despite being widely applied, these methods only identify a limited number of microorganisms, as only an approximate 1% of microbes are cultured currently (191). Culture-independent methods for investigating microbial communities have advanced remarkably during the past decade. Advanced high throughput sequencing (HTS) technologies, also known as next-generation sequencing (NGS), have been developed and employed for investigating the structures and functions of microbial communities. The main improvements are metabarcoding and metagenomics (198). The HTS can be carried out on various sequencing platforms, such as Roche 454, Ion torrent, Illumina, and PacBio (207), by reading each nucleotide of amplicons or fragments of whole genomes to efficiently generate sequencing data. After

processing and analysis of the sequencing data, whole microbial communities in samples can be profiled at different taxonomic levels.

Metabarcoding, also known as amplicon sequencing, generally contains 16S/18S rRNA gene sequencing and ITS region sequencing, in which the selected regions of the DNA extracted from samples are amplified and sequenced (208, 209). The raw sequences are clustered into operational taxonomic units (OTUs) at 97% similarity or amplicon sequence variants (ASVs) with corresponding abundances. Thereafter, the OTUs are assigned to the taxonomic reference databases (Silva, Greengenes, OTT, RDP, or NCBI) for taxonomic identification (210). Metabarcoding approaches have been widely applied to profile the microbial communities in fresh fruits and vegetables (186, 192, 193, 202). Due to the relatively short read length (200-500 bp), however, the resolution for taxonomic identification of amplicon sequences is limited to the genus level (211).

Metagenomic sequencing is a method to sequence the total DNA extracted from a sample. The metagenomics procedure includes DNA extraction, library preparation, sequencing, assembly or binding, and taxonomic annotation (212). Compared to the metabarcoding method, metagenomics provides a higher resolution of microorganism identification, which can reach the species level (213). In addition, the functions of microbes, co-occurring networks between microbes or microbes and hosts, or single nucleotide polymorphisms (SNP) linked to antibiotic resistance can be identified (214, 215). Ottesen et al. (2013a) reported applying the metagenomic sequencing method to investigate the microbial communities in different anatomical niches of tomato plants including flowers, top leaves, tomatoes, stems, bottom leaves, and roots (203). In addition, Ottensen et al. (2013b) applied the shotgun sequencing approach for detecting *Salmonella* to address microbial communities in tomatoes before and after enrichment (216). However, the application of this method in studying fresh produce microbiota is still rare because of the higher cost of library preparation and sequencing as well as more complex sequencing data analyses compared to metabarcoding methods.

Since the realization that fresh produce harbors beneficial microorganisms and the development of high throughput sequencing, studies on investigating the structures of fresh produce microbiota and

their functions have indicated that microbial communities play essential roles in produce health and physiology. To deeply explore the functions of fresh produce microbiota on safety and quality, meta-omics (including metagenomics, metatranscriptomics, metaproteomics, and metabolomics) are just recently being applied in studies of postharvest fresh produce (217). This will extensively increase our understanding of interactions between microorganisms and/or microbes and hosts at the levels of genomics, transcriptomics, proteomics, and metabolomics. Metatranscriptomic sequencing is applied to investigate the gene expression of microbial communities under different conditions. This method follows a similar workflow as the metagenomic sequencing approach but targets mRNA instead of the total DNA, which includes RNA extraction, rRNA removal, and reverse transcription into cDNA (complementary DNA), cDNA sequencing, taxonomic profiling and quantification, and functional profiling and quantification (218). In addition, metaproteomics and metabolomics are employed to investigate the changes in proteins and metabolites produced by microbial communities under different conditions, respectively (219, 220). The workflow of metaproteomics includes protein extraction, protein digestion, peptide separation, mass spectrometer analysis, protein assembly, and functional analysis (221). The metabolomics workflow follows metabolites extraction, data acquisition, identification of metabolites, and pathway analysis (222).

Based on these review papers, research profiling the structures of fresh fruits and the key factors impacting the fruit microbiota published during the past decade has been well summarized. The summary for fresh vegetable microbiota profiles and factors, by contrast, is still lacking. Thus, I summarized the profiling of microbial communities in different types of fresh vegetables during postharvest, key factors influencing the composition and diversity of fresh vegetables, interactions between fresh vegetable microbiota and pathogenic bacteria and/or spoilage bacteria, and indicators identified as being associated with contamination and/or spoilage of fresh vegetables. In addition, the current sequencing techniques, sequencing data platforms, and differential abundance analysis approaches were summarized.

Bacterial communities of fresh vegetables

Over the past 10 years, research investigating the microbial communities association with fresh vegetables can be classified into three stages. The first stage is profiling of the fresh produce microbiota, which is about the observation-based question, "Who is there?" The second stage is monitoring the changes in composition and diversity of microbial communities in fresh produce, which is related to the question "What are they doing?" The third stage, exploring the interactions between microbes and microbes and/or microbes and their hosts, is associated with the question "How will they respond?" (223). This review summarizes studies related to investigating bacterial communities of fresh vegetables that were published from 2010 to 2022. The published papers were collected from the Web of Science platform following criteria such as "(Fresh produce OR vegetables) AND (Microbiome OR Microbiota OR Bacterial communities) AND (Postharvest OR Foodborne pathogens)." In total, 133 articles were searched. After the filtration of papers that did not match the topic, 12 papers were obtained. An additional 12 papers related to bacterial communities associated with fresh vegetables were obtained from references in those papers. Therefore, 24 papers were reviewed in total.

Table 8 shows all studies that profiled the bacterial communities in various fresh vegetables, including alfalfa sprouts, broccoli, cucumber, cilantro, fresh herbs, iceberg lettuce, mungo bean sprouts, mushroom, perilla leaf, Romaine lettuce, spinach, radish sprouts, rapeseed sprouts, rocket salad, sugar beet, sweet pepper, and spring mix salad. Most fresh vegetables shared similar dominant phyla, mainly containing Proteobacteria, Bacteroidetes, Firmicutes, and Actinobacteria. Proteobacteria (with relative abundances ranging from ca. 25% to 99%) were almost always the most dominant phyla across all the types of vegetables, except for the fresh herbs (ca. 13% to 84%). In fresh herbs, Firmicutes was the dominant phyla, which may be an indication that the physical structure, physiological properties, and chemical release of fresh herbs allow the Firmicutes to dominate. Although the bacteria at the phylum level were similar across fresh vegetables, the composition of bacterial communities at the genus level varied dramatically. For example, Asakura et al. (2016) illustrated that *Pseudomonas*, *Paenibacillus*, and *Pantoea* had the most relative abundance in radish sprouts (194). In a study by Kim et al. (2018), *Pseudomonas*, *Hafnia*, and *Acinetobacteria* were the top three most abundant genera (224). Mamphogoro

et al. (2020) noted that *Acinetobacter*, *Pseudomonas*, and *Burkholderia* dominated the bacterial communities of green sweet pepper (225). Gu et al. (2022) reported that fresh Romaine lettuce harbored *Pseudomonas*, unclassified Leuconostocaceae, and *Weissella* as the top three most abundant genera (197). Even for the same type of fresh vegetables, depending on the study, the dominant bacterial communities at the genus level presented large differences. The top three dominant genera in fresh spinach reported by Lopez-Velasco et al. (2011) included unclassified Enterobacteriaceae, *Pseudomonas*, and unclassified Oxalobacteraceae (192). Gu et al. reported that *Pseudomonas*, *Acinetobacter*, and *Flavobacterium* dominated the bacterial communities addressed in their study (75). Similarly, remarkable variances in the dominant genera were observed in the studies profiling Romaine lettuce microbiota (81, 197, 226). The differences in the most abundant bacterial communities of the same fresh vegetable from different studies indicate that the composition and diversity of bacterial communities are affected by various factors in addition to the type of fresh vegetable.

Key factors impacting bacterial communities in fresh vegetables

The effect of environmental biotic or abiotic factors, such as soil types, farming practices, cultivation strategies, the washing process, sanitizer types, and fresh produce types, on the composition and diversity of bacterial communities in fresh fruits and vegetables have been extensively investigated during the past decade. A previous review published by Zhang et al. (2021) summarized factors impacting the microbial communities of fresh fruits (200). These factors covered climate (e.g., rainfall), environment (e.g., soil), genetics, maturation (e.g., ripening and senescence), cultivar, tissue type, hormone, and management practices (e.g., sanitation and storage). Studies carried out by Abdelfattah et al. (2018) and Wassermann et al. (2019) showed that the composition of microbial communities of fruits was shaped by the growth environment and management practices (227, 228). Abdelfattah et al. (2016) reported remarkable differences in alpha and beta diversities between organic and conventional apples (229). In addition, different parts of organic apples presented unique microbes. The impact of farming practices (organic and conventional), on fruit microbiota was also addressed in the research from

Wassermann et al. (2019) (228). For postharvest processing, Abdelfattah et al. have reported the impact of washing, waxing, and storage conditions on the microbiota in different parts of the apple (stem tips, peel, and calyx) (230). The tissues from the three parts of apples also played an important role in determining the diversity and structure of microbial communities. The washing and waxing process for apples also had a remarkable effect on apple microbiota. In addition, the diversity and composition of apple microbial communities was strongly altered over different storage times or at different temperatures. However, a summary of factors impacting the microbial communities associated with fresh vegetables has not yet been reported. In this review, I summarized key factors influencing the composition and diversity of bacterial communities, mainly including soil textures, farming practice, irrigation, type of irrigation water, irrigation time, planting season, harvest time, farm location, cultivar, leaf maturity, health status, produce type, leaf surface sterilization, inoculation, brand, washing process, packaging, storage time, storage temperature, and distribution routes. Among these factors, multiple studies have shown that produce type had the largest impact on the structure of bacterial communities in fresh vegetables (193, 202, 231–235).

In a 2013 study by Leff and Fierer, the 16S rRNA gene pyrosequencing method was applied to investigate the impact of produce types (sprouts, spinach, lettuce, mushroom, and peppers) and farming practice types (organic vs. conventional) on the structures of bacterial communities (193). The permutational multivariate analysis of variance test results demonstrated that the composition and diversity of vegetable microbiota were most remarkably affected by produce type, followed by farming practice. Specifically, the differences in the dominant OTUs for each type of vegetable were illustrated, including *Pantoea* sp. and *Klebsiella*/*Raoulterlla* for both bean sprouts and spinach; *Acinetobacter* sp. for alfalfa sprouts; *Xanthomonas* sp., *Pantoea* sp., *Pectobacterium* sp., *Leuconostoc* sp., and *Janthnobacterium* sp. premonent for lettuce; *Pseudomonas* sp. and *Pedobacter* sp. for mushrooms; and *Pantoea* sp. for pepper. Jang et al. (2021) carried out a study to evaluate the influence of three sprout types (alfalfa, radish, and rapeseed), three brands (A, B, and C), and two distribution routes (offline and online) on bacterial ecology (235). The alpha and beta diversity analyses showed that sprout type had the

most effect on the diversity of bacterial communities. When the taxa in the three types of sprouts were compared, the top three dominant genera for alfalfa sprouts were unclassified Enterobacteriaceae, *Acinetobacter*, and *Janthinobacterium*; for radish sprouts the dominant genera were *Chryseobacterium*, *Acinetobacter*, and *Flavobacterium*; and for rapeseed sprouts they were *Chryseobacterium*, *Acinetobacter*, and *Flavobacterium*. In addition to the produce type, storage time showed a significant influence on the structures of produce microbiota. Lopez-Velasco et al. (2011) monitored the dynamic changes in structures of bacterial communities in fresh spinach stored at 4 ℃ or 10 ℃ for 15 days, showing that bacterial richness and evenness were reduced after storage (192). The unclassified Enterobacteriaceae and *Pseudomonas* had significantly increased after 15 days of storage at both temperatures. *Escherichia* were dominant in the bacterial communities in packaged spinach after being stored at 10 ℃ for 15 days but not after storage at 4 ℃, indicating that storing spinach at 4 ℃ can effectively prevent the growth of *Escherichia* spp. Liao and Wang, 2021 showed that after *E. coli* O157:H7 inoculation, the brand factor played the most important role in changing the composition and diversity of bacterial communities in spring mix salad, compared to the storage factor (81). Based on the analysis of composition of microbiomes test (236), the relative abundances of 18 bacteria was significantly influenced by the brand. In Brand A samples, *Comamonas*, *Brochothrix*, *Dyadobacter*, *Psychrobacter*, *Herminiimonas*, *Arthrobacter*, and *Shewanella* had the highest relative abundances; Brand B samples had *Saccharibacillus*, *Cmobacterium*, *Paenibacillus*, *Massilia*, and *Rheinheimera* with the highest relative abundances; while Brand C samples harbored the highest relative abundances of *Methylobacterium*, *Xanthomonas*, *Sphingomonas*, *Alkanindiges*, *Variovorax*, and *Aeromonas*.

In addition, several studies have investigated the impact of foodborne pathogen contamination on the structures and diversity of bacterial communities (81, 197, 233, 237), which reflected the interactions between foodborne pathogens and vegetable microbiota at the perspective of bacterial relative abundance. Williams et al. (2013) investigated the impact of *E. coli* O157:H7 inoculation and its surviving cells on the composition and diversity of bacterial communities in the phyllosphere of Romaine lettuce plants

(237). The Romaine lettuce plants were inoculated with 7 Log CFU/plant of an attenuated strain of *E. coli* O157:H7. Principal coordinate analysis based on Unifrac distances between bacterial communities showed that both *E. coli* O157:H7 inoculation and *E. coli* O157:H7 persistence markedly impacted the bacterial diversity of Romaine lettuce plants harvested at 14 days post-inoculation. The decrease in *E. coli* O157:H7 surviving on lettuce plants might be correlated with the presence of *Erwinia*. The effect of 6 Log CFU/g of inoculated *E. coli* O157:H7 and bacterial communities in commercial spring mix salad homogenates was evaluated during storage at 4 ℃ for 15 days (81). Since fluctuations of *E. coli* O157:H7 behaviors were observed on Day 5 and Day 10, samples were classified as *E. coli* O157:H7 growth group (G1) and *E. coli* O157:H7 reduction group (G2). Based on differential abundance analysis with the Pearson correlation analysis, it was found that *Brochothrix*, *Dyadobacter*, *Shewanella*, *Yersinia*, and *Methylotenera* were positively correlated with G1; and *Fusobacterium*, *Canobacterium*, *Aeromonas*, and *Planomicrobium* were negatively correlated with G2. These findings indicate the former may be applied as indicators of *E. coli* O157:H7 contamination in spring mix salad, and the latter could be applied as potential antagonists against the pathogen. However, further isolation and validation of the candidates are still needed. Based on the above studies, investigating the impact of various environmental factors on bacterial composition and diversity in fresh vegetables can enhance our knowledge of changes in microbiota profiles under different conditions, utilizing effective factors to control vegetable safety and quality, and identifying potential antagonistic bacteria against foodborne pathogens.

Microbiota sequence processing

With the rapid development of DNA sequencing techniques and the decrease in sequencing cost (238), the sizes of sequencing data have been continuously increasing. To process the high-throughput sequencing data of microbiota, numerous bioinformatic tools have been developed over the past decades. In order to process the increasingly massive microbiota sequencing data, the quantitative insights into microbial ecology (QIIME) method was developed (239). QIIME is an open-source software pipeline to process the raw amplicon sequencing data (bacterial and archaeal 16S rRNA gene sequence, eukaryotic

18S rRNA gene sequence, and fungal internal transcribed spacer region sequence) mainly via data demultiplexing and quality filtering, operational taxonomic unit (OTU) picking, representative sequence picking, diversity analysis, taxonomic analysis, phylogenetic analysis, and OTU networking analysis. QIIME supports various microbial community analyses, such as alpha diversity analysis, beta diversity analysis, phylogenetic analysis, taxonomic analysis, and network analysis. QIIME 1 provides various graphics to visualize the processed microbiota data (239). However, as of January 1, 2018, QIIME 1 was succeeded by QIIME 2. More information can be viewed from its website (http://qiime.org/).

The QIIME 2 pipeline is an upgraded version of QIIME 1, which has been completely re-engineered and rewritten and is a reproducible and interactive platform for microbiota data analysis (240). QIIME 2 was developed based on a large number of plugins contributed by third-party developers. The main workflow of QIIME 2 is similar to that of QIIME 1, including data importing, demultiplexing, primer removal, denoising, diversity analysis, taxonomic analysis, and differential abundance analysis. Different from QIIME 1, QIIME 2 generates amplicon sequence variants (ASV) clustered at 100% similarity of sequences instead of OTUs clustered at 97% similarity after the denoising step (241). The denoising step in the workflow is one of the key steps to filter out noisy reads, remove chimeric sequences, eliminate singletons, join denoised paired end reads, then dereplicate sequences, and generate an ASV table with corresponding abundances of each ASV. To implement this step, two methods can be employed as plugins in the QIIME 2 platform, DADA2 (242) and Deblur (243). With consistent contribution, plugins with new functionality have been continuously developed, including time-series analysis, paired-sample comparison, and machine learning analysis. In addition, the plugins for analyzing metagenomics (244, 245) and metabolomics (246, 247) have been released, including q2-shogun (244), q2-metaphlan2 (245), q2-cscs (246), and q2-metabolomics (247). QIIME 2 also has an interactive data visualization online module (https://view.qiime2.org/), which makes it easy for researchers to visualize data analysis outcomes. Therefore, the QIIME 2 platform has been developed to be applied not only for amplicon sequencing data analysis but also for meta-omics data analysis to help researchers better

understand microbiota features. Currently, QIIME 2 is the most widely applied platform for microbiota data analysis.

Mothur is a comprehensive software for analyzing microbiota sequence data (Schloss et al., 2009). It was established based on several previous bioinformatic tools, including DOTUR (248), SONS (249), TreeClimber (250), ∫-LIBSHUFF (251), and UniFrac (252), to form a powerful package for analyzing amplicon sequencing data (253). In addition, this software added a few other features: First, more than 25 calculators were used for measuring alpha and beta diversity. Second, dendrogram, Venn diagrams, and heat maps were applied to visualize data. Third, the function of data quality filtering was applied, and fourth, NAST-based sequence aligner was employed. Finally, the command could be called from a file with the list of commands or directly input in the command line. Mothur can produce OTU tables with the corresponding abundances of each OTU (254). More information can be seen at https://mothur.org/.

USEARCH is a software package combining searching and clustering algorithms to analyze sequences (255). The algorithms applied are heuristic for rapidly identifying one or a few successful hits instead of all different sequences. For searching or clustering query sequences against the reference database, the sequences in the database are ordered based on the decreasing number of letters/words in common between sequences, indicating similar sequences containing a larger number of common letters/words. Based on this order, the hit of a query sequence will be quickly found in the database. Thus, this method significantly reduces computing time compared to the common searching approach, BLAST. Finally, the USEARCH package clusters sequences into an OTU table using the UPARSE-OTU algorithm (256). More information can be viewed at https://drive5.com/usearch/.

Phyloseq is a software project to explore microbiota profiles based on the R programming language and environment (257). This software supports various types of data processing, including data importing, data filtering, data subsetting, data merging, results combination, results comparison, alpha and beta diversity analyses, ordination analysis, gap statistic, and generation of publication-grade figures. Data generated from other platforms or outputs obtained from other R packages can be implemented on

the phyloseq R package, such as biom-format data, qiime-exported data, mothur-exported data, data2-denoised data, and fast parallel UniFrac data. The plots made by Phyloseq include the ordination plot, alpha diversity plot, phylogenetic tree plot, bar plot, and heatmap plot. More information can be seen at https://joey711.github.io/phyloseq/.

Sourmarsh is different from the above pipelines because it clusters raw sequences into OTUs or ASVs and is a pipeline written in C++, Rust, and Python programming. Sourmash applies a MinHash-based sketching algorithm to produce k-mer hash signatures of DNA, RNA, and protein sequences (258, 259). Sourmash signatures can be searched, clustered, stored, and annotated with taxonomy. Due to the low memory required, evaluation of sequence similarity between large datasets using sourmash signatures or searching sequences from genomes or metagenomes are processed very quickly. The sourmash pipeline has been mainly applied in genomic or metagenomic studies, while it is rarely used in amplicon sequence studies currently (https://sourmash.readthedocs.io/en/latest/).

QIIME 1, mothur, and USEARCH software packages cluster sequences into OTU tables, while the QIIME 2 and Phyloseq pipelines generate ASV tables. The OTU is defined by 97% similarity of reads clustered as a taxonomic unit for downstream taxonomic analysis. The ASV is clustered based on exact sequence variants, 100% similarity of sequences. Sourmash software utilizes all the sequences to generate a k-mer hash table. As viewed in **Table 8**, QIIME 1 and mothur pipelines were used starting from 2013. Nine studies applied QIIME 1 and five studies used mothur to process the amplicon sequences. Most of them were reported before 2019. Nine studies after 2019 employed the QIIME 2 pipeline to process amplicon sequencing data. This suggests that the QIIME 2 pipeline has gradually become dominant in the microbiota data process. Although the QIIME 2 pipeline tends to be the main tool, the parameters in different software plugins need to be optimized for specific microbiota sequence data, e.g., position at which forward and reverse read sequences should be truncated due to decrease in quality.

Differential abundance data analysis

After taxonomic analysis during microbiota data analysis, differential abundance data analysis can be carried out to analyze the difference in microbial abundances between different groups and to identify features or indicators with significantly greater abundance associated with specific groups (260). Currently, a variety of differential abundance data analysis methods have been developed and applied in different fields of microbiota studies, including linear discriminant analysis (LDA) effect size (LEfSe) (261), analysis of composition of microbiomes (ANCOM) (236), analysis of compositions of microbiomes bias correction (ANCOM-BC) (262), and machine learning algorithms, which have different advantages and disadvantages.

The linear discriminant analysis (LDA) effect size (LEfSe) algorithm was developed by Segata et al. (2011) to identify indicators from high-dimensional class comparisons, which was specifically designed for metagenomic data (261). In this algorithm, the non-parametric Kruskal-Wallis test was first used to determine the features with significantly differential abundance among different classes. Subsequently, a set of pairwise comparisons between subclasses were conducted using the Wilcoxon rank-sum test. Finally, the LDA method was applied to estimate and rank the effect sizes of the determined differential abundant features. In default, the features with LDA scores greater than 2.0 were considered as the differential abundant indicators. LEfSe can be applied to different biological data, such as amplicon sequences, whole genome sequences, or mRNA sequences, and can determine various features, such as taxa, genes, metabolites, or proteins (261). LEfSe can be implemented in the online Galaxy platform (http://huttenhower.sph.harvard.edu/galaxy) or LEfSe R package ('microbiomeMarker'). This algorithm has been widely applied in numerous microbiota studies. However, a drawback of this method is that the Kruskal-Wallis test in LEfSe does not appropriately take into consideration the compositional effect in the microbiota data, thus the differential abundance analysis may have high false discovery rates (262).

ANCOM is a statistical framework used for comparison of the composition of two or more microbiota groups to identify features with significantly differential abundances (236). The ANCOM test accounts for the underlying structure of microbiota data without making distribution assumptions. The

identification of features was carried out by the calculation of the centered log ratio (clr) F statistic and the W statistic. The clr F statistic is measured to evaluate the effect size of features between groups, in which the microbial abundance data was clr-transformed. The W statistic is the number of times the log-ratio of a taxon with every other taxon being tested was identified to be significantly different across groups (226). In a previous simulation study conducted by Mandal et al. (2015), only the ANCOM test almost always achieved a low false discovery rate (FDR) and maintained higher statistical power compared to the other two differential abundance analysis methods, t-test and Zero Inflated Gaussian test (236). The drawbacks of ANCOM include that the p-value for each taxon tested is not provided, that confidence intervals of differential abundance for each taxon is not given, and that computation can be intensive (262). ANCOM can be implemented in the QIIME 2 pipeline (https://docs.qiime2.org/) or R package ('ancom.R') (263).

ANCOM-BC is an algorithm to identify features with significantly differential abundances, estimate the sampling fractions, and correct the bias introduced by the differences among samples (262). ANCOM-BC assumes that the tested sample is a fraction of a unit volume of an ecosystem. To account for the sampling fraction, a sample-specific offset parameter is introduced into a linear regression model in ANCOM-BC. The advantages of ANCOM-BC are: appropriate p values are provided after the statistical test; confidence intervals of the differential abundance of tested features are offered; the FDRs are well controlled; the statistical power is maintained; and simple computation is applied (262). ANCOM-BC can be implemented by using its R package ('ANCOMBC').

Application of differential abundance analysis in fresh vegetable microbiota studies

For the 24 fresh vegetable microbiota studies listed in **Table 8**, only six studies carried out differential abundance analysis on the microbial taxonomic data, including one study using Fisher's exact test (264), one applying the LEfSe test (235), three employing ANCOM tests (195, 196, 226), and one study using both the LEfSe and ANCOM tests (81). The other studies simply described the differences or

changes in diversity and relative abundance of bacterial communities in fresh vegetables among different samples, under various conditions, or during storage without any statistical tests.

Jang et al. (2021) applied the LEfSe test to compare differences in the relative abundance of sprout microbiota among three types of sprouts, including alfalfa sprouts, radish sprouts, and rapeseed sprouts (235). It was found that Enterobacteriaceae and Enterobacteriales had significantly higher relative abundance in alfalfa sprouts; *Massilia*, Methylophilales, *Methylotenera*, Rhodospirillales, Alcaligenaceae, *Kaistia*, *Agrobacterium*, Rickettsiales, *Peredibacter*, *Raphanus*, *Sphingomonas*, and Rhizobiaceae had significantly higher relative abundance in radish sprouts; and *Rhodobacter*, *Aeromonas*, Aeromonadales, *Tolumonas*, Neisseriales, *Comamonas*, *Pseudomonas*, and Burkholderiales had markedly higher relative abundance in rapeseed sprouts. These features identified with remarkably higher relative abundance were considered as indicators in the corresponding samples (261).

Kusstatscher et al. (2019) employed the ANCOM test to identify health and disease indicators in the microbiota of sugar beets (195). *Plectospharella*, *Flavobacterium*, and *Pseudarthrobacter* were identified as health indicators, and *Penicillium*, *Lactobacillus*, and *Candida* were found to be highly associated with the decaying status of sugar beets. In the study from Tenzin et al. (2022), ANCOM was used to identify significantly more abundant features (ASVs or families) in ready-to-eat spinach treated with different types of washing agents and features in spinach stored at 4 °C on days 0, 5, and 10 (196). The results showed no indicators were identified for samples treated with different washing agents, but 17 ASVs and five families were found depending on the sampling day, including seven ASVs and three families (Sphingobacteriaceae, Xanthomonadaceae, and Flavobacteriaceae) that were identified for Day 0 samples, four ASVs and one family (Pseudomonadaceae) for Day 5 samples, and six ASVs for Day 10 samples.

Differences in bacterial populations present in spring mix salad samples from three brands (A, B, and C), samples collected at different storage times (Day 0 vs. Day 15), and samples inoculated with *E. coli* O157:H7 (non-inoculated vs. inoculated) were analyzed using the ANCOM test (81). *Comamonas*, *Brochothrix*, *Dyadobacter*, *Psychrobacter*, *Herminiimonas*, *Arthrobacter*, and *Shewanella* were identified

as indicators for Brand A; *Saccharibacillus*, *Camobacterium*, *Paenibacillus*, *Massilia*, and *Rheinheimera* were identified as indicators for Brand B; and *Methylobacterium*, *Xanthomonas*, *Sphingomonas*, *Alkanindiges*, *Variovorax*, and *Aeromonas* were identified as indicators for Brand C. During 15-days of storage, *Pigmentiphage*, *Brevundimonas*, and *Frigoribacterium* had significant increases in relative abundance, while *Exiguobacterium* had a significant decrease in relative abundance. In addition, the LEfSe test was employed to investigate the contribution of bacterial features to the growth of inoculated *E. coli* O157:H7 in spring mix salad samples. During 5 days of storage at 4 °C, 12 genera were identified as indicators for samples where growth of *E. coli* O157:H7 was observed. These genera were *Wautersiella*, *Yersinia*, *Methylotenera*, *Dyadobacter*, *Variovorax*, *Ochrobactrum*, *Lactococcus*, *Novosphingoblum*, *Rhizobium*, *Brochothrix*, *Shewanella*, and *Arthrobacter*. Ten genera were identified as indicators for samples where a decrease of *E. coli* O157:H7 populations occurred during storage: *Weissella*, *Fusobacterium*, *Planomicrobium*, *Saccharibacillus*, *Trichococcus*, *Aeromonas*, *Paenibacillus*, *Massilia*, *Carnobacterium*, and *Janthinobacterium*.

The above studies suggest that the application of differential abundance analysis in analyzing fresh vegetable microbiota data can not only evaluate the impact of various factors on microbial communities in fresh vegetables but also identify the critical features or indicators for the associated samples. Indicators associated with food safety and quality can potentially be developed as biological control agents to improve the safety and quality of fresh produce. Thus, it is suggested that future fresh produce microbiota studies apply proper differential abundance analysis methods to identify critical features associated with groups of samples (e.g., contaminated vs. non-contaminated). Although indicators have been identified by these differential abundance analysis methods, the identified indicators may need to be further isolated and validated for their availability.

Machine learning

With the development of computing and mathematical algorithms, machine learning methods used in solving regression, classification, and clustering problems have become popular for big data

analyses in various fields, including engineering (265), healthcare (266), finance (267), agriculture (268), and recently emerging in food science (269). The progress in machine learning approaches shines a light on the use of microbiota datasets generated from food safety and quality research. Currently, limited studies have applied machine learning algorithms in fresh produce microbiota studies. In my Chapter four study, commonly used machine learning algorithms, including Random forests (RF), *k*-nearest neighbors (*k*-NN), and neural network (NN), were employed to establish fresh produce safety and quality classifiers using amplicon sequencing datasets (ASVs and k-mer hashes) to predict produce contamination and quality reduction (270). The critical features associated with fresh produce safety and quality were identified.

RF is an ensemble learning method for classification and regression, which constructs a number of decision trees during the training time and outputs different classes (classification) or mean predictions (regression) of individual decision trees (271). While powerful, its main drawback is its slow run time (272). Liao et al. (2022) established the RF-based fresh produce safety and quality models using ASV and k-mer hash datasets through the "randomForest" package in R programming of version 4.1.0 (270, 273). *k*-NN is a non-parametric algorithm used for classification or regression (274). The principle is to assign samples to labels based on a reference set of labeled samples (training samples) in a d-dimensional input feature space. The classification of an unlabeled sample is carried out based on the labels of the nearest k training samples that are nearest to the unlabeled sample. Compared to RF, *k*-NN is mainly limited in terms of its predictive power. Liao et al. (2022) also constructed the *k*-NN-based safety and quality models using ASV and k-mer hash datasets related to fresh produce safety and quality through using the 'caret' package in R programming (270). NN is a simplified mathematical model that is inspired by the structure and function of biological neural networks (275). In this algorithm, the weighted sum of inputs (input layer) arriving at each neuron is passed through an activation function (hidden layer) to generate an output signal (output layer) (276). NN can learn and model the complex non-linear relationship between input and output. In addition, Liao et al. (2022) built the NN-based models using ASV and k-mer hash datasets associated with fresh produce safety and quality by using the 'nnet' R package (270).

In that study, the publicly available microbiota datasets associated with fresh produce safety (3 datasets) and quality (3 datasets) in ASV and k-mer formats were used to train RF-based, *k*-NN-based, and NN-based models to predict fresh produce contamination and quality reduction. The study reported that RF-based models using 7-mer hash datasets had the best performance for predicting fresh produce safety (82% accuracy) and quality (82% accuracy). In addition, the consistent and generalizable indicators for pathogen-contaminated produce and decreasing-quality produce were identified by using the ANCOM-BC test. For produce safety-related samples, five indicators were identified for contaminated produce samples. The indicators were *Faecalibacterium, Peredibacter*, *Bacteroides*, *Listeria*, and *Escherichia-Shigella*. In addition, two indicators (*Pseudomonas* and *Rheinheimera*) were identified for non-contaminated produce samples. For produce-quality-related samples, five indicators were identified for decreasing-quality samples (*Clostridium, Acetobacter*, *Lactobacillus*, *Gluconobacter*, and *Leuconostoc*) and six indicators were identified for good-quality samples (*Nocardioides*, *Pire4_lineage*, *Paracoccus, Parabastomonas*, *Pedobacter*, and *Sphigomonas)*. RF-based models contain the function of feature/variable selection, which ranks the features based on their mean decrease in accuracy (MDA) obtained from the models, showing that 332 and 138 genera were identified to make positive contributions to IPS and IPQ classification, respectively. These features covered all the indicators identified by the ANCOM-BC test, indicating the feature selection of RF-based models is more sensitive to detect the differences in features between sample groups in integrated datasets. A similar result was reported by Sheh et al. (2022) that the ANCOM test identified two indicators while RF-based models selected nine features associated with the development of strictures (277).

Conclusion and future suggestions

Although numerous studies have been carried out for profiling microbial communities of different types of fresh produce, studies on investigating the interactions between foodborne pathogens and/or spoilage microbes and microbial communities are very limited. Through additional studies, more indicators associated with fresh produce safety and quality could be identified, and this knowledge could

potentially be applied to develop biological control strategies to improve fresh produce safety and quality. At present, antagonist isolation methods based on culture-dependent methods are random and limited. Thus, methods of isolation of the identified indicators need to be developed, which could facilitate developing more effective biocontrol strategies to control fresh produce safety and quality.

Currently, most fresh produce microbiota studies are carried out using amplicon sequencing techniques, especially 16S rRNA gene sequencing. Although the amplicon sequencing method is cost-effective, this method is limited to investigating the diversity and composition of microbial communities. Deeper exploration of the interactions between microbes in microbial communities or microbes and hosts requires meta-omics techniques, such as metagenomics for investigating antimicrobial resistance bacteria and antimicrobial resistance genes; metatranscriptomics for studying the changes in gene expression levels; and metaproteomics and metabolomics for exploring the protein and metabolic pathways. With the development of sequencing technologies, the cost of these advanced sequencing approaches has been continuously decreasing in the past decade, and fresh produce microbiota studies using meta-omics could be a trend in the future.

Finally, for the microbiota data process and analysis, since 2018 QIIME 2 has been increasingly popular for amplicon sequencing data analysis. Due to the open-source property of the QIIME 2 pipeline, more plugins of meta-omics data analysis tools have been implemented. QIIME 2 will continue to be the most widely used pipeline in the near future. On the other hand, up to more than 50% of amplicon sequences are discarded during the denoising and/or alignment steps, which contain numerous microbial features related to food safety and food quality. To save the sequence loss, k-mer hash data generated by the sourmash pipeline can be used instead of OTU or ASV. To identify indicators of produce safety or quality, machine learning methods will be increasingly applied due to the updated and optimized algorithms for the classification and clustering of samples, taxa, genes, proteins, or metabolites.

References

1. Dhingra D, Michael M, Rajput H, Patil RT. 2012. Dietary fibre in foods: a review. J Food Sci Technol 49:255–266.
2. Rickman JC, Barrett DM, Bruhn CM. 2007. Nutritional comparison of fresh, frozen and canned fruits and vegetables. Part 1. Vitamins C and B and phenolic compounds. J Sci Food Agric 87:930–944.
3. World Health Organization. 2003. Diet, nutrition, and prevention of chronic diseases. https://apps.who.int/iris/bitstream/handle/10665/42665/WHO_TRS_916.pdf;sequence=1 [Accessed November 1, 2022].
4. United States Department of Agriculture (USDA). 2021. https://www.ers.usda.gov/data-products/food-availability-per-capita-data-system/ [Accessed November 1, 2022].
5. Abdulsudi IZ, Yoshinori K. 2010. A review of microbiological safety of fruits and vegetables and the introduction of electrolyzed water as an alternative to sodium hypochlorite solution. Afr J Food Sci 4:778-789.
6. Harris LJ, Farber JN, Beuchat LR, Parish ME, Suslow TV, Garrett EH, Busta FF. 2003. Outbreaks associated with fresh produce: incidence, growth, and survival of pathogens in fresh and fresh-cut produce. Comp Rev Food Sci Food Safety 2:78–141.
7. Mandrell RE. 2011. Tracing pathogens in fruit and vegetable production chains, p. 548–595. *In* Tracing pathogens in the food chain. Elsevier.
8. CDC. 2009. About national outbreak reporting system (NORS). https://www.cdc.gov/nors/about.html [Accessed November 1, 2022].
9. CDC. 2022. National Outbreak Reporting System (NORS) Dashboard. https://wwwn.cdc.gov/norsdashboard/ [Accessed November 1, 2022].
10. Jackson BR, Griffin PM, Cole D, Walsh KA, Chai SJ. 2013. Outbreak-associated Salmonella enterica serotypes and food Commodities, United States, 1998-2008. Emerging Infect Dis 19:1239–1244.

11. Croxen MA, Law RJ, Scholz R, Keeney KM, Wlodarska M, Finlay BB. 2013. Recent advances in understanding enteric pathogenic *Escherichia coli*. Clin Microbiol Rev 26:822–880.

12. Tenaillon O, Skurnik D, Picard B, Denamur E. 2010. The population genetics of commensal *Escherichia coli*. Nat Rev Microbiol 8:207–217.

13. Jang J, Hur HG, Sadowsky MJ, Byappanahalli MN, Yan T, Ishii S. 2017. Environmental *Escherichia coli*: ecology and public health implications-a review. J Appl Microbiol 123:570–581.

14. Ishii S, Sadowsky MJ. 2008. *Escherichia coli* in the Environment: Implications for Water Quality and Human Health. Microbes Environ 23:101–108.

15. Hudson JA, Billington C, Cornelius AJ, Wilson T, On SLW, Premaratne A, King NJ. 2013. Use of a bacteriophage to inactivate *Escherichia coli* O157:H7 on beef. Food Microbiol 36:14–21.

16. Fotadar U, Zaveloff P, Terracio L. 2005. Growth of *Escherichia coli* at elevated temperatures. J Basic Microbiol 45:403–404.

17. Clavero MR, Beuchat LR. 1996. Survival of *Escherichia coli* O157:H7 in broth and processed salami as influenced by pH, water activity, and temperature and suitability of media for its recovery. Appl Environ Microbiol 62:2735–2740.

18. Alteri CJ, Mobley HLT. 2012. *Escherichia coli* physiology and metabolism dictates adaptation to diverse host microenvironments. Curr Opin Microbiol 15:3–9.

19. Al-Zyoud W, Nasereddin A, Aljarajrah H, Saket M. 2019. Culturable gut bacteria lack *Escherichia coli* in children with phenylketonuria. New Microbes New Infect 32:100616.

20. Meng J, LeJeune JT, Zhao T. 2012. Enterohemorrhagic *Escherichia coli*. Food microbiol: Fundamentals and frontiers 12: 287-309.

21. Byrne L, Jenkins C, Launders N, Elson R, Adak GK. 2015. The epidemiology, microbiology and clinical impact of Shiga toxin-producing *Escherichia coli* in England, 2009-2012. Epidemiol Infect 143:3475–3487.

22. Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM. 2011. Foodborne illness acquired in the United States-major pathogens. Emerging Infect Dis 17:7–15.

23. Hoffmann S, Batz MB, Morris JG. 2012. Annual cost of illness and quality-adjusted life Year Losses in the United States Due to 14 Foodborne Pathogens. J Food Prot 75:1292–1302.

24. Croxen MA, Finlay BB. 2010. Molecular mechanisms of *Escherichia coli* pathogenicity. Nat Rev Microbiol 8:26–38.

25. Pilla G, Tang CM. 2018. Going around in circles: virulence plasmids in enteric pathogens. Nat Rev Microbiol 16:484–495.

26. Nawrocki EM, Mosso HM, Dudley EG. 2020. A toxic environment: a growing understanding of how microbial communities affect *Escherichia coli* O157:H7 Shiga toxin expression. Appl Environ Microbiol 86.

27. In J, Foulke-Abel J, Zachos NC, Hansen A-M, Kaper JB, Bernstein HD, Halushka M, Blutt S, Estes MK, Donowitz M, Kovbasnjuk O. 2016. Enterohemorrhagic *Escherichia coli* reduce mucus and intermicrovillar bridges in human stem cell-derived colonoids. Cell Mol Gastroenterol Hepatol 2:48–62.e3.

28. Denamur E, Clermont O, Bonacorsi S, Gordon D. 2021. The population genetics of pathogenic *Escherichia coli*. Nat Rev Microbiol 19:37–54.

29. Naseer U, Løbersli I, Hindrum M, Bruvik T, Brandal LT. 2017. Virulence factors of Shiga toxin-producing *Escherichia coli* and the risk of developing haemolytic uraemic syndrome in Norway, 1992-2013. Eur J Clin Microbiol Infect Dis 36:1613–1620.

30. Gardette M, Le Hello S, Mariani-Kurkdjian P, Fabre L, Gravey F, Garrivier A, Loukiadis E, Jubelin G. 2019. Identification and prevalence of in vivo-induced genes in enterohaemorrhagic *Escherichia coli*. Virulence 10:180–193.

31. Projahn M, Lamparter MC, Ganas P, Goehler A, Lorenz-Wright SC, Maede D, Fruth A, Lang C, Schuh E. 2021. Genetic diversity and pathogenic potential of Shiga toxin-producing *Escherichia coli* (STEC) derived from German flour. Int J Food Microbiol 347:109197.

32. Pizarro-Cerdá J, Cossart P. 2019. Microbe Profile: *Listeria monocytogenes*: a paradigm among intracellular bacterial pathogens. Microbiology (Reading, Engl) 165:719–721.

33. Santos T, Théron L, Chambon C, Viala D, Centeno D, Esbelin J, Hébraud M. 2018. MALDI mass spectrometry imaging and in situ microproteomics of *Listeria monocytogenes* biofilms. J Proteomics 187:152–160.

34. Taylor AJ, Stasiewicz MJ. 2019. Persistent and sporadic *Listeria monocytogenes* strains do not differ when growing at 37 °C, in planktonic state, under different food associated stresses or energy sources. BMC Microbiol 19:257.

35. Bonsaglia ECR, Silva NCC, Fernades Júnior A, Araújo Júnior JP, Tsunemi MH, Rall VLM. 2014. Production of biofilm by *Listeria monocytogenes* in different materials and temperatures. Food Control 35:386–391.

36. da Silva EP, De Martinis ECP. 2013. Current knowledge and perspectives on biofilm formation: the case of *Listeria monocytogenes*. Appl Microbiol Biotechnol 97:957–968.

37. Orsi RH, den Bakker HC, Wiedmann M. 2011. *Listeria monocytogenes* lineages: Genomics, evolution, ecology, and phenotypic characteristics. Int J Med Microbiol 301:79–96.

38. Hernandez-Milian A, Payeras-Cifre A. 2014. What is new in listeriosis? Biomed Res Int 2014:358051.

39. Cossart P. 2011. Illuminating the landscape of host-pathogen interactions with the bacterium *Listeria monocytogenes*. Proc Natl Acad Sci USA 108:19484–19491.

40. Radoshevich L, Cossart P. 2018. *Listeria monocytogenes*: towards a complete picture of its physiology and pathogenesis. Nat Rev Microbiol 16:32–46.

41. Pizarro-Cerdá J, Kühbacher A, Cossart P. 2012. Entry of *Listeria monocytogenes* in mammalian epithelial cells: an updated view. Cold Spring Harb Perspect Med 2.

42. Birmingham CL, Canadien V, Kaniuk NA, Steinberg BE, Higgins DE, Brumell JH. 2008. Listeriolysin O allows *Listeria monocytogenes* replication in macrophage vacuoles. Nature 451:350–354.

43. Camejo A, Carvalho F, Reis O, Leitão E, Sousa S, Cabanes D. 2011. The arsenal of virulence factors deployed by *Listeria monocytogenes* to promote its cell infection cycle. Virulence 2:379–394.

44. Ward TJ, Gorski L, Borucki MK, Mandrell RE, Hutchins J, Pupedis K. 2004. Intraspecific phylogeny and lineage group identification based on the prfA virulence gene cluster of *Listeria monocytogenes*. J Bacteriol 186:4994–5002.

45. Travier L, Guadagnini S, Gouin E, Dufour A, Chenal-Francisque V, Cossart P, Olivo-Marin J-C, Ghigo J-M, Disson O, Lecuit M. 2013. ActA promotes *Listeria monocytogenes* aggregation, intestinal colonization and carriage. PLoS Pathog 9:e1003131.

46. Moumita S, Das B, Patnaik A, Balasubramanian P, Jayabalan R. 2019. Genomics and proteomics features of *Listeria monocytogenes*, p. 124–139. *In* Paramithiotis, S, Patra, JK (ed), Food Molecular Microbiology. CRC Press, Boca Raton, FL.

47. Singh V. 2013. *Salmonella* serovars and their host specificity. J Vet Sci Anim Husb 1:10-15744.

48. Chlebicz A, Śliżewska K. 2018. Campylobacteriosis, salmonellosis, yersiniosis, and listeriosis as zoonotic foodborne diseases: A review. Int J Environ Res Public Health 15.

49. Andino A, Hanning I. 2015. *Salmonella enterica*: survival, colonization, and virulence differences among serovars. Sci World J2015:520179.

50. Liu Z, Liao C, Golson K, Phillips S, Wang L. 2020. Survival of common foodborne pathogens on dried apricots made with and without sulfur dioxide treatment. Food Control 107569.

51. Santillana Farakos S, Hicks JW, Frye JG, Frank JF. 2014. Relative survival of four serotypes of *Salmonella enterica* in low-water activity whey protein powder held at 36 and 70°C at various water activity levels. J Food Prot 77:1198–1200.

52. Eng S-K, Pusparajah P, Ab Mutalib N-S, Ser H-L, Chan K-G, Lee L-H. 2015. *Salmonella* : A review on pathogenesis, epidemiology and antibiotic resistance. Front Life Sci 8:284–293.

53. Abulreesh HH. 2012. *Salmonellae* in the environment. InTech 19-50.

54. Gilchrist JJ, MacLennan CA, Hill AVS. 2015. Genetic susceptibility to invasive *Salmonella* disease. Nat Rev Immunol 15:452–463.

55. De Souza Santos M, Orth K. 2019. The role of the type III secretion system in the intracellular lifestyle of enteric pathogens. Microbiol Spectr 7.

56. Galán JE. 2021. *Salmonella* Typhimurium and inflammation: a pathogen-centric affair. Nat Rev Microbiol 19:716–725.

57. Ilyas B, Tsai CN, Coombes BK. 2017. Evolution of *Salmonella*-host cell interactions through a dynamic bacterial genome. Front Cell Infect Microbiol 7:428.

58. Goto R, Miki T, Nakamura N, Fujimoto M, Okada N. 2017. *Salmonella* Typhimurium PagP- and UgtL-dependent resistance to antimicrobial peptides contributes to the gut colonization. PLoS One 12:e0190095.

59. Dos Santos AMP, Ferrari RG, Conte-Junior CA. 2019. Virulence factors in *Salmonella* typhimurium: the sagacity of a bacterium. Curr Microbiol 76:762–773.

60. 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.

61. Luna-Guevara JJ, Arenas-Hernandez MMP, Martínez de la Peña C, Silva JL, Luna-Guevara ML. 2019. The role of pathogenic *Escherichia coli* in fresh vegetables: behavior, contamination factors, and preventive measures. Int J Microbiol 2019:2894328.

62. Alegbeleye OO, Singleton I, Sant'Ana AS. 2018. Sources and contamination routes of microbial pathogens to fresh produce during field cultivation: A review. Food Microbiol 73:177–208.

63. Mogren L, Windstam S, Boqvist S, Vågsholm I, Söderqvist K, Rosberg AK, Lindén J, Mulaosmanovic E, Karlsson M, Uhlig E, Håkansson Å, Alsanius B. 2018. The hurdle approach-A

holistic concept for controlling food safety risks associated with pathogenic bacterial contamination of leafy green vegetables. A Review. Front Microbiol 9:1965.

64. Wade W. 2003. Proteolytic yeasts isolated from raw, ripe tomatoes and metabiotic association of *Geotrichum candidum* with *Salmonella*. Int J Food Microbiol 86:101–111.

65. Francis GA, Gallone A, Nychas GJ, Sofos JN, Colelli G, Amodio ML, Spano G. 2012. Factors affecting quality and safety of fresh-cut produce. Crit Rev Food Sci Nutr 52:595–610.

66. Gil MI, Selma MV, Suslow T, Jacxsens L, Uyttendaele M, Allende A. 2015. Pre- and postharvest preventive measures and intervention strategies to control microbial food safety hazards of fresh leafy vegetables. Crit Rev Food Sci Nutr 55:453–468.

67. Guan J, Lacombe A, Tang J, Bridges DF, Sablani S, Rane B, Wu VCH. 2021. Use of mathematic models to describe the microbial inactivation on baby carrots by gaseous chlorine dioxide. Food Control 123:107832.

68. Kyere EO, Foong G, Palmer J, Wargent JJ, Fletcher GC, Flint S. 2019. Rapid attachment of *Listeria monocytogenes* to hydroponic and soil grown lettuce leaves. Food Control 101:77–80.

69. Marik CM, Zuchel J, Schaffner DW, Strawn LK. 2020. Growth and survival of *Listeria monocytogenes* on intact fruit and vegetable surfaces during postharvest handling: a systematic literature review. J Food Prot 83:108–128.

70. Oliveira M, Abadias M, Colás-Medà P, Usall J, Viñas I. 2015. Biopreservative methods to control the growth of foodborne pathogens on fresh-cut lettuce. Int J Food Microbiol 214:4–11.

71. Rajwar A, Srivastava P, Sahgal M. 2016. Microbiology of fresh produce: route of contamination, detection methods, and remedy. Crit Rev Food Sci Nutr 56:2383–2390.

72. Sheng L, Zhu M-J. 2021. Practical in-storage interventions to control foodborne pathogens on fresh produce. Comp Rev Food Sci Food Safety 20:4584–4611.

73. Smith JL, Fratamico PM. 2018. Emerging and re-emerging foodborne pathogens. Foodborne Pathog Dis 15:737–757.

74. Farber JN, Harris LJ, Parish ME, Beuchat LR, Suslow TV, Gorney JR, Garrett EH, Busta FF. 2003. Microbiological safety of controlled and modified atmosphere packaging of fresh and fresh-cut produce. Comp Rev Food Sci Food Safety 2:142–160.

75. Gu G, Ottesen A, Bolten S, Ramachandran P, Reed E, Rideout S, Luo Y, Patel J, Brown E, Nou X. 2018. Shifts in spinach microbial communities after chlorine washing and storage at compliant and abusive temperatures. Food Microbiol 73:73–84.

76. Caleb OJ, Mahajan PV, Al-Said FA-J, Opara UL. 2013. Modified atmosphere packaging technology of fresh and fresh-cut produce and the microbial consequences—A Review. Food Bioprocess Technol 6:303–329.

77. Oliveira M, Usall J, Solsona C, Alegre I, Viñas I, Abadias M. 2010. Effects of packaging type and storage temperature on the growth of foodborne pathogens on shredded “Romaine” lettuce. Food Microbiol 27:375–380.

78. Possas A, Pérez-Rodríguez F. 2022. New insights into cross-contamination of fresh-produce. Current Opinion in Food Science 100954.

79. Gu G, Ottesen A, Bolten S, Luo Y, Rideout S, Nou X. 2020. Microbiome convergence following sanitizer treatment and identification of sanitizer resistant species from spinach and lettuce rinse water. Int J Food Microbiol 318:108458.

80. Azizoglu RO, Osborne J, Wilson S, Kathariou S. 2009. Role of growth temperature in freeze-thaw tolerance of *Listeria* spp. Appl Environ Microbiol 75:5315–5320.

81. Liao C, Wang L. 2021. Evaluation of the bacterial populations present in Spring Mix salad and their impact on the behavior of *Escherichia coli* O157:H7. Food Control 107865.

82. Erickson MC, Liao J-Y, Payton AS, Cook PW, Den Bakker HC, Bautista J, Pérez JCD. 2019. Pre-harvest internalization and surface survival of *Salmonella* and *Escherichia coli* O157:H7 sprayed onto different lettuce cultivars under field and growth chamber conditions. Int J Food Microbiol 291:197–204.

83. Erickson MC, Liao J-Y, Payton AS, Cook PW, Ortega YR. 2019. Survival and internalization of *Salmonella* and *Escherichia coli* O157:H7 sprayed onto different cabbage cultivars during cultivation in growth chambers. J Sci Food Agric 99:3530–3537.

84. Gutiérrez-Rodríguez E, Gundersen A, Sbodio A, Koike S, Suslow TV. 2019. Evaluation of post-contamination survival and persistence of applied attenuated *E. coli* O157:H7 and naturally-contaminating *E. coli* O157:H7 on spinach under field conditions and following postharvest handling. Food Microbiol 77:173–184.

85. Jang H, Matthews KR. 2018. Survival and interaction of *Escherichia coli* O104:H4 on Arabidopsis thaliana and lettuce (*Lactuca sativa*) in comparison to *E. coli* O157:H7: Influence of plant defense response and bacterial capsular polysaccharide. Food Res Int 108:35–41.

86. Roy D, Melotto M. 2019. Stomatal response and human pathogen persistence in leafy greens under preharvest and postharvest environmental conditions. Postharvest Biol Technol 148:76–82.

87. Jang H, Matthews KR. 2018. Influence of surface polysaccharides of *Escherichia coli* O157:H7 on plant defense response and survival of the human enteric pathogen on Arabidopsis thaliana and lettuce (*Lactuca sativa*). Food Microbiol 70:254–261.

88. Chhetri VS, Janes ME, King JM, Doerrler W, Adhikari A. 2019. Effect of residual chlorine and organic acids on survival and attachment of *Escherichia coli* O157: H7 and *Listeria monocytogenes* on spinach leaves during storage. LWT 105:298–305.

89. Leonard SR, Simko I, Mammel MK, Richter TKS, Brandl MT. 2021. Seasonality, shelf life and storage atmosphere are main drivers of the microbiome and *E. coli* O157:H7 colonization of post-harvest lettuce cultivated in a major production area in California. Environ Microbiome 16:25.

90. Mendes-Oliveira G, Luo Y, Zhou B, Gu G, Teng Z, Bolten S, Park E, Pearlstein D, Turner ER, Millner PD, Nou X. 2022. Use of a silver-based sanitizer to accelerate *Escherichia coli* die-off on fresh-cut lettuce and maintain produce quality during cold storage: Laboratory and pilot-plant scale tests. Food Res Int 157:111170.

91. Patel J, Keelara S, Green J. 2018. Inactivation of *Escherichia coli* O157:H7 and *Salmonella* on fresh herbs by plant essential oils. Foodborne Pathog Dis 15:332–338.

92. Song YS, Stewart D, Reineke K, Wang L, Ma C, Lu Y, Shazer A, Deng K, Tortorello ML. 2019. Effects of package atmosphere and storage conditions on minimizing risk of *Escherichia coli* O157:H7 in packaged fresh baby spinach. J Food Prot 82:844–853.

93. Bardsley CA, Boyer RR, Rideout SL, Strawn LK. 2019. Survival of *Listeria monocytogenes* on the surface of basil, cilantro, dill, and parsley plants. Food Control 95:90–94.

94. Honjoh K, Lin Y, Jo K, Iwaizako Y, Maeda M, Kijima N, Miyamoto T. 2018. Possible contamination routes of *Listeria monocytogenes* in leaf lettuce during cultivation. Food Sci Technol Res 24:911–920.

95. Kyere EO, Popovich DG, Palmer J, Wargent JJ, Fletcher GC, Flint S. 2021. Reduction of the attachment, survival and growth of *L. monocytogenes* on lettuce leaves by UV-C stress. LWT 145:111528.

96. Rossi F, Lathrop A. 2019. Effects of *Lactobacillus plantarum*, *Pediococcus acidilactici*, and *Pediococcus pentosaceus* on the growth of *Listeria monocytogenes* and *Salmonella* on Alfalfa sprouts. J Food Prot 82:522–527.

97. Brierley GL, Parreira VR, Farber JM, Pagotto F. 2020. Growth of *Listeria monocytogenes* inoculated on packaged fresh-cut turnips stored at 4 and 10°C. J Food Prot 83:1296–1301.

98. Culliney P, Schmalenberger A. 2020. Growth potential of *Listeria monocytogenes* on refrigerated spinach and rocket leaves in modified atmosphere packaging. Foods 9.

99. Luciano WA, Griffin S, Targino de Souza Pedrosa G, Alvarenga V, Valdramidis V, Magnani M. 2022. Growth behavior of low populations of *Listeria monocytogenes* on fresh-cut mango, melon and papaya under different storage temperatures. Food Microbiol 102:103930.

100. Huang J, Luo Y, Zhou B, Zheng J, Nou X. 2019. Growth and survival of *Salmonella enterica* and *Listeria monocytogenes* on fresh-cut produce and their juice extracts: Impacts and interactions of food matrices and temperature abuse conditions. Food Control 100:300–304.

101. Jung Y, Guo M, Matthews KR. 2022. The effect of crisping, misting, and storage temperature on the survival or growth of *Listeria monocytogenes* and natural psychrotrophic bacteria on romaine lettuce. Food Sci Technol Int 10820132221101265.

102. Olaimat AN, Abu Ghoush M, Al-Holy M, Abu Hilal H, Al-Nabulsi AA, Osaili TM, Ayyash M, Holley RA. 2021. Survival and growth of *Listeria monocytogenes* and *Staphylococcus aureus* in ready-to-eat Mediterranean vegetable salads: Impact of storage temperature and food matrix. Int J Food Microbiol 346:109149.

103. Parichanon P, Sattayakhom A, Matan N, Matan N. 2022. Antimicrobial activity of lime oil in the vapour phase against *Listeria monocytogenes* on ready-to-eat salad during cold storage and its possible mode of action. Food Control 132:108486.

104. Shahbazi Y. 2018. Application of carboxymethyl cellulose and chitosan coatings containing Mentha spicata essential oil in fresh strawberries. Int J Biol Macromol 112:264–272.

105. Ziegler M, Rüegg S, Stephan R, Guldimann C. 2018. Growth potential of *Listeria monocytogenes* in six different RTE fruit products: impact of food matrix, storage temperature and shelf life. Ital J Food Safety 7:7581.

106. Oblessuc PR, Melotto M. 2020. A Simple assay to assess *Salmonella enterica* persistence in lettuce leaves after low inoculation dose. Front Microbiol 11:1516.

107. Xylia P, Chrysargyris A, Botsaris G, Skandamis P, Tzortzakis N. 2022. *Salmonella* Enteritidis survival in different temperatures and nutrient solution pH levels in hydroponically grown lettuce. Food Microbiol 102:103898.

108. Tokarskyy O, De J, Fatica MK, Brecht J, Schneider KR. 2018. Survival of *Escherichia coli* O157:H7 and *Salmonella* on bruised and unbruised tomatoes from three ripeness stages at two temperatures. J Food Prot 81:2028–2033.

109. Chen Z, Meng J. 2021. Persistence of *Salmonella enterica* and *Enterococcus faecium* NRRL B-2354 on Baby Spinach Subjected to Temperature Abuse after Exposure to Sub-Lethal Stresses. Foods 10.

110. Singh A, Rahman MA, Sharma R, Yemmireddy V. 2021. Papaya ripeness and post-harvest storage conditions affect growth, survival and death kinetics of *Salmonella* and spoilage organisms. Postharvest Biol Technol 181:111659.

111. Gálvez A, Abriouel H, Benomar N, Lucas R. 2010. Microbial antagonists to food-borne pathogens and biocontrol. Curr Opin Biotechnol 21:142–148.

112. Mosqueda-Melgar J, Raybaudi-Massilia RM, Martín-Belloso O. 2007. Influence of treatment time and pulse frequency on *Salmonella* Enteritidis, *Escherichia coli* and *Listeria monocytogenes* populations inoculated in melon and watermelon juices treated by pulsed electric fields. Int J Food Microbiol 117:192–200.

113. Pathanibul P, Taylor TM, Davidson PM, Harte F. 2009. Inactivation of *Escherichia coli* and *Listeria innocua* in apple and carrot juices using high pressure homogenization and nisin. Int J Food Microbiol 129:316–320.

114. Uesugi AR, Moraru CI. 2009. Reduction of *Listeria* on ready-to-eat sausages after exposure to a combination of pulsed light and nisin. J Food Prot 72:347–353.

115. Yoon J-H, Lee S-Y. 2018. Review: Comparison of the effectiveness of decontaminating strategies for fresh fruits and vegetables and related limitations. Crit Rev Food Sci Nutr 58:3189–3208.

116. Hossian et al., 2017 microbial control - Google Scholar.

117. Shen Z, Mustapha A, Lin M, Zheng G. 2017. Biocontrol of the internalization of *Salmonella enterica* and Enterohaemorrhagic *Escherichia coli* in mung bean sprouts with an endophytic Bacillus subtilis. Int J Food Microbiol 250:37–44.

118. Hossain MI, Sadekuzzaman M, Ha S-D. 2017. Probiotics as potential alternative biocontrol agents in the agriculture and food industries: A review. Food Res Int 100:63–73.

119. Fuller R. 1989. Probiotics in man and animals. J Appl Bacteriol 66:365–378.

120. Nurmi E, Nuotio L, Schneitz C. 1992. The competitive exclusion concept: development and future. Int J Food Microbiol 15:237–240.

121. Callaway TR, Edrington TS, Anderson RC, Harvey RB, Genovese KJ, Kennedy CN, Venn DW, Nisbet DJ. 2008. Probiotics, prebiotics and competitive exclusion for prophylaxis against bacterial disease. Anim Health Res Rev 9:217–225.

122. Abd El-Hack ME, El-Saadony MT, Shafi ME, Qattan SYA, Batiha GE, Khafaga AF, Abdel-Moneim A-ME, Alagawany M. 2020. Probiotics in poultry feed: A comprehensive review. J Anim Physiol Anim Nutr (Berl) 104:1835–1850.

123. Di Francesco A, Martini C, Mari M. 2016. Biological control of postharvest diseases by microbial antagonists: how many mechanisms of action? Eur J Plant Pathol 145:711–717.

124. Khubber S, Marti-Quijal FJ, Tomasevic I, Remize F, Barba FJ. 2022. Lactic acid fermentation as a useful strategy to recover antimicrobial and antioxidant compounds from food and by-products. Current Opinion in Food Science 43:189–198.

125. Yeni F, Samut H, Soyer Y. 2021. Effect of non-LAB probiotics on foodborne enteric pathogens: A systematic review. Food Reviews International 1–32.

126. Alegre I, Viñas I, Usall J, Teixidó N, Figge MJ, Abadias M. 2013. Control of foodborne pathogens on fresh-cut fruit by a novel strain of *Pseudomonas graminis*. Food Microbiol 34:390–399.

127. Dhundale V, Hemke V, Desai D, Dhundale P. 2018. Evaluation and exploration of lactic acid bacteria for preservation and extending the shelf life of fruit. Int J of Fruit Sci 18:355–368.

128. Lopez-Velasco G, Tydings HA, Boyer RR, Falkinham JO, 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.

129. Russo P, de Chiara MLV, Vernile A, Amodio ML, Arena MP, Capozzi V, Massa S, Spano G. 2014. Fresh-cut pineapple as a new carrier of probiotic lactic acid bacteria. Biomed Res Int 2014:309183.

130. Wang H, Jiang X. 2022. Isolation and characterization of competitive exclusion microorganisms from animal wastes-based composts against *Listeria monocytogenes*. J Appl Microbiol 132:4531–4543.

131. Balouiri M, Sadiki M, Ibnsouda SK. 2016. Methods for in vitro evaluating antimicrobial activity: A review. J Pharm Anal 6:71–79.

132. Hockett KL, Baltrus DA. 2017. Use of the soft-agar overlay technique to screen for bacterially produced inhibitory compounds. J Vis Exp.

133. Vijayakumar PP, Muriana PM. 2015. A microplate growth inhibition assay for screening bacteriocins against *Listeria monocytogenes* to differentiate their Mode-of-Action. Biomolecules 5:1178–1194.

134. Tinrat S, Sedtananun S. 2022. Novel *rummeliibacillus* sp. isolated from fermented vegetable products as the potential probiotics. JMBFS 11:e4194.

135. Mithun S, Dipak V, Sheela S. 2015. Screening and characterization of novel bacteriocin producing lactobacilli obtained from raw milk samples. World Journal ….

136. Pessione E. 2012. Lactic acid bacteria contribution to gut microbiota complexity: lights and shadows. Front Cell Infect Microbiol 2:86.

137. Egan K, Field D, Rea MC, Ross RP, Hill C, Cotter PD. 2016. Bacteriocins: novel solutions to age old spore-related problems? Front Microbiol 7:461.

138. Webb L, Ma L, Lu X. 2022. Impact of lactic acid bacteria on the control of *Listeria monocytogenes* in ready-to-eat foods. Food Quality and Safety.

139. Mokoena MP, Omatola CA, Olaniran AO. 2021. Applications of lactic acid bacteria and their bacteriocins against food spoilage microorganisms and foodborne pathogens. Molecules 26.

140. Ibrahim SA, Ayivi RD, Zimmerman T, Siddiqui SA, Altemimi AB, Fidan H, Esatbeyoglu T, Bakhshayesh RV. 2021. Lactic acid bacteria as antimicrobial agents: food safety and microbial food spoilage prevention. Foods 10.

141. Bian X, Evivie SE, Muhammad Z, Luo G-W, Liang H-Z, Wang N-N, Huo G-C. 2016. In vitro assessment of the antimicrobial potentials of Lactobacillus helveticus strains isolated from traditional cheese in Sinkiang China against food-borne pathogens. Food Funct 7:789–797.

142. Bungenstock L, Abdulmawjood A, Reich F. 2020. Evaluation of antibacterial properties of lactic acid bacteria from traditionally and industrially produced fermented sausages from Germany. PLoS One 15:e0230345.

143. Heredia-Castro PY, Méndez-Romero JI, Hernández-Mendoza A, Acedo-Félix E, González-Córdova AF, Vallejo-Cordoba B. 2015. Antimicrobial activity and partial characterization of bacteriocin-like inhibitory substances produced by *Lactobacillus* spp. isolated from artisanal Mexican cheese. J Dairy Sci 98:8285–8293.

144. Langa S, Martín-Cabrejas I, Montiel R, Peirotén Á, Arqués JL, Medina M. 2018. Protective effect of reuterin-producing' ' *Lactobacillus reuteri* against *Listeria monocytogenes* and *Escherichia coli* O157:H7 in semi-hard cheese. Food Control 84:284–289.

145. El-Ziney MG, Debevere JM. 1998. The Effect of Reuterin on *Listeria monocytogenes* and *Escherichia coli* O157:H7 in Milk and Cottage Cheese. J Food Prot 61:1275–1280.

146. Fhoula I, Najjari A, Turki Y, Jaballah S, Boudabous A, Ouzari H. 2013. Diversity and Antimicrobial Properties of Lactic Acid Bacteria Isolated from Rhizosphere of Olive Trees and Desert Truffles of Tunisia. BioMed Research International 2013:1–14.

147. Benkerroum N, Ghouati Y, Ghalfi H, Elmejdoub T, Roblain D, Jacques P, Thonart P. 2002. Biocontrol of *Listeria monocytogenes* in a model cultured milk (lben) by in situ bacteriocin production from *Lactococcus lactis* ssp. *lactis*. Int J Dairy Technol 55:145–151.

148. Morandi S, Silvetti T, Vezzini V, Morozzo E, Brasca M. 2020. How we can improve the antimicrobial performances of lactic acid bacteria? A new strategy to control *Listeria monocytogenes* in Gorgonzola cheese. Food Microbiol 90:103488.

149. Campagnollo FB, Margalho LP, Kamimura BA, Feliciano MD, Freire L, Lopes LS, Alvarenga VO, Cadavez VAP, Gonzales-Barron U, Schaffner DW, Sant'Ana AS. 2018. Selection of

indigenous lactic acid bacteria presenting anti-listerial activity, and their role in reducing the maturation period and assuring the safety of traditional Brazilian cheeses. Food Microbiol 73:288–297.

150. Matamoros S, Leroi F, Cardinal M, Gigout F, Chadli FK, Cornet J, Prévost H, Pilet MF. 2009. Psychrotrophic lactic acid bacteria used to improve the safety and quality of vacuum-packaged cooked and peeled tropical shrimp and cold-Smoked Salmon. J Food Prot 72:365–374.

151. Langa S, Martín-Cabrejas I, Montiel R, Landete JM, Medina M, Arqués JL. 2014. Short communication: Combined antimicrobial activity of reuterin and diacetyl against foodborne pathogens. J Dairy Sci 97:6116–6121.

152. Rahmeh R, Akbar A, Kishk M, Al-Onaizi T, Al-Azmi A, Al-Shatti A, Shajan A, Al-Mutairi S, Akbar B. 2019. Distribution and antimicrobial activity of lactic acid bacteria from raw camel milk. New Microbes New Infect 30:100560.

153. Choi A-R, Patra JK, Kim WJ, Kang S-S. 2018. Antagonistic activities and probiotic potential of lactic acid bacteria derived from a plant-based fermented food. Front Microbiol 9:1963.

154. Filannino P, Di Cagno R, Trani A, Cantatore V, Gambacorta G, Gobbetti M. 2017. Lactic acid fermentation enriches the profile of biogenic compounds and enhances the functional features of common purslane (*Portulaca oleracea* L.). J Funct Foods 39:175–185.

155. Kwaw E, Ma Y, Tchabo W, Apaliya MT, Wu M, Sackey AS, Xiao L, Tahir HE. 2018. Effect of lactobacillus strains on phenolic profile, color attributes and antioxidant activities of lactic-acid-fermented mulberry juice. Food Chem 250:148–154.

156. Wu C, Li T, Qi J, Jiang T, Xu H, Lei H. 2020. Effects of lactic acid fermentation-based biotransformation on phenolic profiles, antioxidant capacity and flavor volatiles of apple juice. LWT 122:109064.

157. Pontonio E, Montemurro M, Pinto D, Marzani B, Trani A, Ferrara G, Mazzeo A, Gobbetti M, Rizzello CG. 2019. Lactic acid fermentation of pomegranate juice as a tool to improve antioxidant activity. Front Microbiol 10:1550.

158. Siroli L, Patrignani F, Serrazanetti DI, Tabanelli G, Montanari C, Gardini F, Lanciotti R. 2015. Lactic acid bacteria and natural antimicrobials to improve the safety and shelf-life of minimally processed sliced apples and lamb's lettuce. Food Microbiol 47:74–84.

159. Ghelardi E, Celandroni F, Salvetti S, Gueye SA, Lupetti A, Senesi S. 2015. Survival and persistence of Bacillus clausii in the human gastrointestinal tract following oral administration as spore-based probiotic formulation. J Appl Microbiol 119:552–559.

160. Baccigalupi L, Ricca E, Ghelardi E. 2015. Non-LAB Probiotics: Spore Formers, p. 93–104. *In* Probiotics and prebiotics: current research and future trends. Caister Academic Press.

161. Cutting SM. 2011. Bacillus probiotics. Food Microbiol 28:214–220.

162. Cassir N, Benamar S, La Scola B. 2016. Clostridium butyricum: from beneficial to a new emerging pathogen. Clin Microbiol Infect 22:37–45.

163. Isa K, Oka K, Beauchamp N, Sato M, Wada K, Ohtani K, Nakanishi S, McCartney E, Tanaka M, Shimizu T, Kamiya S, Kruger C, Takahashi M. 2016. Safety assessment of the *Clostridium butyricum* MIYAIRI 588® probiotic strain including evaluation of antimicrobial sensitivity and presence of *Clostridium* toxin genes in vitro and teratogenicity in vivo. Hum Exp Toxicol 35:818–832.

164. Fang K, Jin X, Hong SH. 2018. Probiotic *Escherichia coli* inhibits biofilm formation of pathogenic E. coli via extracellular activity of DegP. Sci Rep 8:4939.

165. Yuksekdag ZN, Onal Darilmaz D, Beyatli Y. 2014. Dairy propionibacterium strains with potential as biopreservatives against foodborne pathogens and their tolerance–resistance properties. Eur Food Res Technol 238:17–26.

166. Wan MLY, Forsythe SJ, El-Nezami H. 2019. Probiotics interaction with foodborne pathogens: a potential alternative to antibiotics and future challenges. Crit Rev Food Sci Nutr 59:3320–3333.

167. Mansilla RT, Vives LB. 2008. Lactic acid bacteria from fresh fruit and vegetables as biocontrol agents of phytopathogenic bacteria and fungi. Int Microbiol 11:231-236.

168. Russo P, Peña N, de Chiara MLV, Amodio ML, Colelli G, Spano G. 2015. Probiotic lactic acid bacteria for the production of multifunctional fresh-cut cantaloupe. Food Res Int 77:762–772.

169. Collazo C, Abadías M, Colás-Medà P, Iglesias MB, Granado-Serrano AB, Serrano J, Viñas I. 2017. Effect of *Pseudomonas graminis* strain CPA-7 on the ability of *Listeria monocytogenes* and *Salmonella enterica* subsp. *enterica* to colonize Caco-2 cells after pre-incubation on fresh-cut pear. Int J Food Microbiol 262:55–62.

170. Iglesias MB, López ML, Echeverría G, Viñas I, Zudaire L, Abadias M. 2018. Evaluation of biocontrol capacity of *Pseudomonas graminis* CPA-7 against foodborne pathogens on fresh-cut pear and its effect on fruit volatile compounds. Food Microbiol 76:226–236.

171. Karlsson ME, Uhlig E, Håkansson Å, Alsanius BW. 2022. Seed inoculation with antagonistic bacteria limits occurrence of *E. coli* O157:H7gfp + on baby spinach leaves. BMC Microbiol 22:131.

172. Lima G, Arru S, De Curtis F, Arras G. 1999. Influence of antagonist, host fruit and pathogen on the biological control of postharvest fungal diseases by yeasts. Journal of Industrial Microbiology and Biotechnology 23:223–229.

173. Janisiewicz WJ, Tworkoski TJ, Sharer C. 2000. Characterizing the mechanism of biological control of postharvest diseases on fruits with a simple method to study competition for nutrients. Phytopathology 90:1196–1200.

174. Poppe L, Vanhoutte S, Höfte M. 2003. Modes of Action of *Pantoea agglomerans* CPA-2, an Antagonist of Postharvest Pathogens on Fruits. Eur J Plant Pathol 109:963–973.

175. Yu S-M, Lee YH. 2015. Genes involved in nutrient competition by *Pseudomonas putida* JBC17 to suppress green mold in postharvest satsuma mandarin. J Basic Microbiol 55:898–906.

176. Collado MC, Meriluoto J, Salminen S. 2007. Development of new probiotics by strain combinations: is it possible to improve the adhesion to intestinal mucus? J Dairy Sci 90:2710–2716.

177. Hibbing ME, Fuqua C, Parsek MR, Peterson SB. 2010. Bacterial competition: surviving and thriving in the microbial jungle. Nat Rev Microbiol 8:15–25.

178. Schiffrin EJ, Blum S. 2002. Interactions between the microbiota and the intestinal mucosa. Eur J Clin Nutr 56 Suppl 3:S60–4.

179. Valeriano VD, Bagon BB, Balolong MP, Kang D-K. 2016. Carbohydrate-binding specificities of potential probiotic *Lactobacillus* strains in porcine jejunal (IPEC-J2) cells and porcine mucin. J Microbiol 54:510–519.

180. Kawai T, Akira S. 2009. The roles of TLRs, RLRs and NLRs in pathogen recognition. Int Immunol 21:317–337.

181. Jessie Lau LY, Chye FY. 2018. Antagonistic effects of *Lactobacillus plantarum* 0612 on the adhesion of selected foodborne enteropathogens in various colonic environments. Food Control 91:237–247.

182. Oelschlaeger TA. 2010. Mechanisms of probiotic actions - A review. Int J Med Microbiol 300:57–62.

183. Ray B, Sandine WE. 2019. Acetic, propionic, and lactic acids of starter culture bacteria as biopreservatives. Food biopreservatives of microbial origin.

184. Cleusix V, Lacroix C, Vollenweider S, Le Blay G. 2008. Glycerol induces reuterin production and decreases *Escherichia coli* population in an in vitro model of colonic fermentation with immobilized human feces. FEMS Microbiol Ecol 63:56–64.

185. Pomares MF, Salomón RA, Pavlova O, Severinov K, Farías R, Vincent PA. 2009. Potential applicability of chymotrypsin-susceptible microcin J25 derivatives to food preservation. Appl Environ Microbiol 75:5734–5738.

186. Jackson C, Stone B, Tyler H. 2015. Emerging perspectives on the natural microbiome of fresh produce vegetables. Agriculture 5:170–187.

187. Kusstatscher P, Cernava T, Abdelfattah A, Gokul J, Korsten L, Berg G. 2020. Microbiome approaches provide the key to biologically control postharvest pathogens and storability of fruits and vegetables. FEMS Microbiol Ecol 96.

188. Iglesias MB, Echeverría G, Viñas I, López ML, Abadias M. 2018. Biopreservation of fresh-cut pear using' ' *Lactobacillus rhamnosus* GG and effect on quality and volatile compounds. LWT - Food Science and Technology 87:581–588.

189. Kim W-I, Choi SY, Han I, Cho SK, Lee Y, Kim S, Kang B, Choi O, Kim J. 2020. Inhibition of *Salmonella enterica* growth by competitive exclusion during early alfalfa sprout development using a seed-dwelling *Erwinia persicina* strain EUS78. Int J Food Microbiol 312:108374.

190. Abe Sato ST, Marques JM, da Luz de Freitas A, Sanches Progênio RC, Nunes MRT, Mota de Vasconcelos Massafra J, Gomes Moura F, Rogez H. 2020. Isolation and genetic identification of endophytic lactic acid bacteria from the amazonian açai fruits: probiotics features of selected strains and their potential to inhibit pathogens. Front Microbiol 11:610524.

191. Martiny AC. 2019. High proportions of bacteria are culturable across major biomes. ISME J 13:2125–2128.

192. 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.

193. Leff JW, Fierer N. 2013. Bacterial communities associated with the surfaces of fresh fruits and vegetables. PLoS One 8:e59310.

194. Asakura H, Tachibana M, Taguchi M, Hiroi T, Kurazono H, Makino S-I, Kasuga F, Igimi S. 2016. Seasonal and growth-dependent dynamics of bacterial community in radish sprouts. J Food Saf 36:392–401.

195. Kusstatscher P, Zachow C, Harms K, Maier J, Eigner H, Berg G, Cernava T. 2019. Microbiome-driven identification of microbial indicators for postharvest diseases of sugar beets. Microbiome 7:112.

196. Tenzin S, Ogunniyi AD, Ferro S, Deo P, Trott DJ. 2020. Effects of an eco-friendly sanitizing wash on spinach leaf bacterial community structure and diversity. Appl Sci 10:2986.

197. Gu G, Kroft B, Lichtenwald M, Luo Y, Millner P, Patel J, Nou X. 2022. Dynamics of *Listeria monocytogenes* and the microbiome on fresh-cut cantaloupe and romaine lettuce during storage at refrigerated and abusive temperatures. Int J Food Microbiol 364:109531.

198. Gorni C, Allemand D, Rossi D, Mariani P. 2015. Microbiome profiling in fresh-cut products. Trends Food Sci Technol 46:295–301.

199. Droby S, Wisniewski M. 2018. The fruit microbiome: A new frontier for postharvest biocontrol and postharvest biology. Postharvest Biol Technol 140:107–112.

200. Zhang H, Serwah Boateng NA, Ngolong Ngea GL, Shi Y, Lin H, Yang Q, Wang K, Zhang X, Zhao L, Droby S. 2021. Unravelling the fruit microbiome: The key for developing effective biological control strategies for postharvest diseases. Comp Rev Food Sci Food Safety 20:4906–4930.

201. Glenn DM, Bassett C, Dowd SE. 2015. Effect of pest management system on "Empire" apple leaf phyllosphere populations. Sci Hortic (Amsterdam) 183:58–65.

202. 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.

203. 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.

204. Pinto C, Pinho D, Sousa S, Pinheiro M, Egas C, Gomes AC. 2014. Unravelling the diversity of grapevine microbiome. PLoS One 9:e85622.

205. 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.

206. Shoaib M, Muzammil I, Hammad M, Bhutta ZA, Yaseen I. 2020. A Mini-Review on Commonly used Biochemical Tests for Identification of Bacteria. IJRP 54.

207. Kircher M, Kelso J. 2010. High-throughput DNA sequencing--concepts and limitations. Bioessays 32:524–536.

208. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res 41:e1.

209. Consortium FB. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc Natl Acad Sci USA 109:6241-6246.

210. Balvočiūtė M, Huson DH. 2017. SILVA, RDP, Greengenes, NCBI and OTT - how do these taxonomies compare? BMC Genomics 18:114.

211. Ceuppens S, De Coninck D, Bottledoorn N, Van Nieuwerburgh F, Uyttendaele M. 2017. Microbial community profiling of fresh basil and pitfalls in taxonomic assignment of enterobacterial pathogenic species based upon 16S rRNA amplicon sequencing. Int J Food Microbiol 257:148–156.

212. Sharpton TJ. 2014. An introduction to the analysis of shotgun metagenomic data. Front Plant Sci 5:209.

213. Hildebrand F. 2021. Ultra-resolution Metagenomics: When Enough Is Not Enough. mSystems e0088121.

214. Fani F, Brotherton M-C, Leprohon P, Ouellette M. 2013. Genomic analysis and reconstruction of cefotaxime resistance in *Streptococcus pneumoniae*. J Antimicrob Chemother 68:1718–1727.

215. Treven P. 2015. Strategies to develop strain-specific PCR based assays for probiotics. Benef Microbes 6:887–898.

216. Ottesen AR, Gonzalez A, Bell R, Arce C, Rideout S, Allard M, Evans P, Strain E, Musser S, Knight R, Brown E, Pettengill JB. 2013. Co-enriching microflora associated with culture based methods to detect *Salmonella* from tomato phyllosphere. PLoS One 8:e73079.

217. Saminathan T, García M, Ghimire B, Lopez C, Bodunrin A, Nimmakayala P, Abburi VL, Levi A, Balagurusamy N, Reddy UK. 2018. Metagenomic and metatranscriptomic analyses of diverse watermelon cultivars reveal the role of fruit associated microbiome in carbohydrate metabolism and ripening of mature fruits. Front Plant Sci 9:4.

218. Mehta S, Crane M, Leith E, Batut B, Hiltemann S, Arntzen MØ, Kunath BJ, Delogu F, Sajulga R, Kumar P, Johnson JE, Griffin TJ, Jagtap PD. 2021. ASaiM-MT: a validated and optimized ASaiM workflow for metatranscriptomics analysis within Galaxy framework. F1000Res 10:103.

219. Schneider T, Riedel K. 2010. Environmental proteomics: analysis of structure and function of microbial communities. Proteomics 10:785–798.

220. Cevallos-Cevallos JM, Reyes-De-Corcuera JI, Etxeberria E, Danyluk MD, Rodrick GE. 2009. Metabolomic analysis in food science: a review. Trends Food Sci Technol 20:557–566.

221. Schmidt A, Forne I, Imhof A. 2014. Bioinformatic analysis of proteomics data. BMC Syst Biol 8 Suppl 2:S3.

222. Alonso A, Marsal S, Julià A. 2015. Analytical methods in untargeted metabolomics: state of the art in 2015. Front Bioeng Biotechnol 3:23.

223. Boon E, Meehan CJ, Whidden C, Wong DH-J, Langille MGI, Beiko RG. 2014. Interactions in the microbiome: communities of organisms and communities of genes. FEMS Microbiol Rev 38:90–118.

224. Kim MS, Bae JW, Park EJ. 2018. Postharvest processing decreases the richness of bacterial taxa in the phyllosphere of broccoli. J Appl Microbiol 125:295–305.

225. Mamphogoro TP, Maboko MM, Babalola OO, Aiyegoro OA. 2020. Bacterial communities associated with the surface of fresh sweet pepper (*Capsicum annuum*) and their potential as biocontrol. Sci Rep 10:8560.

226. Liao C, Wang L. 2022. The microbial quality of commercial chopped romaine lettuce before and after the “use by” date. Front Microbiol 13:850720.

227. Abdelfattah A, Malacrinò A, Wisniewski M, Cacciola SO, Schena L. 2017. Metabarcoding: A powerful tool to investigate microbial communities and shape future plant protection strategies. Biological Control.

228. Wassermann B, Kusstatscher P, Berg G. 2019. Microbiome response to hot water treatment and potential synergy with biological control on stored apples. Front Microbiol 10:2502.

229. Abdelfattah A, Wisniewski M, Droby S, Schena L. 2016. Spatial and compositional variation in the fungal communities of organic and conventionally grown apple fruit at the consumer point-of-purchase. Hortic Res 3:16047.

230. Abdelfattah A, Whitehead SR, Macarisin D, Liu J, Burchard E, Freilich S, Dardick C, Droby S, Wisniewski M. 2020. Effect of washing, waxing and low-temperature storage on the postharvest microbiome of apple. Microorganisms 8.

231. 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.

232. Jarvis KG, Daquigan N, White JR, Morin PM, Howard LM, Manetas JE, Ottesen A, Ramachandran P, Grim CJ. 2018. Microbiomes associated with foods from plant and animal sources. Front Microbiol 9:2540.

233. Zhang Y, Jewett C, Gilley J, Bartelt-Hunt SL, Snow DD, Hodges L, Li X. 2018. Microbial communities in the rhizosphere and the root of lettuce as affected by *Salmonella*-contaminated irrigation water. FEMS Microbiol Ecol 94.

234. Keshri J, Krouptiski Y, Abu-Fani L, Achmon Y, Bauer TS, Zarka O, Maler I, Pinto R, Sela Saldinger S. 2019. Dynamics of bacterial communities in alfalfa and mung bean sprouts during refrigerated conditions. Food Microbiol 84:103261.

235. Jang MJ, Kim SY, Ricke SC, Rhee MS, Kim SA. 2021. Microbial ecology of alfalfa, radish, and rapeseed sprouts based on culture methods and 16S rRNA microbiome sequencing. Food Res Int 144:110316.

236. Mandal S, Van Treuren W, White RA, Eggesbø M, Knight R, Peddada SD. 2015. Analysis of composition of microbiomes: a novel method for studying microbial composition. Microb Ecol Health Dis 26:27663.

237. 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.

238. Jarvis KG, White JR, Grim CJ, Ewing L, Ottesen AR, Beaubrun JJ-G, Pettengill JB, Brown E, Hanes DE. 2015. Cilantro microbiome before and after nonselective pre-enrichment for Salmonella using 16S rRNA and metagenomic sequencing. BMC Microbiol 15:160.

239. Caporaso JG, 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. Nat Methods 7:335–336.

240. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu Y-X, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF,

Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37:852–857.

241. 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.

242. 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.

243. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, Knight R. 2017. Deblur rapidly resolves single-nucleotide community sequence patterns. mSystems 2.

244. Hillmann B, Al-Ghalith GA, Shields-Cutler RR, Zhu Q, Gohl DM, Beckman KB, Knight R, Knights D. 2018. Evaluating the information content of shallow shotgun metagenomics. mSystems 3.

245. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, Tett A, Huttenhower C, Segata N. 2015. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods 12:902–903.

246. Sedio BE, Rojas Echeverri JC, Boya P CA, Wright SJ. 2017. Sources of variation in foliar secondary chemistry in a tropical forest tree community. Ecology 98:616–623.

247. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu W-T, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu C-C, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE,

Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw C-C, Yang Y-L, Humpf H-U, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, P CAB, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJN, Granatosky EA, Glukhov E, Ryffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard P-M, Phapale P, Nothias L-F, Alexandrov T, Litaudon M, Wolfender J-L, Kyle JE, Metz TO, Peryea T, Nguyen D-T, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson BO, Pogliano K, Linington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N. 2016. Sharing and community curation of mass spectrometry data with global natural products social molecular networking. Nat Biotechnol 34:828–837.

248. Schloss PD, Handelsman J. 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl Environ Microbiol 71:1501–1506.

249. Schloss PD, Handelsman J. 2006. Introducing SONS, a tool for operational taxonomic unit-based comparisons of microbial community memberships and structures. Appl Environ Microbiol 72:6773–6779.

250. Schloss PD, Handelsman J. 2006. Introducing TreeClimber, a test to compare microbial community structures. Appl Environ Microbiol 72:2379–2384.

251. Schloss PD, Larget BR, Handelsman J. 2004. Integration of microbial ecology and statistics: a test to compare gene libraries. Appl Environ Microbiol 70:5485–5492.

252. Lozupone C, Knight R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 71:8228–8235.

253. Schloss PD. 2020. Reintroducing mothur: 10 Years Later. Appl Environ Microbiol 86.

254. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541.

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

256. Edgar RC. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 10:996–998.

257. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8:e61217.

258. Titus Brown C, Irber L. 2016. sourmash: a library for MinHash sketching of DNA. JOSS 1.

259. Pierce NT, Irber L, Reiter T, Brooks P, Brown CT. 2019. Large-scale sequence comparisons with sourmash. [version 1; peer review: 2 approved]. F1000Res 8:1006.

260. Yang S-C, Lin C-H, Sung CT, Fang J-Y. 2014. Antibacterial activities of bacteriocins: application in foods and pharmaceuticals. Front Microbiol 5:241.

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

262. Lin H, Peddada SD. 2020. Analysis of compositions of microbiomes with bias correction. Nat Commun 11:3514.

263. Kaul A, Mandal S, Davidov O, Peddada SD. 2017. Analysis of microbiome data in the presence of excess zeros. Front Microbiol 8:2114.

264. Margot H, Stephan R, Tasara T. 2016. Mungo bean sprout microbiome and changes associated with culture-based enrichment protocols used in detection of Gram-negative foodborne pathogens. Microbiome 4:48.

265. Ward L, O'Keeffe SC, Stevick J, Jelbert GR, Aykol M, Wolverton C. 2018. A machine learning approach for engineering bulk metallic glass alloys. Acta Mater 159:102–111.

266. Chen M, Hao Y, Hwang K, Wang L, Wang L. 2017. Disease prediction by machine learning over big data from healthcare communities. IEEE Access 5:8869–8879.

267. Ding X, Zhang Y, Liu T, Duan J. 2015. Deep learning for event-driven stock prediction. In Twenty-fourth international joint conference on artificial intelligence.

268. Behmann J, Mahlein A-K, Rumpf T, Römer C, Plümer L. 2015. A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. Precision Agric 16:239–260.

269. Tonda A, Boukhelifa N, Chabin T, Barnabé M, Génot B, Lutton E, Perrot N. 2018. Interactive machine learning for applications in food science, p. 459–477. *In* Zhou, J, Chen, F (ed), Human and machine learning. Springer International Publishing, Cham.

270. Liao C, Wang L, Quon G. 2022. Alignment-free microbiome-based classification of fresh produce safety and quality. BioRxiv.

271. Breiman L. 2001. Random forests. Machine Learning. 45:5-32.

272. Wiesmeier M, Barthold F, Blank B, Kögel-Knabner I. 2011. Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. Plant Soil 340:7–24.

273. Liaw A, Wiener M. 2002. Classification and regression by randomForest. R news. 2:18-22.

274. Altman NS. 1992. An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat 46:175–185.

275. Tino P, Benuskova L, Sperduti A. 2015. Artificial neural network models, p. 455–471. *In* Kacprzyk, J, Pedrycz, W (ed), Springer handbook of computational intelligence. Springer Berlin Heidelberg, Berlin, Heidelberg.

276. 2001. Kalman filtering and neural networks. John Wiley & Sons, Inc., New York, USA.

277. Sheh A, Artim SC, Burns MA, Molina-Mora JA, Lee MA, Dzink-Fox J, Muthupalani S, Fox JG. 2022. Alterations in common marmoset gut microbiome associated with duodenal strictures. Sci Rep 12:5277.

278. Gullian-Klanian M, Sánchez-Solis MJ. 2018. Growth kinetics of *Escherichia coli* O157:H7 on the epicarp of fresh vegetables and fruits. Braz J Microbiol 49:104–111.

279. Yuan J, Wang L. 2018. Survival of *Escherichia coli* O157:H7, *Salmonella* spp., and *Listeria monocytogenes* on Fresh and Sliced Green and Golden Kiwifruits. Foodborne Pathog Dis 15:560–567.

280. Gómez-Aldapa CA, Portillo-Torres LA, Villagómez-Ibarra JR, Rangel-Vargas E, Téllez-Jurado A, Cruz-Gálvez AM, Castro-Rosas J. 2018. Survival of foodborne bacteria on strawberries and antibacterial activities of *Hibiscus sabdariffa* extracts and chemical sanitizers on strawberries. J Food Saf 38.

281. Lee S, Han A, Yoon J-H, Lee S-Y. 2022. Growth evaluation of *Escherichia coli* O157:H7, *Salmonella* typhimurium, and *Listeria monocytogenes* in fresh fruit and vegetable juices via predictive modeling. LWT 162:113485.

282. Macarisin D, Sheth I, Hur M, Wooten A, Kwon HJ, Gao Z, De Jesus A, Jurick W, Chen Y. 2019. Survival of outbreak, food, and environmental strains of Listeria monocytogenes on whole apples as affected by cultivar and wax coating. Sci Rep 9:12170.

283. Maks N, Ye MU, Swanson S, Lee A, Freeman BB, Deng K. 2019. Evaluation of Inactivating Norovirus, Hepatitis A, and *Listeria monocytogenes* on Raspberries by Sanitizer Spray. J Food Prot 82:869–877.

284. De Jesus AJ, Sheth I, Kwon HJ, Gao Z, Palmer J, Hur M, Hammack TS, Macarisin D, Chen Y. 2020. Survival of a serotype 4b strain and a serotype 1/2a strain of *Listeria monocytogenes*, isolated from a stone fruit outbreak investigation, on whole stone fruit at 4 °C. Int J Food Microbiol 334:108801.

285. Dong L, Li Y. 2021. Fate of *Salmonella* Typhimurium and *Listeria monocytogenes* on Whole Papaya during Storage and Antimicrobial Efficiency of Aqueous Chlorine Dioxide Generated with HCl, Malic Acid or Lactic Acid on Whole Papaya. Foods 10.

286. Kuttappan D, Muyyarikkandy MS, Mathew E, Amalaradjou MA. 2021. *Listeria monocytogenes* Survival on peaches and nectarines under conditions simulating commercial stone-fruit packinghouse operations. Int J Environ Res Public Health 18.

287. Cabrera-Díaz E, Martínez-Chávez L, Gutiérrez-González P, Pérez-Montaño JA, Rodríguez-García MO, Martínez-Gonzáles NE. 2022. Effect of storage temperature and time on the behavior of *Salmonella*, *Listeria monocytogenes*, and background microbiota on whole fresh avocados (Persea americana var Hass). Int J Food Microbiol 369:109614.

288. Yin H-B, Chen C-H, Colorado-Suarez S, Patel J. 2022. Biocontrol of *Listeria monocytogenes* and *Salmonella enterica* on Fresh Strawberries with Lactic Acid Bacteria During Refrigerated Storage. Foodborne Pathog Dis 19:324–331.

289. Rodrigues Marques Ferreira ÍH, de Souza Pedrosa GT, Jung J, Ferreira de Melo AN, Campagnollo FB, Schaffner DW, Magnani M. 2020. Modeling *Salmonella enterica* fate in fresh-cut pepper (*Capsicum annuum* L.) during storage as a function of temperature and relative humidity. LWT 133:109849.

290. Chiu Y-C, Shen C, Farnham MW, Ku K-M. 2020. Three-dimensional epicuticular wax on plant surface reduces attachment and survival rate of *Salmonella* during storage. Postharvest Biol Technol 166:111197.

291. Tarlak F, Johannessen G, Bascón Villegas I, Bolívar A, Posada-Izquierdo GD, Pérez-Rodríguez F. 2020. Modelling of the Behaviour of *Salmonella enterica* serovar Reading on Commercial Fresh-Cut Iceberg Lettuce Stored at Different Temperatures. Foods 9.

292. Ha J, Park E, Kim J, Lee S, Kim S, Lee J, Choi Y, Yoon Y, Oh H, Kim Y, Lee Y, Seo Y, Kang J. 2020. Prevalence of *Salmonella* in cucumbers, antibiotic and acid resistances and description of the kinetic behavior with dynamic model during storage. J Food Saf 40.

293. Jung J, Schaffner DW. 2021. Modeling the survival of *Salmonella* on whole cucumbers as a function of temperature and relative humidity. Food Microbiol 100:103840.

294. Jung J, Schaffner DW. 2021. Quantification of Survival and Transfer of *Salmonella* on Fresh Cucumbers during Waxing. J Food Prot 84:456–462.

295. Grivokostopoulos NC, Makariti IP, Tsadaris S, Skandamis PN. 2022. Impact of population density and stress adaptation on the internalization of *Salmonella* in leafy greens. Food Microbiol 106:104053.

296. Wang W, Zhou Y, Xiao X, Yang G, Wang Q, Wei W, Liu Y, Yang H. 2018. Behavior of *Salmonella* typhimurium on fresh strawberries under different storage temperatures and wash treatments. Front Microbiol 9:2091.

297. Mathew EN, Muyyarikkandy MS, Kuttappan D, Amalaradjou MA. 2018. Attachment of *Salmonella enterica* on Mangoes and Survival Under Conditions Simulating Commercial Mango Packing House and Importer Facility. Front Microbiol 9:1519.

298. Danyluk MD, Friedrich LM, Dunn LL, Zhang J, Ritenour MA. 2019. Reduction of *Escherichia coli*, as a surrogate for *Salmonella* spp., on the surface of grapefruit during various packingline processes. Food Microbiol 78:188–193.

299. González-López C. 2019. Attachment and colonization of *Salmonella* on “Rayada”, “Golden Delicious”, and “Red Delicious” apples. J Sci Food Agri 99:1166-1171. ….

300. Akter T, Shahriar A, Alo M, Rahman T, Emran TB. 2020. Survival assessment of pathogenic bacteria with antibiotic resistance traits from fresh summer royal grape: in vitro microbial challenge test. JMBFS 10:344–349.

301. Ortiz-Solà J, Valero A, Viñas I, Colás-Medà P, Abadias M. 2020. Microbial interaction between *Salmonella enterica* and main postharvest fungal pathogens on strawberry fruit. Int J Food Microbiol 320:108489.

302. Ortiz-Solà J, Abadias I, Colàs-Medà P, Anguera M, Viñas I. 2021. Inactivation of *Salmonella enterica*, *Listeria monocytogenes* and murine norovirus (MNV-1) on fresh strawberries by conventional and water-assisted ultraviolet light (UV-C). Postharvest Biol Technol 174:111447.

303. Jin TZ, Chen W, Gurtler JB, Fan X. 2020. Effectiveness of edible coatings to inhibit browning and inactivate foodborne pathogens on fresh-cut apples. J Food Saf 40.

304. He Y, Chen R, Qi Y, Salazar JK, Zhang S, Tortorello ML, Deng X, Zhang W. 2021. Survival and transcriptomic response of *Salmonella enterica* on fresh-cut fruits. Int J Food Microbiol 348:109201.

305. Juneja VK, Osoria M, Altuntas EG, Salazar JK, Kumar GD, Sehgal S, Baker D. 2021. Inactivation of *Listeria monocytogenes*, *Escherichia coli* O157:H7, and *Salmonella* spp. on dates by antimicrobial washes. J Food Process Preserv 45.

306. Luo W, Chen M, Chen A, Dong W, Hou X, Pu B. 2015. Isolation of lactic acid bacteria from pao cai, a Chinese traditional fermented vegetable, with inhibitory activity against *Salmonella* associated with fresh-cut apple, using a modelling study. J Appl Microbiol 118:998–1006.

307. Jeon D-Y, Yum S-J, Seo DW, Kim SM, Jeong HG. 2019. Leaf-associated microbiota on perilla (*Perilla frutescens* var. frutescens) cultivated in South Korea to detect the potential risk of food poisoning. Food Res Int 126:108664.

308. Mira Miralles M, Maestre-Carballa L, Lluesma-Gomez M, Martinez-Garcia M. 2019. High-throughput 16S rRNA sequencing to assess potentially active bacteria and foodborne pathogens: A case example in ready-to-eat food. Foods 8.

309. Gu G, Yin H-B, Ottesen A, Bolten S, Patel J, Rideout S, Nou X. 2019. Microbiomes in ground water and alternative irrigation water, and spinach microbiomes impacted by irrigation with different types of water. Phytobiomes J 3:137–147.

Table 1. Multistate outbreaks of pathogenic *Escherichia coli* infections linked to fresh produce in the U.S. from 2010 to 2021*

Year	Outbreaks	Serotype	Illnesses	Hospitalizations (hospitalization rate %)	Deaths (death rate %)	Food Vehicle
2010	1	O145	31	14 (45.2)	0	Romaine lettuce
2011	2	O157:H7	86	40 (46.5)	0	Romaine lettuce
2012	3	O26	29	7 (24.1)	0	clover sprouts
		O157:H7	33	13 (39.4)	0	prepackaged leafy greens
		O145	16	6 (37.5)	0	lettuce
2013	3	O157:H7	47	18 (38.3)	1 (2.1)	prepackaged leafy greens, Romaine lettuce
		O26	26	5 (19.2)	0	lettuce
2014	5	O157:H7	31	16 (51.6)	0	prepackaged salad, salad, and spinach
		O121	19	5 (26.3)	0	clover sprouts
		O111	16	2 (12.5)	0	cabbage
2015	3	O145	7	5 (71.4)	0	prepackaged leafy greens
		O157:H7	21	13 (61.9)	0	prepackaged salad and Romaine lettuce
2016	2	O157:NM (H-)	11	2 (18.2)	0	alfalfa sprouts
		O157:H7	11	4 (36.4)	0	iceberg lettuce
2017	3	O26	8	3 (37.5)	0	spinach
		O157:H7	93	27 (29.0)	0	baby leaf, leafy greens, mizuna, and spring salad
2018	5	O157:H7	358	152 (42.5)	5 (1.4)	Romaine lettuce
2019	8	O157:H7	262	125 (47.7)	0	Romaine lettuce
		O26:H11	16	3 (18.8)	0	spinach
2020	2	O103	51	3 (5.9)	0	clover sprouts
		O157:H7	40	4 (10.0)	0	leafy greens
2021	2	O157:H7	25	8 (32.0)	1 (4.0)	prepackaged salad

* NORS and CDC website currently only includes data linked to fresh produce up to 2021 (9).

Table 2. Multistate outbreaks of *Listeria monocytogenes* infections linked to fresh produce in the U.S. from 2010 to 2021*

Year	Illnesses	Hospitalizations (hospitalization rate %)	Deaths (death rate %)	Food Vehicle
2011	147	143 (97.3)	33 (22.4)	cantaloupe
2014	2	2 (100)	1 (50)	peaches
	5	3 (60)	2 (40)	mung bean sprouts
	35	34 (97.1)	7 (20)	caramel apple
	9	9 (100)	3 (33.3)	prepackaged leafy greens
2016	19	19 (100)	1 (5.3)	lettuce
	36	31 (86.1)	4 (11.1)	enoki mushroom
	10	9 (90)	1 (10)	avocado
2017	3	3 (100)	0	caramel apple
	10	7 (70)	3 (30)	prepackaged leafy greens
	5	4 (80)	1 (20)	enoki mushroom
2018	7	5 (71.4)	1 (14.3)	stone fruit
2020	36	31 (86.1)	4 (11.1)	enoki mushrooms
2021	18	16 (88.9)	3 (16.7)	prepackaged leafy greens
	10	10 (100)	1 (10)	prepackaged leafy greens

* NORS and CDC website currently only includes data linked to fresh produce up to 2021 (9).

Table 3. Multistate outbreaks of *Salmonella enterica* infections linked to fresh produce in the U.S. from 2010 to 2021*

Year	Outbreaks	Serotype	Illnesses	Hospitalizations (hospitalization rate %)	Deaths (death rate %)	Food Vehicle
2010	2	Javiana	30	8 (26.7)	0	tomatoes
		I 4,[5],12:i:-	140	31 (22.1)	0	alfalfa sprouts
2011	5	Agona	106	10 (9.4)	0	papaya
		Enteritidis	27	3 (11.1)	0	alfalfa sprouts
		Panama	20	3 (15)	0	cantaloupe
		Uganda	25	4 (16)	0	cantaloupe
		Newport	166	0	0	tomatoes
2012	8	Braenderup and Worthington	129	33 (25.6)	0	mango
		Cubana	19	0	0	sprouts
		Javiana	49	14 (28.6)	0	cucumber
		Newport and Typhimurium	261	94 (36)	3 (1.1)	cantaloupe
		Newport	72	17 (23.6)	1 (1.4)	cantaloupe and Romaine lettuce
		Typhimurium	14	1 (7.1)	0	cantaloupe
2013	4	Saintpaul	215	40 (18.6)	0	cherry tomato, cucumber, and grape tomato
		Thompson	13	6 (46.2)	1 (7.7)	papaya
		Virchow	7	1 (14.3)	0	unspecified fruit
2014	7	Baildon	20	6 (30)	0	cantaloupe
		Enteritidis	115	19 (16.5)	0	mung bean sprouts
		Javiana	36	8 (22.2)	0	cucumber
		Minnesota	4	1 (25)	0	mango
		Newport	275	48 (17.5)	1 (0.4)	cucumber
		Paratyphi B	21	5 (23.8)	0	peppers
		Saintpaul	27	10 (37)	0	grapes
2015	5	Cubana, Kentucky, and Muenchen	32	10 (31.3)	1 (3)	alfalfa seeds

		Hartford	19	4 (21.1)	0	tomatoes
		Newport	119	17 (14.3)	0	salsa
		Paratyphi B	11	2 (18.2)	0	lettuce
		Poona	907	204 (22.5)	6 (0.7)	cucumber
2016	11	Abony and Reading	36	7 (19.4)	0	alfalfa sprouts
		Anatum	32	8 (25)	0	peppers
		Braenderup	32	6 (18.8)	0	mung bean sprouts
		Enteritidis	114	17 (14.9)	0	avocado, bean sprouts, prepackaged leafy greens, and salad mix
		Javiana	29	6 (20.7)	0	onions
		Minnesota	10	3 (30)	0	cantaloupe
		Oslo	14	3 (21.4)	0	cucumber
		Saintpaul	10	3 (30)	0	cucumber
2017	12	Agona, Gaminara, Kiambu, Senftenberg, and Thompson	213	68 (31.9)	1 (0.5)	papaya
		Anatum	20	5 (25)	1 (5)	papaya
		Enteritidis	151	31 (20.5)	0	Romaine lettuce
		Braenderup	55	18 (32.7)	0	papaya
		Chailey	14	2 (14.3)	0	coconut
		Cubana and Montevideo	62	3 (4.8)	0	sprouts
		Infantis	48	15 (31.3)	0	mango
		Infantis and Newport	4	2 (50)	0	papaya
		Javiana	27	4 (14.8)	0	leafy greens
		Newport	44	11 (25)	1 (2.3)	melon and watermelon
		Urbana	7	4 (57.1)	0	papaya
2018	5	Adelaide	77	36 (46.8)	1 (1.3)	melon
		Infantis	7	2 (28.6)	0	English cucumber
		Javiana	266	48 (18)	0	onions

		Newport	41	9 (21.9)	0	alfalfa sprouts
2019	8	Braenderup	24	9 (37.5)	0	tomatoes
		Carrau	149	45 (30.2)	0	melon
		Infantis	5	0	0	vegetable tray
		Javiana	179	75 (41.9)	0	mixed cut fruit and prepackaged salad mix
		Montevideo	32	5 (15.6)	0	iceberg lettuce
		Saintpaul	97	20 (20.6)	0	strawberries
		Uganda	81	27 (33.3)	0	papaya
2020	6	Enteritidis	101	28 (27.7)	0	peaches
		Miami	65	28 (43.1)	0	tomatoes
		Newport	1212	184 (15.2)	0	cantaloupe, red onion, and watermelon
		Rubislaw	37	14 (37.8)	0	mango
		Stanley	55	6 (10.9)	0	wood ear mushrooms
2021	2	Oranienburg	1040	260 (25)	0	onions
		Typhimurium	31	3 (9.7)	0	prepackaged salad mix

* NORS and CDC outbreaks currently only includes data linked to fresh produce up to 2021 (9).

Table 4. *Escherichia coli* O157:H7 survival in fresh produce from papers published between 2018 and 2022.

Food vehicle (Stage)	Conditions	Sampling points (post inoculation)	Inoculation level	Major results	References
Romaine lettuce (Preharvest)	21 ± 2 °C during the day and 19 ± 2 °C at night for 4 weeks before inoculation (Field)	Days 0, 1, and 5	6.3 Log CFU/g	• Wild-type and mutant *E. coli* O157:H7 reduced by 2.8 and 3.6 Log CFU/g 5 days post-inoculation, respectively. • Specific surface polysaccharides (cellulose, colanic acid, and lipopolysaccharide) of *E. coli* O157:H7 differentially induce the plant defense response.	(87)
Seven cultivars of green leafy lettuce: Gabriella, Green Star, Muir, New Red Fire, Starfighter, Tropicana, and Two Star; one cultivar of Romain lettuce: Coastal Star (Preharvest)	10.8 to 22.3 °C for 2-3 weeks before inoculation (Field)	Hours 0 and Days 5 and 9	4.4 – 4.7 Log CFU/plant for Gabriella, Green Star, Muir, and New Red Fire	• *E. coli* O157:H7 was only recovered by enrichment from Costal Star and Gabriella after 9 days post-inoculation. • No significant effect of different lettuce cultivars on survival of *E. coli* O157:H7 was observed.	(82)
	20 °C for a 10-h light cycle and 15 °C for a 14-h dark cycle (Chamber)	Days 0, 1, 4, 8, and 12	2.3-4.6 Log CFU/plant for Gabriella and New Red Fire, Green Star, Starfighter, Tropicana, Muir, Two Star, and Romaine	• No effect of cultivar was manifested on the fate of *E. coli* O157:H7 when plants were sampled for up to 12 days.	

Cabbage (Preharvest)	18 °C for a 10-h light cycle at and 17 °C for a 14-h dark cycle (Chamber)	Days 0, 3, 6, and 9 for small cabbage. Days 0, 1, 3, and 10 for cabbage in early head formation	3.4 to 4.3 Log CFU/plant for a small cabbage. 5.6-5.8 Log CFU/ml for cabbage in the early head (the initial level was not reported)	• For small cabbage: 33% to 88% of plants were positive by enrichment on day 9 sampling. For cabbage in early head: 30 to 100% of plants positive by enrichment on day 10 sampling. • For small cabbage, *E. coli* O157:H7 populations decreased by 3.4 Log CFU/plant within the first three days post-contamination, and the cells can only be recovered by enrichment. • For cabbage in early head: *E. coli* O157:H7 dropped by 0.5 to 2.1 Log CFU/plant from day 1 to day 3, and the cells can only be recovered by enrichment on day 10.	(83)
Spinach (Preharvest)	7.5-15 °C in Salinas and 7.5-17.5 °C in Soledad (Field)	Days 0, 1, 7, 12-14, and 18-21	2 and 4 Log CFU/ml	• Salinas: 2 Log CFU/ml inoculum dose: 0.7 Log CFU/g for both 25 and 150 g sample size up to 14 days post-inoculation. 4 Log CFU/ml inoculum dose: 0.7 and 1.3 Log CFU/g for 25 and 150 g sample size up to 14 days post-inoculation, respectively. Soledad (150 g sample size): 2 Log CFU/ml inoculum dose: 1.5 and 0.9 Log CFU/g depending on the field trial. 4 Log CFU/ml: 1.0 and 1.4 Log CFU/g depending on the field trial. • Differences in the applied inoculum dose of 100-fold, resulted in indistinguishable	(84)

				population densities of 1 Log CFU/g by 14 days post-inoculation.	
Spinach and Romaine (Preharvest)	22 °C and 12 h daily light (Chamber)	Days 0, 1, 3, and 7	4 - 4.5 Log CFU/cm^2	• 2.0 - 2.2 Log CFU/cm^2 on day 7. • *E. coli* O157:H7 population declined from day 1 to day 7 by 2 Log CFU/cm^2.	(86)
Basil, Butter lettuce, Cilantro, Romaine, Spinach (Postharvest)	4 °C	Days 0, 1, 3, 7, 14, and 21	3.2 Log CFU/cm^2	• 2 Log CFU/cm^2 on day 21. • The survival of *E. coli* O157:H7 was only tested on butter lettuce, as all other produce spoiled before end of experiment.	
Fresh herbs (Postharvest)	4 °C	Days 0, 2, 7, and 14	5.0 - 5.4 Log CFU/g	• 0.6 - 1.5 Log CFU/g of *E. coli* O157:H7 was reduced by washing. • *E. coli* O157:H7 declined by 0.9 to 4.1 Log CFU/g during storage on fresh herbs.	(91)
Beet, cucumber, habanero chili, onion, and radish (Postharvest)	35 °C	Hours 0, 3, 5, 7 and 9	2.7 Log CFU/ml	• Tomato: 5.4 Log CFU/ml; Radish: 6 Log CFU/ml; Cucumber: 5.5 Log CFU/ml; Beet: 5.6 Log CFU/ml; Habanero chili:5.6 log CFU/ml; Onion: 5.2 Log CFU/ml after 9 hours. • The interaction between *E. coli* O157:H7 and onion was less supportive of the growth. • The time for *E. coli* O157:H7 to attain the maximum growth rate was longer for tomato and cucumber as compared to other samples.	(278)
Tomato (Postharvest)	10 °C with RH between 25 and 60%	Days 0, 1, 3, and 7	4 Log CFU/tomato	• 0 to 0.3 Log CFU/tomato depending on the bruised or	(108)

	20 °C RH between 25 and 60%			unbruised and ripeness stages of tomato on day 7. • Initial 2-h drying led to > 1 Log CFU/tomato reduction of *E. coli* O157:H7. • *E. coli* O157:H7 counts declined on the bruised and unbruised tomatoes by 2.5 Log over the 7-day storage period. • Bruising had no significant effect ($P > 0.05$) on the survival of *E. coli* O157:H7 (CFU per tomato) compared with unbruised tomatoes. • Temperature has a significant effect on the survival of *E. coli* O157:H7 on breaker and pink tomatoes, but not on red tomatoes.	
Baby spinach (Postharvest)	5 °C	Days 0, 2, 4, 6, 8, 10, 12, and 14	2 and 4 Log CFU/g	• Using the aerobic atmosphere as an example: the growth or survival of *E. coli* O157:H7 decreased with decreasing temperature. • Population of *E. coli* O157:H7 only increased at 15 °C under both inoculation levels, whereas *E. coli* O157:H7 maintained a stable population or showed a moderate increase at 12 °C. • At the lower temperatures (5 and 7 °C), the population levels of the pathogen decreased for both inoculation levels.	(92)
	7 °C	Days 0, 3, 7, 10, and 14	2 and 4 Log CFU/g		
	12 °C	Days 0, 2, 4, 6, 8, 10, 12, and 14	2 and 4 Log CFU/g		
	15 °C	Days 0, 4, 7, 11, and 14	2 and 4 Log CFU/g		

Spinach (Postharvest)	4 °C	Hours 0, 0.25, 24, and 48 for the washing residue study. Hours 0, 20, and 48 for the attachment study	5 - 6 Log CFU/cm^2 for the effect of residues on the survival during storage. 5 - 6 Log CFU/cm^2 for the attachment study	• After 48-hour storage, 2.6 Log CFU/cm^2 of *E. coli* O157 on chlorine-treated spinach leaves was reduced and 1.4 Log CFU/cm^2 of *E. coli* O157 on lactic acid-treated leaves was decreased. • For the washing residue study: no significant change was observed in the bacterial survival up to 48 h of storage. • For the attachment study: there was no significant change in the attachment strength of *E. coli* O157:H7 on the spinach surface.	(88)
Lettuce (Postharvest)	6 °C	Days 0, 7, and 14	4 Log CFU/g	• 3.5 Log CFU/g for cultivars DA and TT on day 14. • *E. coli* O157:H7 population declined on average five-fold (0.5 Log) over 2 weeks of cold storage. • Inoculation of lettuce with *E. coli* O157:H7 did not affect the lettuce microbiota under cold storage.	(89)
	24 °C	Hours 0, 14-24, and 38-66	4 Log CFU/g	• 8 Log CFU/for cultivars DA and TT at the end of storage. • *E. coli* O157:H7 population increased by > 300-fold (~ 3 Log) within 1 day, and over 6000-fold within 2 days. • Change in *E. coli* O157:H7 population size was a variable that correlated with microbiota composition during storage at 24 °C.	

Spring mix (Postharvest)	4 °C	Days 0, 5, 10, and 15	6 Log CFU/g	• 1.5 Log CFU/g of *E. coli* O157 reduced from Day 0 to Day 15	(81)
Lettuce (Postharvest)	5 °C	Days 0,3,7, and 14	6.5 Log CFU/g	• 1.5 Log of *E. coli* O157:H7 were reduced from Day 0 to Day 14.	(90)
Fresh and sliced green and golden kiwifruits (Postharvest)	Whole fruit:23 ± 2 °C	Days 0, 5, 10, 15, 20, 25, and 30	4 or 7 Log CFU/unit	• For whole kiwi fruits with high inoculation level, after drying, 2.4 Log CFU/kiwi of *E. coli* O157:H7 on green kiwi reduced from Day 0 to Day 30, whereas 2.7 Log CFU/kiwi of *E. coli* O157:H7 on golden kiwi reduced after 20-day storage. • For whole kiwi fruits with the low inoculation level: *E. coli* O157:H7 fell below 2.0 Log CFU/kiwi on green kiwi on Day 30, and on golden kiwifruit on Day 10.	(279)
	Sliced fruit: 4 °C	Every 6 h for the first 48 h and Days 3, 5, and 7		• For sliced fruits: about 3 Log reduction was observed on both golden and green kiwifruit with the high inoculation level when stored for 7 days. No significant decrease in pathogen concentration was observed at the low inoculation level.	
Strawberry (Postharvest)	3 ± 2 °C and 22 ± 2 °C	Days 0 and 15	4 Log CFU/g	• Until Day 15, both *E. coli* O157:H7 and non-O157 STEC decreased by about 1.3 Log	(280)

		CFU/g at 3 ± 2 ℃ and about 2 Log CFU/g at 22 ± 2 °C, respectively.

ND: not detected.

Table 5. *Listeria monocytogenes* survival in fresh produce from papers published between 2018 and 2022

Pathogen strain	Food vehicle (Storage)	Conditions (°C)	Sampling points (post inoculation)	Inoculation level	Major results	References
L. monocytogenes (3-strain cocktail)	Leaf lettuce (Preharvest)	20 °C (Chamber)	Weeks 0, 2, 4, 6, 8, and 10	4, 6 and 8 Log CFU/g	• Populations of *L. monocytogenes* in the soil fell below the detection limit (2 Log CFU/g) at 8–10 weeks after the start of cultivation. • The possibility of internalization of *L. monocytogenes* into leaf lettuce plants was low. • Spraying of water contaminated with bacteria at greater than 3.2 log CFU/plant led to the survival of bacteria on the leaf lettuce leaves after seven days. • Leaf damage prolonged the survival period of bacteria on the leaves.	(94)
L. monocytogenes (3-strain cocktail)	Alfalfa Sprouts (Preharvest)	20 °C (Incubator)	Days 0, 1, 3, and 5	Low: 0.7 Log CFU/g and high: 2.7 Log CFU/g	• Both the low and high inoculation levels of LM leveled off to about 6.5 to 7.0 Log CFU/g in alfalfa seeds after 5 days of sprouting.	(96)
L. monocytogenes (5 nalidixic acid-resistant strain cocktail)	Herb plants: basil, cilantro, dill, and parsley (Preharvest)	21 °C (Field)	Hours 0, 5 and Days 1, 2, 3, 7, 14, 21, and 28.	Basil: 4.6 Log CFU/g, cilantro: 4.9 Log CFU/g, dill: 4.2 Log CFU/g, parsley: 5.1 Log CFU/g	• Basil: 0.9 Log CFU/g; Cilantro: 0.6 Dill: 1.1 Log CFU/g; Parsley: < 1.1 Log CFU/g after 28 days • Overall, *L. monocytogenes* populations did not grow on the studied herb plants; however, *L. monocytogenes* was able to survive on the surface of the studied herb plants for up to 28 days, except parsley plants, which fell below the limit of detection 7 days post-inoculation.	(93)

L. monocytogenes O8A07 and O8A08	Lettuce leaves (Preharvest)	20 °C (Field)	Hours 0, 24, 48, 72, and 96	3 Log CFU/g	• O8A07: 4.6 Log CFU/g; O8A08: 4.7 Log CFU/g at 96 h. • LM could grow in the lettuce leaves at 20 °C for 96 h. • The different strains of LM had no significant impact on the growth of LM in this study.	(95)
L. monocytogenes (3-strain cocktail)	Iceberg lettuce, rocket, and spinachrocket (Postharvest)	8 °C	Days 0, 2, 5, and 7	Spinach: low: 0.9 Log CFU/g and high:1.7-2.0 Log CFU/g Rocket: low: 0.9 Log CFU/g and high:1.6-2.0 Log CFU/g Lettuce: low: 1.0 Log CFU/g and high:1.3-2.0 Log CFU/g	• Lettuce: low: 2.6 Log CFU/g and high: 3.1-4.0 Log CFU/g, rocket: low: 2.7 Log CFU/g and high: 3.1-3.7 Log CFU/g, and spinach: low: 3.3 Log CFU/g and high: 4.1-4.6 Log CFU/g on Day 7.	(98)
L. monocytogenes (5-strain cocktail)	Fresh-cut turnips (Postharvest)	4 °C 10 °C	Days 0, 5, and 10	Brand A: 3.5 Log CFU/g Brand B: 3.5 Log CFU/g	• Brand A: 3.0 Log CFU/g; Brand B: 3.2 Log CFU/g on Day 10. • *L. monocytogenes* growth was supported on turnips stored at 10 °C. • *L. monocytogenes* survived on turnips stored at 4 °C. • Brand A: 5.16 Log CFU/g; Brand B: 4.7 Log CFU/g on Day 10.	(97)
L. monocytogenes (5-strain cocktail)	Arabic salad (Postharvest)	4, 10, and 24 °C	Days 0, 1, 2, 3, 4, and 5	5.6 Log CFU/g	• The inoculated *L. monocytogenes* could survive all conditions used in this study, except for toum sauce salad at 24 °C.	(102)

					• Addition of lemon juice and salt to Arabic salad significantly reduced the numbers of pathogens. • 24 °C: 3.7 Log CFU/g; 10 °C: 7.5 Log CFU/g; 4 °C: 6 Log CFU/g on Day 5.	
	Arabic salad with salt and lemon juice (Postharvest)			5.8 Log CFU/g	• 24 °C: 3.0 Log CFU/g; 10 °C: 5.4 Log CFU/g; 4 °C: 5.6 Log CFU/g on Day 5.	
	Tahini salad (Postharvest)			6.0 Log CFU/g	• 24 °C: 4.2 Log CFU/g; 10 °C: 8.0 Log CFU/g; 4 °C: 6.6 Log CFU/g on Day 5.	
	Coleslaw salad (Postharvest)			5.4 Log CFU/g	• 24 °C: 2.3 Log CFU/g; 10 °C: 3.1 Log CFU/g; 4 °C: 2.9 Log CFU/g on Day 5	
	Toum sauce Salad (Postharvest)			5.5 Log CFU/g	• 24 °C: < 2 Log CFU/g; 10 °C: 3.0 Log CFU/g; 4 °C: 2.8 Log CFU/g on Day 5	
L. monocytogenes (3 nalidixic acid-resistant strain cocktail)	Romaine lettuce Non-crisped (NC) Water crisped (WC) Non-misted (NM) Misted (M) (Postharvest)	5 °C	Days 0, 1, 4, and 7	5.5 Log CFU/g	• WC/NM: 3.7 Log CFU/g; WC/M: 2.9 Log CFU/g; NC/NM: 5.0 Log CFU/g; NC/M: 3.9 Log CFU/g on Day 7. • Regardless of the storage temperature crisping or misting can effectively reduce *L. monocytogenes* population on romaine lettuce during storage	(101)
L. monocytogenes Scott A	RTE vegetable salad (30% carrot, 15% green oak lettuce, 25% red	4 °C	Days 0, 1, 3, 5, and 7	0.5 Log CFU/g	• 2.0 Log CFU/g on Day 7.	(103)

	cabbage, and 30% tomato) (Postharvest)					
L. monocytogenes (3-strain cocktail)	Beet, cabbage, carrot, celery, grapefruit, kale, lemon, and red cabbage juice (Postharvest)	10 °C	Hours 0, 12, 24, 48, 72, 96, 120, and 144	2.24 Log CFU/mL	• Beet: 6.0 Log CFU/mL, cabbage: 1.8 Log CFU/mL, carrot: 1.0 Log CFU/mL, celery: 4.8 Log CFU/mL, crapefruit: 0.8 Log CFU/mL, kale: 6.3 Log CFU/mL, lemon: 1.0 Log CFU/mL, and red cabbage: 1.8 Log CFU/mL at 144 hours.	(281)
L. monocytogenes (5-strain cocktail)	Whole Green and Golden Kiwifruits (Postharvest)	23 °C	Days 0, 5, 10, 15, 20, 25, and 30	ca. 4 and 7 Log CFU/kiwi	• For the whole kiwifruits, both of high-level inoculation and low-level inoculation of LM fell below the limit of detection on Day 10 on golden kiwifruit and on Day 15 on green kiwifruit, respectively and could survive on Day 30 storage.	(279)
	Sliced Green and Golden Kiwifruits (Postharvest)	4 °C	Hours 0, 6, 12, 18, 24, 30, 36, 42, 48 and Days 3, 5, and 7	ca. 4 and 7 Log CFU/slice	• For the sliced kiwifruit, 3 Log CFU/slice reduction was obtained for high-level inoculation of LM on both green and golden kiwifruit slice after 7-day storage, while no significant decrease of LM concentration was observed for low-level inoculation of LM.	
L. monocytogenes (3-strain cocktail, cold adapted)	Cantaloupe Coconut Mango (Postharvest)	4 °C for 72 h then 8 °C 4 °C for 72 h then 12 °C	Days 0, 3, 4, 5, and 6	Cantaloupe: 4.0 Log CFU/g Coconut: 4.0 Log CFU/g Mango: 3.9 Log CFU/g	• Cantaloupe: 6.4 Log CFU/g, coconut: 7.3 Log CFU/g, and mango: 4.0 Log CFU/g on Day 6. • Cantaloupe: 6.1 Log CFU/g, coconut:8.3 Log CFU/g, and mango:3.9 Log CFU/g on Day 6.	(105)

L. monocytogenes (ATCC 19118)	Strawberry (Postharvest)	4 °C	Days 0, 2, 3, 6, 8, 10, and 12	5 Log CFU/g	• 6.26 Log CFU/g on Day 12.	(104)
L. monocytogenes (3-strain cocktail)	fresh-cut cantaloupe, honeydew, pineapple, radish, and watermelon (Postharvest)	4 °C	Days 0, 1, 2, 4, and 7	2.5-2.6 Log CFU/g	• Cantaloupe: 4.1 Log CFU/g, honeydew: 3.7 Log CFU/g, pineapple: 0.7 Log CFU/g, radish: 2.3 Log CFU/g, and watermelon: 4.1 Log CFU/g; on Day 7.	(100)
		8 °C			• Cantaloupe: 6.7 Log CFU/g, Honeydew: 5.4 Log CFU/g, pineapple: < 5 CFU/g, radish: 3.1, and watermelon: 6.7 Log CFU/g on Day 7.	
		12 °C			• Cantaloupe: 7.7 Log CFU/g, honeydew: 5.9 Log CFU/g, pineapple: < 5 CFU/g, radish: 3.2 Log CFU/g, and watermelon: 7.8 Log CFU/g; on Day 7.	
		35 °C for 2 h followed by 4°C for the reminder 7 days			• Cantaloupe: 4.1 Log CFU/g, honeydew: 4.0 Log CFU/g, pineapple: 0.9 Log CFU/g, radish: 2.9 Log CFU/g, and watermelon: 4.1 Log CFU/g on Day 7.	
L. monocytogenes (6-strain cocktail)	Unwaxed apples (3 cultivars) (Postharvest)	3 °C	Days 0, 1, 3, 7, 16, 31, 62, 93, and 160	Red Delicious: 3.99 ± 0.07 Log CFU/apple Granny Smith: 3.85 ± 0.09 Log CFU/apple Fuji: 5.13 ± 0.06 Log CFU/apple	• Fuji: 1.3 Log CFU/apple, Granny Smith: 0.6 Log CFU/apple, and Red Delicious: 0.4 Log CFU/apple on Day 160. • Apple cultivar- did not affect the prolonged survival of *L. monocytoge*nes during cold storage. • Apple coating with commercial wax significantly facilitated *L. monocytoge*nes survival in experimentally contaminated fruits	(282)

	Waxed apples (3 cultivars) (Postharvest)		Days 0, 1, 3, 7, 16, 31, 62, 93 and 160	Red Delicious: 2.94 ± 0.12 Log CFU/apple Granny Smith: 3.39 ± 0.09 Log CFU/apple Fuji: 3.92 ± 0.15 Log CFU/apple	• Fuji: 2.4 Log CFU/apple, Granny Smith: 2.4 Log CFU/apple, and Red Delicious: 2.6 Log CFU/apple on Day 160.	
L. monocytogenes (5-strain cocktail)	Raspberries (Postharvest)	-20 °C	Days 0, 1, 14, 35,62 and 90	5.6 Log CFU/g	• 3.4 Log CFU/g on Day 90. • *L. monocytogenes* can survive on the frozen berries at -20 °C.	(283)
L. monocytogenes ST382	White peach (Postharvest)	4 °C	Days 0, 3, 7, 10, 14, 17, 19, 21, and 26	3.5 Log CFU/fruit (high) 2.6 Log CFU/fruit (low)	• High: 2.0 Log CFU/fruit and low: 1.4 Log CFU/fruit on Day 26.	(284)
	Yellow peach (Postharvest)			3.6 Log CFU/fruit (high) 2.5 Log CFU/fruit (low)	• High: 2.0 Log CFU/fruit and low: 1.6 Log CFU/fruit on Day 26.	
	Yellow nectarine (Postharvest)			3.4 Log CFU/fruit (high)	• High: 2.3 Log CFU/fruit and low: 1.7 Log CFU/fruit on Day 26.	

L. monocytogenes CC5	White peach (Postharvest)	4 °C	Days 0, 3, 7, 10, 14, 17, 19, 21, and 26	2.4 Log CFU/fruit (low) 3.6 Log CFU/fruit (high)	• High: 1.8 Log CFU/fruit and low: 1.8 Log CFU/fruit on Day 26.	
	Yellow peach (Postharvest)			2.5 Log CFU/fruit (low) 3.5 Log CFU/fruit (high)	• High: 2.5 Log CFU/fruit and low: 1.6 Log CFU/fruit on Day 26.	
	Yellow nectarine (Postharvest)			2.6 Log CFU/fruit (low) 3.2 Log CFU/fruit (high)	• High: 2.5 Log CFU/fruit and low: 1.5 Log CFU/fruit on Day 26.	
L. monocytogenes (F2365)	Papayas (Postharvest)	21 °C	Days 0, 1, 7, 10, and 14	2.4 Log CFU/fruit (low) 4.67 Log CFU/fruit	• 5.9 Log CFU/fruit on Day 14. • *L. monocytogenes* maintains its population at 7 °C and grew at 21°C on the surface of whole papayas.	(285)
L. monocytogenes (4-strain cocktail)	Peaches and nectarines (Postharvest)	28°C	Hours 0, 2, 6, 12, and 18	High: 5.46 ± 0.05 Log CFU/fruit and low: 3.75 ± 0.12 Log CFU/fruit	• High: 5.7 Log CFU/fruit and low: 3.9 Log CFU/fruit at 18 h. • *L. monocytogenes* remained at the same levels on both stone fruits during the unloading and staging	(286)

					section and waxing and fungicide application section.	
L. monocytogenes (6 rifampin-resistant strain cocktail)	Fresh Hass avocados (Postharvest)	5 °C	Days 0, 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, and 48	7.4 ± 0.15 Log CFU/avocado	• 3.2 Log CFU/avocado on Day 48. • *L. monocytogenes* survived for a long period of time on the surface of avocados under refrigeration and at room temperature	(287)
		25 °C	Hours 0, 4, 8, 12, 24, 28, 32, 36, 48, 72, 96, 120, 144, 168, 192, 216, 240, and 264	7.5 ± 0.11 Log CFU/avocado	• 5.1 Log CFU/avocado after 264 h.	
L. monocytogenes C1-387	Cut melon, papaya, mango, and mix (cut melon, mongo, and papaya) (Postharvest)	4 °C	Days 0, 1, 3, 5, 7, and 10	1-4 cells/sample	• Mango: <0.3 Log CFU/g (detection limit), melon: 2 Log CFU/g; papaya: 2 Log CFU/g, and mix: 1.9 Log CFU/g on Day 10. • Even at low levels, *L. monocytogenes* cells could survive and grow in fresh-cut fruits and fruit mix during refrigerated storage. • The combination of storage condition and low pH can slow the growth	(99)
		8 °C			• Mango: 1.2 Log CFU/g, melon: 2 Log CFU/g, papaya: 2.1 Log CFU/g, and mix:1.9 Log CFU/g on Day 10.	
		12 °C			• Mango: 1.5 Log CFU/g, melon: 4 Log CFU/g, papaya: 3.7 Log CFU/g, and mix: 3.4 Log CFU/g on day 10.	

		16 °C			• Mango: 1.7 Log CFU/g, melon: 5 Log CFU/g, papaya:4.3 Log CFU/g, and and mix :4.5 Log CFU/g on day 10.	
L. monocytogenes (3-strain cocktail)	Strawberry (Postharvest)	4 °C	Days 0, 1, 3, and 7	7 Log CFU/g	• 5.6 ± 0.2 Log CFU/g on day 7.	(288)
		10 °C			• 4.9 ± 0.5 Log CFU/g on day 7.	

Table 6. *Salmonella enterica* survival in fresh produce from papers published from 2018 to 2022

Pathogen serotype/strain	Food vehicle	Conditions	Sampling points (post inoculation)	Inoculation level	Major results	References
S. enterica cocktail including serotype Enteritis and Newport	Cabbage (Preharvest)	18 °C for a 10-h light cycle and 17 °C for a 14-h dark cycle (Chamber)	Days 0, 3, 6, and 9	4.1-4.9 Log CFU/plant	• Population level of *S. enterica* decreased to ca. 1.3 – 3.9 Log CFU/plant after 9 days, which varied among different cabbage cultivars.	(82)
S. enterica cocktail including serotype Enteritis and Newport	Lettuce (Preharvest)	20 °C for a 10-h light cycle and 15 °C for a 14-h dark cycle (Chamber)	Days 0, 1 4, and 8	2.8-3.3 and 4.1-5.1 Log CFU/plant for small and large size lettuce, respectively grown in growth chamber	• Population level of *S. enterica* decreased to 3.2 to 3.7 Log CFU/plant after 8 days on the phyllosphere tissue of large size lettuce cultivar but was not detected on the phyllosphere tissue of small size lettuce cultivar in the growth chamber.	(83)
		10.8 to 22.3 °C (Field)	Days 0, 5, and 9	4.4-4.7 Log CFU/plant for lettuce grown in field	• Samples showed *S. enterica* positive after enrichment on lettuce cultivar Gabriella and New Red Fire after 9 days.	
S. enterica including serotype Enteritis and Orangieburg	Alfalfa sprouts (Preharvest)	20 °C (Incubator)	Days 0, 1, 3, and 5	Low inoculation level: 1.5 Log CFU/g High inoculation level: 3.0 Log CFU/g	• Population level of *S. enterica* increased to 6.2 Log CFU/g after 5 days. • Population level of *S. enterica* increased to 7.0 Log CFU/g after 5 days.	(96)
S. Typhimurium	Lettuce (Preharvest)	18 °C (Chamber)	Hour 0, 2 then Days 1, 10, and 20	3, 5, and 7 Log CFU/ml	• Population level of *S. enterica* was positively related with inoculation level at each sampling points.	(106)

					• *S. enterica* generally decreased over 20 days on lettuce and maintained higher level on Rede tide than Lollo Rossa and Salinas.	
S. Enteritis NCTC 5188	Lettuce (Preharvest)	21 °C for a photoperiod of 16 h and 19 °C for 8 h darkness (Chamber)	Day 0 and 28-29 for old plants. Days 0, 14 and 21 for small-medium, plants	Low concentration: 3 Log CFU/ml. High concentration: 6 Log CFU/ml	• Higher pH combined with higher inoculation levels resulted in enhanced persistence of *S. enterica* in the roots for lettuce plant in different growth stage. • *S. enterica* was not detected from leaves of small-medium lettuce plants and old plants at the end of cultivation.	(107)
S. enterica cocktail including serotype Agona, Montevideo,Newport, Poona, and Saintpaul	Tomato (Postharvest)	10 °C 20 °C	Days 0, 1, 3, and 7	4.0 Log CFU/tomato	• Population level of *S. enterica* decreased to 1.8, 1.8 and 1.7 Log CFU/tomato in tomato at three ripeness stages (breaker/green, pink, and red) stored at 10 °C. • Population level of *S. enterica* decreased to 1.3, 1.6, and 1.6 Log CFU/tomato in tomato at three ripeness stages (breaker/green, pink, and red) stored at 20 °C. • No obvious differences in *S. enterica* levels between unbruised and bruised tomato were observed at the end of the 7-day storage. • *S. enterica* had the highest levels in red tomato at the end of the 7-day storage.	(108)
S. enterica including serotype Braenderup,	Basil, cilantro, dill, parsley, and	4 °C	Days 0, 2, 7, and 14	5.0-5.5 Log CFU/g	• *S. enterica* reduced to less than ca. 3.0, 2.8, 3.9, 4.0, and 2.0 Log CFU/g on basil, parsley, tarragon, cilantro, and dill, respectively after 14 days.	(91)

Newport, Negev, Thompson, and Tennessee	tarragon (Postharvest)					
S. enterica cocktail including serotype Kasenyi, Napoli, and Veneziana	Lettuce (Postharvest)	8 °C	Hours 0 and 120	5.0 and 4.7 Log CFU/cm^2 loosely and strongly attached cell, respectively	• Population level of *S. enterica* gradually increased to ca. 6.0 Log CFU/cm^2 after 120 hours.	(96)
S. Typhimurium	Basil, butter lettuce, cilantro, Romaine, and spinach (Postharvest)	4 °C	Days 0, 1, 3, and 7	3.5 Log CFU/cm^2 leaf	• *S. enterica* persisted at higher population level on leafy greens in condition with relative humidity 95% than 65% on Day 1 and 3 and reached similar population levels (ca. 2-3 Log CFU/) on Day 7.	(86)
	Butter lettuce (Postharvest)	25 °C	Days 0, 1, 3, 7, 14, and 21	3.5 Log CFU/cm^2 leaf	• *S. enterica* persisted at 2.0 Log CFU/ after 21 days.	
S. enterica cocktail including serotype Enteritis and Typhimurium	Pepper (Postharvest)	7, 14, and 21 °C	Hours 0, 12, 24, 48, 72, 96, 120 and 144	2.5 and 4.5 Log CFU/g	• In general, the higher temperature combined with higher relative humidity, the more *S. enterica* survived on pepper.	(289)
S. enterica cocktail including serotype Tennessee and Typhimurium	Broccoli and collard (Postharvest)	4 °C	Days 0, 5, 9, and 14	4.5 – 6.0 Log CFU/	• Population level of *S. enterica* was below limit detection after 14 days for broccoli but remained 3.0 Log CFU/ for collard.	(290)

S. Reading	Lettuce (Postharvest)	4 °C 8 °C 15 °C	Days 0, 1, 2, 3, 4, 7, and 10	4.9 Log CFU/g 4.9 Log CFU/g 4.9 Log CFU/g	• The *S. enterica* population level was 2.1 Log CFU/g after 10 days • Population level of *S. enterica* was 2.9 Log CFU/g after 10 days • Population level of *S. enterica* was 5.7 Log CFU/g after 10 days	(291)
S. enterica cocktail, including SMFM2018-C15-1, SMFM2018-C17-1, and SMFM2018-C18-2	Fresh-cut cucumber (Postharvest)	10 °C	Hours 0 and 96	3.0 Log CFU/g	• Population level of *S. enterica* increased rapidly at 20, 25, and 30 °C but remained at 2.8 Log CFU/g at 10 °C.	(292)
S. enterica cocktail including serotype Enteritidis, Newport, and Typhimurium	Baby spinach (Postharvest)	4 °C for 7 days; 37 °C for 2h and 4 °C for 7 days	Days 0, 1, 2, 4, and 7	7.5 Log CFU/sample unit	• The unstressed *S. enterica* population decreased rapidly on the first day and remained at ca. 5.5 log CFU/sample unit under both cold and temperature abuse conditions. • The desiccation-stressed *S. enterica* most enhanced its persistence on baby spinach compared with *S. enterica* pre-exposed to other stresses in temperature abused condition.	(109)
S. enterica cocktail including serotype Newport, Montevideo, Saintpaul, and Stanley.	Cucumber (Postharvest)	7, 14, and 21 °C	Hours 0, 12, 24, 72, 120, 168, and in some cases 240	6.0 Log CFU/cucumber	• Population level of *S. enterica* declined under all of conditions and the survival of *S. enterica* was highest at 100% relative humidity followed by 50 and 15%.	(293)

S. enterica cocktail including serotype Newport, Montevideo, Saintpaul, and Stanley.	Cucumber (Postharvest)	7 and 21 °C	Days 0 and 7	6.0 Log	• Population level of *S. enterica* decreased by 3.3 and 2.9 Log CFU/cucumber at 21 and 7 °C, respectively.	(294)
S. Enteritis	Arugula, chicory, lettuce, and spinach (Postharvest)	5 and 20 °C	Hours 2 and 48	3, 4, 5, 6, and 7 Log CFU/ml	• The higher inoculation level, the higher population detected on leafy greens after 48 h. • The population of *S. enterica* mainly increased at 20 °C after 48 h and remained at a stable level at 5 °C after 48 h regardless of types of vegetable. • Growth of desiccation- and acid-stressed *S. enterica* were higher than other stressed and non-stressed counterparts in the majority of cases.	(295)
S. Typhimurium	Beet, cabbage, carrot, celery, grapefruit, kale, lemon, and red cabbage (Postharvest)	10 °C	Hours 0, 12, 24, 48, 72, 96, 120, and 144	2.0 Log CFU/ml	• Population level of *S.* Typhimurium increased sharply and decreased after 96 hours in carrot and celery juice. • Population level of *S.* Typhimurium increased continuously and remained relatively constant in kale and beet juice (5.0 and 4.4 Log CFU/ml). • Population level of *S.* Typhimurium gradually decreased in grapefruit, cabbage, and red cabbage juice. • Population level of *S.* Typhimurium decreased below limit detection in lemon juice immediately after inoculation.	(281)

S. Typhimurium	Strawberry (Postharvest)	4 °C	Hours 0 and 72	2.1 Log CFU/g	• Population level of *S. enterica* decreased to 1.3 log CFU/g	(296)
		25 °C		4.6 Log CFU/g	• Population level of *S. enterica* increased to 7.8 log CFU/g	
S. enterica cocktail including serotype Baildon, Braenderup, Montevideo, Newport, Poona, and Saintpaul	Mango (Postharvest)	20 °C	Hours 0 and 24	6.6 Log CFU/mango 4.3 Log CFU/mango	• Population level of *S. enterica* remained at 6.5 Log CFU/mango • Population level of *S. enterica* remained at 4.3 Log CFU/mango	(297)
S. enterica Montevideo, Typhimurium, and Typhi	Strawberry (Postharvest)	22 °C 3 °C	Days 0 and 15	4.0 Log CFU/g	• Population level of *S. enterica* decreased to ca. 2.0 Log CFU/g. • Population level of *S. enterica* decreased by ca. 1.3 Log CFU/g.	(280)
S. enterica including serotype Enteritis, Newington, Stanley, and Thompson	Whole kiwi fruit (Postharvest)	22 °C	Days 0, 5, 10, 15, 20, 25, and 30	7.0 Log CFU/kiwi fruit 4.0 Log CFU/kiwi fruit	• Culturable *S. enterica* fell below limit of detection on Day 10 but remained enrichment positive on Day 30 for both green and golden kiwi fruits. • On Day 10, culturable *S. enterica* on green kiwi fruit was higher than golden kiwi fruits and remained enrichment positive on Day 30 for both green and golden kiwi fruits.	(279)
	Sliced kiwi fruit (Postharvest)	4 °C	Days 0 and 7	4.0 and 7.0 Log CFU/kiwifruit (low and high inoculation	• In the high-inoculation scenario, population level of *S. enterica* was decreased by ca. 3.0 Log CFU/kiwifruit after 7 days.	

				levels, respectively)	• In the low-inoculation scenario, population level of *S. enterica* was decreased by ca. 1.0 Log CFU/kiwifruit after 7 days.	
S. enterica. cocktail	Grapefruit (Postharvest)	20 °C	Days 0, 1, 5, 7, and 14	7 log CFU/grapefruit	• Population level of *S. enterica* reduced by 4.7 Log CFU/grapefruit	(298)
S. enterica cocktail including serotype Montevideo and Typhimurium	Apple (Postharvest)	22 °C	Hours 0, 3, 12, 24, and 48	2.3-3.8 Log CFU/apple	• Population level of *S. enterica* increased by up to 5 log CFU/apple in Rayada and Red Delicious, whereas 4.8 log CFU/apple in Golden Delicious at 48 h	(299)
		15 °C	Hours 0, 3, 24, 48, and 96		• Population level of *S. enterica* steadily increased to 4.1 and 4.3 log CFU/apple	
		5 °C			• Population level of *S. enterica* steadily decreased to 3.0 log CFU/apple	
S. enterica including serotype Newport, Poona, and Typhimurium	Fresh-cut cantaloupe, honeydew, pineapple, radish, and watermelon (Postharvest)	4 °C	Days 0 and 7	3.3 Log CFU/ g	• Population level of *S. enterica* decreased sharply to 0.2 Log CFU/g in 2 days on pineapple but persisted at ca. 2.6 Log CFU/g on rest of fruits.	(100)
		8 or 12 °C			• Population level of *S. enterica* increased significantly on cantaloupe, honeydew and watermelon, reaching up to 4.4 and 5.3 Log CFU/g, at 8 °C and 12 °C, respectively. • Population level of *S. enterica* continuously decreased to 0.2 Log CFU/g or lower on pineapple by the end of 7-day storage at 8 °C and 12 °C. • Population level of *S. enterica* stayed at 2.7 Log CFU/g at 8 °C and slightly	

					increased to 3.2 Log CFU/g on radish at 12 °C.	
		35 °C for 2 h, then stored at 4 °C			• Population level of *S. enterica* decreased sharply to 0 Log CFU/g in 2 days on pineapple (comparably slower than storage at 4 °C without temperature abuse) but persisted at ca. 2.5 Log CFU/g on rest of fruits.	
S. enterica cocktail	Grape (Postharvest)	37 °C	Days 0 and 15	2.0 Log CFU/g	• Population level of *S. enterica* reduced by 1.0 Log CFU/ml	(300)
S. enterica cocktail including serotype Agona, Gaminara, Michigan, Montevideo, and Enteritidis	Strawberry (Postharvest)	4 °C 20 °C	Days 0, 2, 6, 9, 12 and 14 Hours 0, 8, 24, 30, and 48	4.5-4.6 Log CFU/wound	• Population level of *S. enterica* reduced by 3.4 Log CFU/wound • Population level of *S. enterica* reduced by 1.2 Log CFU/wound	(301)
S. enterica cocktail including serotype Poona, Panama, and Stanley,	Fresh-cut apple (Postharvest)	4 °C	Days 0, 1, 7, 14, 21 and 35	4.2 Log CFU/	• The remaining population level on fresh-cut apple slices was 2.2 Log CFU/	(302)

S. enterica including serotype, Poona, Panama, and Stanley	Sliced apple (Postharvest)	4 °C	Days 0, 1, 7, 14, 21, and 35	4.2 Log CFU/	• Population level of *S. enterica* decreased to 2.2 Log CFU/ after 35 days.	(303)
S. Typhimurium	Papaya (Postharvest)	7 °C	Days 0, 1, 7, 10, and 14	5.5 Log CFU/g	• Population level of *S. enterica* decreased to 4.1 log CFU on day 7 and then increased to 6.2 log CFU at the end of the storage period.	(285)
		21 °C			• Population level of *S. enterica* gradually increased to 7.34 log CFU on Day 14.	
S. enterica cocktail including serotype Agona, Newport, Poona, and St. Paul	Papaya (Postharvest)	4 °C	Days 0, 1, 3, 5, 7, 9, 11, and 14	4.0-5.0 Log CFU/g	• When RH = 90%, population level of *S. enterica* was decreased by 0.35, 0.46, 1.00, and 0.85 log CFU/g for 0, 25, 50, and 75 % fruit ripeness. • When RH = 55%, population level of *S. enterica* was decreased by 1.4, 1.5, 1.4, and 1.4 log CFU/g were observed for 0, 25, 50, and 75 % fruit ripeness.	(110)
		12 °C	Days 0, 1, 3, 5, 7, 9, 11, and 14		• When RH = 90% and 55%, population level of *S. enterica* was decreased by <2.0 Log CFU/g for papaya with different ripeness.	
		21 °C	Days 0, 1, 3, 5, and 7		• When RH = 90%, population level of *S. enterica* was increased by 1.59 to 2.06 Log CFU/g on papaya for 50 and 75% fruit ripeness levels but remained at the similar level at 0 and 25% of ripeness levels after 7 days.	

					• The overall growth of *S. enterica* was higher in 50 and 75% fruit ripeness than 0 and 25% fruit ripeness papaya. • When RH = 55%, population level of *S. enterica* decreased by 2.11, 1.54, 1.86, 0.90 Log CFU/g at 0, 25, 50, 75% of fruit ripeness. •	
S. enterica cocktail including serotype Newport and Typhimurium	Fresh-cut apple, cantaloupe, cucumber, and tomato (Postharvest)	4 °C	Days 0, 1, 3, 5, and 7	7.0 and 2.0 Log CFU/g for high and low inoculation levels, respectively	• Less than 0.2 log reduction observed on tomato and cantaloupe and 0.5 log reduction on cucumber and apple at the end Day 7. • *S. enterica* behaved similarly when inoculation level was low (< 0.5 log variation) on tomato and cantaloupe.	(304)
S. enterica cocktail including serotype Enteritidis, Hadar, Montevideo, and Thompson,	Dates (Postharvest)	4 °C	Days 0, 10, 20, and 31	8.1 Log CFU/g	• Population level of *S. enterica* remained at 8.2 Log CFU/g on Day 31.	(305)
S. enterica cocktail including serotype Agona, Gaminara, Montevideo, Michigan, and Poona.	Avocado (Postharvest)	5 °C 25 °C	Days 0, 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, and 48	7.4 Log CFU/ avocado 7.2 Log CFU/ avocado	• Population level of *S. enterica* decreased to 1.7 Log CFU/ avocado after 48 days. • Population level of *S. enterica* decreased to 4.4 Log CFU/ avocado after 48 days.	(287)
S. enterica including	Strawberry (Postharvest)	4 °C	Days 0, 1, 3 and 7	6.0 Log CFU/g	• Population level of *S.* Newport, *S.* Tennessee, and *S.* Thompson	(288)

serotype Newport, Tennessee, and Thompson	10 °C	decreased to 3.5, 3.7 and 4.1 Log CFU/g on Day 7, respectively. Population level of *S.* Newport, *S.* Tennessee, and *S.* Thompson decreased to 3.3, 3.6 and 3.8 Log CFU/g on Day 7, respectively.

NA: not available

Table 7. Summary of studies on the application of antagonistic cultures in inhibiting major foodborne pathogens in fresh produce between 2010 and 2022

Targeted foodborne pathogen	Antagonistic culture (isolation source)	Antagonist group	Antimicrobial activity test	Food vehicle	Major results	References
E. coli O157:H7 (~4 Log CFU/ml)	*Bacillus* spp., *Bacillus pumilus, Erwinia perscina, Paenibacillus polymyxa, Pantoea agglomerans, Pseudomonas* spp. (Fresh spinach) (~ 4 Log CFU/ml for each strain).	Non-LAB	Spot-on-lawn	Spinach	• 1.3 – 2.8 Log CFU/ml reduction in *E. coli* O157:H7 populations was observed when co-cultured with antagonists in TSB broth compared to the control (only *E. coli* O157:H7) stored at 25 °C after 48 hours. • 0.3 – 1.3 Log CFU/g reduction in *E. coli* O157:H7 was measured when co-inoculated with antagonists on spinach leaves after 24 h at 25 °C compared to the control. • Cellobiose was used by *Erwinia perscina* to enhance inhibition of *E. coli* O157:H7.	(128)
E. coli O157:H7 gfp + (~8 Log CFU/ml)	*Rhodococcus* sp. (Spinach) (~8 Log CFU/ml)	Non-LAB	Perpendicular streak	Spinach leaves	• Seed inoculation mitigates more culture *E. coli* O157:H7 gfp + than	(171)

					canopy spray inoculation. • *Rhodococcus* sp. reduced by the 1.5 Log CFU/g of *E. coli* O157:H7 gfp+ co-inoculated with the antagonist on seeds at 20 °C (day) and 16 °C (night) after 28 days.	
L. monocytogenes (~ 4, 6, and 8 Log CFU/g)	*Lactiplantibacillus plantarum* B2 and *Lactobacillus fermentum* PBCC11.5 (~ 8 Log CFU/g for each strain)	LAB	N.A.	Fresh-cut cantaloupe	• When the concentration of inoculated *L. monocytogenes* was the same as that of the probiotic bacteria (~ 8 Log CFU/g), no reduction in *L. monocytogenes* population in fresh-cut cantaloupe caused by LAB compared to that in the control group was observed after 9 days at 4 °C. • When the concentration of inoculated *L. monocytogenes* was 6 Log CFU/g, the growth reduction of 1 Log CFU/g compared to the	(168)

					control was obtained at the end of storage. • When the concentration of inoculated *L. monocytogenes* was 4 Log CFU/g, *L. fermentum* PBCC11.5 caused 3 Log CFU/g growth reduction of *L. monocytogenes*, while *L. plantarum* B2 achieved 2 Log CFU/g of *L. monocytogenes* reduction compared to the control at the end of storage.	
S. enterica? Enteritidis (~ 4 Log CFU/g)	*Lactiplantibacillus plantarum* and *Lactobacillus brevis* (Chinese traditional fermented vegetables) (~ 7 Log CFU/g)	LAB	Well diffusion assay	Fresh-cut apples	• *L. plantarum* RD1 strain exhibited the best inhibitory activity against *S.* Enteritidis on fresh-cut apples stored at 10 °C after 7 days, which was that • RD1 reduced by 1.4 Log CFU/g of *S.* Enteritidis at the end of storage.	(306)
S. enterica ATCC4931r (~3 Log CFU/g)	*Erwinia persicina* strain EUS78 (various vegetable seeds) (~ 5-8 Log CFU/g)	Non-LAB	Spot-on-lawn	Alfalfa sprout	• EUS78 remarkably inhibited the growth of *S. enterica* ATCC4931r on	(189)

					alfalfa sprouts with up to 5.0 Log CFU/g after 7 days at 25 °C.	
E. coli O157:H7 and *L. monocytogenes* (~7 Log CFU/g for each strain)	*Lactiplantibacillus plantarum* and *Lactobacillus fermentum* (sourdough) (~ 8 Log CFU/g for each strain)	LAB	N.A.	Pineapples	• *L. plantarum* reduced the amount of *E. coli* O157:H7 and *L. monocytogenes* by about 1.1 Log CFU/g on pineapple plugs at 5 °C after day 7 compared to the *E. coli* O157:H7 and *L. monocytogenes* control groups. • *L. fermentum* induced the reduction in 2.0 Log CFU/g of *L. monocytogenes* populations compared to that in *L. monocytogenes* control group at 5 °C after day 7.	(129)
Enterohaemorrhagic *E. coli* (~ 7 Log CFU/g) and *S. enterica* cocktail (~6 Log CFU/g)	*Bacillus subtilis* (LCA1, M24, and 168) (inner tissue of mung bean seeds or lettuce stems) (~ 5 Log CFU/g for LCA1 and M24; 4 Log CFU/g for 168)	Non-LAB	Spot-on-lawn	Mung bean sprouts	• All three *B. subtilis* strains reduced by the 1.1-1.4 Log CFU/g of *S. enterica* in seed at 22 °C by Day 5. • LCA1-cat strain reduced by the greatest amount of	(117)

						S. enterica, which was 1.4 Log CFU/g when compared to the control, after 5 days of sprouting at 22 °C. • LCA1-cat strain reduced by the greatest amount of EHEC, which was 2.0 Log CFU/g when compared to the control, after 5 days of sprouting at 22 °C.	
	E. coli ATCC 25922 and *S.* Typhimurium ATCC 14028 (~6 Log CFU/ml for each strain)	*Lactiplantibacillus plantarum* (B135 and Z183) and *Pediococcus pentosaceus* (B125) (Amazonian Acai fruits) (~ 6 Log CFU/ml for each strain)	LAB	Spot-on-lawn	Amazonian Acai fruits juice	• B125, B135, and Z183 could inhibit the growth of *E. coli* in Acai juice stored at room temperature after 3 days. Among them, B135 inhibited the greatest number of *E. coli* with 8.0 Log CFU/ml after 3 days of storage compared to the control. • All three strains reduced 6.0 Log CFU/ml of *Salmonella* Typhimurium in Acai juice stored at	(190)

					room temperature after 3 days.	
L. monocytogenes and *S.* Enteritidis (~5 Log CFU/ml for each strain)	*Pseudomonas graminis* strain CPA-7 (apple surface) (~ 7 Log CFU/ml)	Non-LAB	N.A.	Fresh-cut pear	• CPA-7 reduced the levels of *L. monocytogenes* and *S.* Enteritidis by 5.5 and 3.1 Log CFU/g compared to the controls on fresh-cut pear after 7 days at 10 °C.	(169)
L. monocytogenes and *Salmonella* and (~ 5 Log CFU/ml for each strain)	*Lacticaseibacillus rhamnosus* GG (~ 8 Log CFU/ml)	LAB	N.A.	Fresh-cut pear	• 1.8 Log units of *L. monocytogenes* were reduced by *L. rhamnosus* GG co-inoculated on fresh-cut pear stored at 5 °C after 9 days compared to the control. However, no reduction of *Salmonella* was observed at the end of the storage.	(188)
L. monocytogenes and *Salmonella* (~5 Log CFU/ml for each strain)	*Pseudomonas graminis* CPA-7 (apple surface) (~ 7 Log CFU/ml)	Non-LAB	N.A.	Fresh-cut pear	• *Salmonella* and *L. monocytogenes* had the most reduction with 2.2 and 4.0 Log units by CPA-7 co-inoculated on fresh-cut pear after 10 days at 10 °C compared to the controls.	(170)

E. coli O157:H7 (~5 Log CFU/plug), *Listeria innocua* (~4 Log CFU/plug), and *S. enterica* (~5 Log CFU/plug)	*Pseudomonas graminis* CPA-7 (apples, peaches, and pineapples) (~ 6-8 Log CFU/ml)	Non-LAB	*In vitro* test	Fresh-cut apples and peaches	• For fresh-cut apples, *P. graminis* CPA-7 reduced the populations of *E. coli* O157:H7, *S. enterica*, and *L. innocua* by 4.5, 4.7, and 5.9 Log units, respectively, compared to the control after 2 days at 20 °C. • For fresh-cut peaches, *P. graminis* CPA-7 reduced the populations of *E. coli* O157:H7, *S. enterica* and *L. innocua* by 4.3, 2.8, and 4.0 Log units, respectively, compared to the control after 2 days at 20 °C.	(126)
E. coli O157:H7, *L. monocytogenes*, and *S. enterica* (~ 7 Log CFU/ml for each strain)	*Pseudomonas* sp. strain M309 (whole vegetables, fresh-cut vegetables, and sprouts) (~ 7 Log CFU/ml)	Non-LAB	*In vivo* assay	Fresh-cut lettuce	• *Pseudomonas* sp. strain M309 reduced *S. enterica* and *L. monocytogenes* counts by 3.3 and 2.2 Log units compared to the control on lettuce disks at 20 ℃ at the	(70)

					end of 3-day storage. • M309 strain could also reduce *S. enterica* and *E. coli* O157 on lettuce disks by 3.5 and 3.7 Log units compared to the control at 10 °C after 3 days.	
E. coli, *Staphlococcus aureus*, *Bacillus cereus*, *Pseudomonas aeruginosa*, *Proteus vulgaris*, *S.* Typhi, *Serritia* spp., and *Xanthomonas campestris* (the level is N.A.)	*Salimicrobium*, *Bacillus*, *Paenibacillus*, and *Lactobacillus* (curd and cow dung samples) (the level is N.A.)	Non-LAB and LAB	Agar well diffusion	Apple, Banana, Chikku, Fig, Pomegranate, and Strawberry	• LAB had antagonistic activity against the selected pathogenic bacteria. • The bacteriocin-like inhibitory substance produced by LAB coating on various fruits was able to extend their shelf-life at room temperature.	(127)
E. coli ATCC 25922, *S.* Typhimurium ATCC 13311, *S.* Paratyphi A, *Pseudomonas aeruginosa* ATCC 27853, *Staphylococcus aureus* ATCC 25923, *Bacillus cereus* DMST 5040, and *Bacillus subtilis* (~ 8 Log CFU/ml)	*Rummeliibacillus pycnus* STR 0103f and *Rummeliibacillus stabekisii* STR 0404f (fermented vegetable products) (~ 8 Log CFU/ml)	LAB	Agar well diffusion	N.A.	• The two strains elicited the broad spectrum against both Gram-negative and Gram-positive pathogens. • The cultures can tolerate bile salt tolerance at 0.3% bile salt for 24 h with survival rates of ~ 67.77 and 68.72%.	(134)

					• The cultures can survive under acid conditions (pH 3-5) with survival rates of about 60.53-76.87%.

LAB: lactic acid bacteria; N.A.: not available.

Table 8. Characteristics of fresh vegetable microbiota studies published from 2010-2022.

Food vehicle	Sequencing platform	Sequencing region (primers)	Pipeline	Key factors	Most abundant taxa (Top three)	Differential abundance analysis	References
Spinach	Roche 454 Pyrosequencing	V4 (515f and 806r)	• RDP pyrosequencing pipeline	• Storage (15 days)* • Packaging (fresh vs. packaged) • Temperature (4 vs. 10 °C)	Phyla: Proteobacteria Actinobacteria, and Acidobacteria (fresh spinach); Proteobacteria, Actinobacteria, and Deinococcus-Thermus (4 °C for 15 days); Proteobacteria and Actinobacteria (10 °C for 15 days). Genera: Unclassified Enterobacteriaceae, *Pseudomonas*, and unclassified Oxalobacteraceae (fresh spinach); Unclassified Enterobacteriaceae, *Pseudomonas*, and *Methylobacterium* (4 °C for 15 days); Unclassified Enterobacteriaceae, *Pseudomonas*, and *Escherichia* (10 °C for 15 days).	N.A.	(192)

Fresh produce	Roche 454 Pyrosequencing	V5-V6 (799f and 1115r)	• QIIME 1	• Produce types* (Sprouts, spinach, lettuce, mushroom, and peppers) • Farming practices (organic vs. conventional)	Phyla: Proteobacteria, Actinobacteria, and Firmicutes. Genera: *Pantoea*, *Klebsiella*, and *Xanthomonas* (across all samples).	N.A.	(193)
Leaf greens	Roche 454 Pyrosequencing	V6-V8 (939f and 1392r)	• Mothur	• Produce types* (Romaine lettuce, baby spinach, green leaf lettuce, iceberg lettuce, and red leaf lettuce) • Farming practices (organic vs. conventional) • Leaf surface sterilization (sterilized vs. non-sterilized)	Phyla: Proteobacteria, Bacteroidetes, and Firmicutes. Genera: *Pseudomonas*, *Ralstonia*, and *Flavobacterium* (across all samples).	N.A.	(202)
Romaine lettuce	Roche 454 Pyrosequencing	V5-V9 (799f and 1492r)	• QIIME 1	• Planting seasons* (Early: June vs. Late: August and October) • Irrigation (sprinkler vs. drip) • Leaf age • Inoculation (attenuated *Escherichia coli* O157:H7)	Phyla: Proteobacteria, Firmicutes, and Actinobacteria. Genera: *Leuconostoc*, *Lactococcus*, and *Bacillus* (early season) *Pantoea*, unclassified Enterobacteriaceae, and *Pseudomonas* (late season).	N.A.	(237)

				• Harvest time (7, 14, and 21 days after inoculation)			
Fresh herbs	Illumina MiSeq Sequencing	V4 (515f and 806r)	• Mothur	• Herb types* (Basil, Chives, Dill, Mint, Rosemary, and Thyme)	Phyla: Firmicutes, Bacteroidetes, and Proteobacteria. Genera: Unclassified *Bacillaceae*, *Streptococcus*, and *Prevotella*.	N.A.	(186)
Leafy greens	Illumina MiSeq Sequencing	V4 (515f and 806r)	• Mothur	• Farm locations (Farm Vestfold vs. Farm Buskerud) • Produce types* (rocket salad and lettuce) • Cultivars (Frillice, Iceberg, and Little Gem) • Leaf maturity (3 weeks vs. harvest)	Phyla: Proteobacteria, Actinobacteria, and Bacteroidetes. Genera: *Pseudomonas*, *Duganella*, and *Pantoea* (lettuce); *Pseudomonas*, *Flavobacterium*, and *Chryseobacterim* (rocket salad).	N.A.	(231)
Cilantro	Roche 454 Pyrosequen-cing and Illumina MiSeq shotgun sequencing	V1-V3 (27f and 533r)	• QIIME 1 and CLC Genomics Workbench software	• Enrichment (before vs. after)	Phyla: Proteobacteria, Bacteroidetes, and Actinobacteria (Before enrichment). Firmicutes, Proteobacteria, and Bacteroidetes (after enrichment). Families:	N.A.	(238)

						Oxalobacteraceae, Pseudomonadaceae, and Flavobacteriaceae (before enrichment); Peptostreptococcaceae, Planococcaceae, and Clostridiaceae (after enrichment).		
Mung bean sprouts	Illumina HiSeq2500 Sequencing	V1-V3 (27f and 534r)	• CLC Genomics Workbench software and NCBI BLAST	• Enrichment strategies (BPW vs. EE-broth) • Temperatures (37 vs. 42 °C)		Phyla: Proteobacteria, Bacteroidetes, and Firmicutes (before enrichment); Proteobacteria, Firmicutes, and Bacteoidetes (BPW at both temperatures for 24 h); Proteobacteria, Bacteroidetes, and Firmicutes (EE-broth at both temperatures for 24 h) Genera: *Janthinobacerium*, *Pseudomonas*, and *Enterobacter* (before enrichment); *Klebsiella*, *Acinetobacter*, and *Bacillus* (BPW at 42 °C for 24 h); *Klebsiella*, *Enterobacter*, and *Cronobacter* (EE-	Fisher's exact test	(264)

					broth at both temperatures for 24 h)		
Radish sprouts	Ion Torrent Pyrosequen-cing	V5-V6 (799f and 1115r)	• CLC Genomics Workbench software, MG-RAST, and NCBI BLAST	• Seasons* (Summer and winter) • Sprouting (pre-sprouted seeds vs. sprouted samples)	Families: Pseudomonadaceae, Oxalobacteraceae, and Flavobacteriaceae (Winter); Pseudomonadaceae, Oxalobacteraceae, and Enterobacteriaceae (Summer). Genera: *Halospirulina* and *Pseudomonas* (pre-spouted seeds) *Pseudomonas*, *Paenibacillus*, and *Pantoea* (sprouted samples).	N.A.	(194)
Cilantro, cucumber, and mung bean sprout	Illumina MiSeq Sequencing	V1-V3 (27f and 534r)	• QIIME 1	• Produce types* (cilantro, cucumber, mung bean sprout) • Brands (brand A and brand B of mung bean sprout)	Genera: *Pseudomonas*, *Flavobacterium*, and *Duganella* (cilantro); *Acinetobacter*, *Pseudomonas*, and *Pantoea* (cucumber); *Leuconostoc*, *Lactococcus*, and *Enterobacter* (brand A of mung bean sprout); *Enterobacter*, unclassified Enterobacteriaceae, and *Pseudomonas*	N.A.	(232)

					(Brand B of mung bean sprout).		
Spinach	Illumina MiSeq Sequencing	V4 (515f and 806r)	• QIIME 2	• Washing process (chlorine-washed vs. unwashed) • Storage (Day 0 vs. Day 7) • Temperatures (4, 10, and 15 °C) • Planting locations* (CA vs. AZ)	Phyla: Proteobacteria, Bacteroidetes, and Firmicutes (CA); Proteobacteria, Bacteroidetes, and Actinobacteria (AZ). Genera: *Pseudomonas*, *Acinetobacter*, and *Flavobacterium* (CA); *Pseudomonas*, *Sphingomonas*, and *Acinetobacter* (AZ).	N.A.	(75)
Broccoli	Roche 454 Pyrosequen-cing	V5-V6 (799f and 1115r)	QIIME 1 •	• Harvest stages (preharvest vs. postharvest)	Phyla: Proteobacteria, Actinobacteria, and Firmicutes. Genera: *Pseudomonas*, *Acinetobacter*, and *Janthinobacterium* (preharvest); *Pseudomonas*, *Hafnia*, and *Acinetobacteria* (postharvest).	N.A.	(224)
Lettuce	Illumina MiSeq Sequencing	V4 (515f and 806r)	• QIIME 1	• Inoculation (*Salmonella* Infantis) • Sample types* (rhizosphere and root)	Phyla: Proteobacteria, Verrucomicrobia, and Actinobacteria (rhizosphere); Proteobacteria,	N.A.	(233)

				• Soil textures (silt vs. loamy sand) • Cultivars (Green Star vs. Green Salad Bowl) • Harvest times (4 vs. 5 weeks)	Bacteroidetes, and Actinobacteria (root).		
Sugar beet	Illumina MiSeq Sequencing	V4-V5 (515f and 926r)	• QIIME 1 and QIIME 2	• Health status* (healthy vs. decaying) • Farm Locations (Austria vs. Germany) • Clamp types (Grossmugl, Mittich, Kleinweichs1, Kleinweichs2, Osterhofen1, and Osterhofen2)	Orders: Micrococcales, Flavobacteriales, and Pseudomonadales (Healthy samples); Lactobacillales, Rhodospirllales, and Flavobacteriales (Decaying samples).	ANCOM	(195)
Alfalfa and mung bean sprouts	Illumina MiSeq Sequencing	V3-V4 (341f and 806r)	• Mothur	• Sprout types* (alfalfa sprouts vs. mung bean sprouts) • Brands (c1 vs. c2) • Storage (3 weeks for alfalfa and 2 weeks for mung bean sprouts at 4 °C)	Genera: *Pseudomonas*, *Pantoea*, and *Janthinobacterium* (Alfalfa sprouts); *Pseudomonas*, *Pantoea*, and *Acinetobacter* (Mung bean sprouts).	N.A.	(234)
Perilla leaf	Illumina MiSeq Sequencing	V5-V6 (799f and 1115r)	• Mothur and QIIME 1	• Harvesting times* (April vs. July) • Sampling sites (A, B, C, D, and E)	Phyla: Proteobacteria, Firmicutes, and Actinobacteria. Genera:	N.A.	(307)

						Unclassified Enterobacteriaceae, *Methylobacterium*, and *Sphingomonas* (April); *Bacillus*, *Exiguibacterium*, and *Acinetobacter* (July).		
	Ready-to-eat salad (arugula, lamb's lettuce, and radicchio)	Illumina MiSeq Sequencing	V3-V4 (341f and 806r)	• Prinseq-lite program, FLASH program, and QIIME 2	• Brands (S1, S2, and S3) • Nucleic acids* (16S rRNA vs. 16S rDNA)	Phyla: Proteobacteria, Bacteroidetes, and Actinobacteria. Genera: *Pseudomonas*, unclassified genera, and *Rahnella* (16S rRNA); *Pseudomonas*, *Flavobacterium*, and *Rahnella* (16S rDNA).	N.A.	(308)
	Spinach	Illumina MiSeq Sequencing	V4 (505f and 806 r)	• QIIME 2	• Types of irrigation water* (Gr, Wa, and Rf) • Irrigation times (Day 0 vs. Day 7)	Phyla: Proteobacteria, Bacteroidetes, and Actinobacteria. Genera: *c-ZB2*, unclassified Coxiellaceae, and *p-OD1* (Gr); *Polynucleobacter*, *Sediminibacterium*, and *Flavobacterium* (Wa); *Bradyrhizobium*, unclassified Comamonadaceae, and *Novosphigonium* (Rf).	N.A.	(309)

Sweet pepper	Illumina MiSeq Sequencing	V4-V5 (515f and 909r)	• QIIME 1	• Habitats (hydrophonic vs. soil) • Treatments (fungicide-treated vs. untreated) • Maturity stages* (red vs. green)	Phyla: Proteobacteria, Firmicutes, and Actinobacteria. Genera: *Acinetobacter*, *Pseudomonas*, and *Burkholderia* (green); *Pseudomonas*, *Acinetobacter*, and *Bacillus* (red).	N.A.	(225)
Spinach	Illumina MiSeq Sequencing	V3-V4 (341f and 806r)	• Phyloseq R packages, and QIIME 2	• Storage (Day 0, 5, and 10 at 4 °C) • Sanitizer types* (no treatment, tap water, PAA 50ppm, ECAS 50 ppm, and ECAS 85ppm)	Phyla: Proteobacteria, Bacteroidetes, and Firmicutes (no treatment); Firmicutes, Actinobacteria, and Proteobacteria (tap water); Proteobacteria, Actinobacteria, and Bacteroidetes (PAA 50ppm); Proteobacteria, Actinobacteria, and Bacteroidetes (ECAS 50ppm); Proteobacteria, Bacteroidetes, and Actinobacteria (ECAS 85ppm).	ANCOM	(196)
Sprouts	Illumina MiSeq Sequencing	V4 (515f and 806r)	• QIIME 2	• Sprout types* (alfalfa, radish, and rapeseed) • Brands (A, B, and C)	Phyla: Proteobacteria, Bacteroidetes, and Firmicutes. Genera:	LEfSe	(235)

				• Distribution routes (online vs. offline)	*Unclassified Enterobacteriacae*, *Acintobacter*, and *Janthinobacterium* (alfalfa sprouts); *Chryseobacterium*, *Acinetobacter*, and *Flavobacterium* (radish sprouts); *Pseudomonas*, *Chryseobacterium*, and *Janthinobacterium* (rapeseed sprouts).		
Spring Mix salad	Illumina MiSeq Sequencing	V3-V4 (314f and 785r)	• QIIME 2	• Inoculation (*E. coli* O157:H7) • Brands* (A, B, and C) • Storage (Day 0, 5, 10, and 15)	Phyla: Proteobacteria, Bacteroidetes, and Actinobacteria (Brands A and C); Proteobacteria, Bacteroidetes, and Firmicutes (Brand B). Genera: *Flavobacterium*, *Pseudomonas*, and *Janthinobacterium* (across all samples)	LEfSe and ANCOM	(81)
Romaine lettuce	Illumina MiSeq Sequencing	V4 (515f and 806r)	• QIIME 2	• Inoculation* (non-inoculated vs. *Listeria monocytogenes*-inoculated) • Storage temperatures (4, 12, and 24 °C)	Phyla: Proteobacteria, Firmicutes, and Actinobacteria. Genera: *Pseudomonas*, unclassified *Leuconostocaceae*, and *Weissella* (non-	N.A.	(197)

					• Storage (Days 0, 4, and 8)	inoculated); *Pseudomonas*, unclassified Enterobacteriaceae, and *Pantoea* (inoculated).		
Romaine lettuce	Illumina MiSeq Sequencing	V3-V4 (314f and 785r)	• QIIME 2		• Brands* (A, B, and C) • Seasons (Early vs. Late) • Storage (Day 0 vs. Day 5)	Genera: *Pseudomonas*, *Leuconostoc*, and *Pantoea* (Brand A); *Pseudomonas*, *Pantoea*, and *Xanthomonas* (Brand B); *Pseudomonas*, *Weissella*, and *Serratia* (Brand C).	ANCOM	(226)

The * highlights the most impactful factor on fresh vegetable microbiota. BPW: buffered peptone water; EE-broth: Enterobacteriaceae enrichment broth; V: Variable regions of 16S rRNA genes; ANCOM: Analysis of composition of microbiomes; QIIME: Quantitative insights into microbial ecology; N.A.: Not available; NCBI: The national center for biotechnology information; EMBL-EBI: The European molecular biology laboratory-European bioinformatics institute; CA: California; AZ: Arizona.

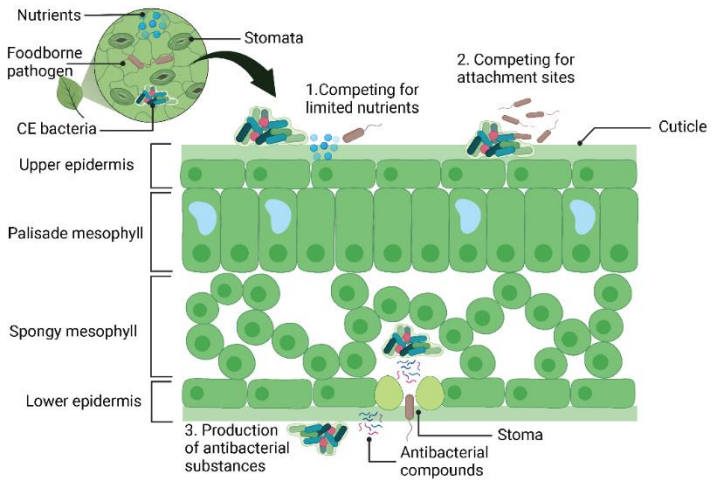

Figure 1. Mechanisms of antagonistic activity of antagonistic bacteria against foodborne pathogens in fresh produce, created with BioRender.com.

Chapter 2. The microbial quality of commercial chopped Romaine lettuce before and after the "Use By" date

Running title: Microbial quality, use-by date, Romaine lettuce

Chao Liao [a] and Luxin Wang [a*]

[a] Department of Food Science and Technology, University of California Davis, Davis, CA 95616, USA

*Corresponding author: Luxin Wang, Associate Professor, University of California Davis.

Email address: lxwang@ucdavis.edu. Institutional address: Department of Food Science and Technology, University of California Davis, Davis, CA 95616

Published in Frontiers in Microbiology 13: 850720-850720.

https://doi.org/10.3389/fmicb.2022.850720

Abstract

In the United States, due to the limited information about the microbial quality and safety of fresh produce after the labeled open dates, unnecessary discarding of fresh produce in good conditions and food loss have been caused. The aim of this study was to address this knowledge gap and evaluate the microbial quality of commercial chopped Romaine lettuce (RL) on the "Use By" dates (UBD) and 5 days after the "Use By" dates (UBD5). The microbial quality was evaluated using culture-dependent and culture-independent methods. Three brands of RL samples, from early and late harvest seasons, were purchased from local grocery stores and stored at 4 °C until 5 days after their UBD. On the UBD and UBD5, bagged lettuce was opened, homogenized, diluted, and plated onto plate count agar and anaerobic agar to obtain total aerobic plate counts (APC) and total anaerobic plate counts (AnPC). For the culture-independent method, DNA was extracted from each sample homogenate and used for 16S rRNA gene sequencing. The culture-dependent results showed that there was no significant change in APC or AnPC between UBD and UBD5 samples. The APC and AnPC ranged from 5.71 ± 0.74 to 7.89 ± 0.10 Log CFU/g and 1.75 ± 0.08 to 7.32 ± 0.61 Log CFU/g, respectively. No significant difference in alpha diversity, based on observed features and Shannon index values, was detected between UBD and UBD5 samples using 16S rRNA sequencing. Similarly, no difference was observed in beta diversity based on the Jaccard distance matrixes and the weighted Unifrac distance matrixes. Taxonomic analysis revealed 128 genera in all RL samples. The top five genera were *Pseudomonas* (with relative abundance ranging from 16.47-92.72%), *Serratia* (0-52.35%), *Weissella* (0-42.42%), *Pantoea* (0.17-21.33%), and *Lactococcus* (0-24.30%). The differential abundance analysis based on the ANCOM test showed that no bacteria were detected to have significantly differential abundance in RL between UBD and UBD5. In summary, both the culture-dependent and culture-independent results showed that there was no significant difference in the microbial quality of RL before and shortly after the UBD.

Keywords: Romaine lettuce; "Use By" date; Microbial food quality; Plate counting; 16S rRNA gene sequencing; Bacterial communities.

Introduction

Although fruits and leafy greens are essential components of a healthy diet, they are also one of the most wasted foods, with up to 63% of fresh fruits and 70% of leafy greens wasted (1). The United States Department of Agriculture (USDA) estimated that approximately 31% of the 430 billion pounds of food produced in 2010 (valued at ca. $161.6 billion) was lost and not available for human consumption at the retail and consumer levels (2). Vegetables (19 %) ranked the second on the list of food groups with high loss rates (2). In 2015, the USDA and Environmental Protection Agency (EPA) announced that the U.S. 2030 Food Loss and Waste Reduction goal was to reduce 50% of food loss and waste by the year 2030 (3, 4). There are many causes and drivers for food loss and waste at the retail level; consumers' confusion over "Use By" and "Best Before" dates and other date labeling is one of them (5, 6). Date labeling has been practiced for decades, and its main aims include informing stock rotation at retail and facilitating potential recalls and traceability. However, the terminologies used vary widely around the world (5).

Such variations in date labeling terms have contributed to substantial misunderstanding by the industry and consumers and lead to significant unnecessary food loss and waste, especially in the United States (5). Recent studies showed that approximately 37% of American consumers usually discarded fresh produce when the open date is past even when the product had no quality and safety problems (7, 8). In the U.S., three date labels are the most applied for food products: "Best if Use By" or "Best Before", "Use By", and "Sell By" (8). Tsiros and Heilman (2005) summarized the three commonly used label phrases: (1) "Best before", which indicates the date after which a product loses the best quality for consumption. (2) "Use by", indicating the date after which a product is no longer remaining sufficient quality and should not be consumed, but not necessarily associated with food safety. (3) "sell by", suggesting the last day on which a product should be sold (8, 9). Among these three phrases, "Use By" generates the greatest value of predicted waste (8).

In contrast to the U.S., the regulation No. 1169/2011 of the European Parliament and the Council of the European Union (EU) mandates defined the use of a date of minimum durability or best before

date, and a use-by date (10). Foods that are highly perishable and likely to constitute an immediate danger to human health should carry a "Use by" date, after which date the "food shall be deemed unsafe" (11). In Australia and New Zealand, date labeling with a best before date or a use by date is required for most packaged foods with a shelf life of less than 2 years (12). "Best before" date indicates "the last date on which you can expect a food to retain all of its quality attributes, provided it has been stored according to any stated storage conditions and the package is unopened" (13). "Use by" date is "the last date on which the food may be eaten safely, provided it has been stored according to any stated storage conditions and the package is unopened" and the foods should not be eaten due to the health and safety reasons after the date (13).

Microorganisms play significant roles in food spoilage and loss (14). The microbiological quality and safety of fresh produce are determined by factors from pre- to post-harvest. To better characterize the microorganisms present in fresh produce during storage, a number of research projects have applied culture-dependent and culture-independent methods to investigate the microbiological quality of fresh produce during storage (15–18). For instance, Jeddi et al. (2014) enumerated the total aerobic bacteria present in 116 samples of fresh-cut vegetables, ready-to-eat salads, and wheat and mung bean sprouts before their open dates (15). The total aerobic mesophilic bacterial counts were 5.3-7.5, 5.5-7.4, 5.5-8.4 and 6.4-8.5 Log CFU/g for fresh-cut vegetable, ready-to-eat salads, and wheat and mung bean sprouts, respectively. Similarly, Berthold-Pluta et al. (2017) reported that the total aerobic mesophilic bacterial counts in lettuces, sprouts, and non-pasteurized fruits and vegetables during cold storage (below 4 °C) before the "Best Before" dates were in ranges of 5.6-7.6, 6.8-8.4, and 2.9-7.7 Log CFU/g, respectively (16).

By using 16S rRNA sequencing, Leff and Fierer (2013) profiled the bacterial communities present on 11 produce varieties purchased from grocery stores, including apples, grapes, lettuce, mushrooms, peaches, peppers, spinach, strawberries, tomatoes, alfalfa sprouts, and mung bean sprouts (19). Their results showed that the bacterial communities associated with each produce type differed remarkable from each other, but certain produce types appeared to share more similar bacterial

communities, including sprouts, spinach, lettuce, tomatoes, peppers, and strawberries. Keshri et al. (2019) determined the dynamics of the bacterial communities of fresh alfalfa and mung bean sprouts right after purchase and every week until the end of their shelf-life (20). They found that both sprout types contained spoilage-related bacteria, *Pseudomonas* and *Pantoea*, at the time of purchase. The abundance of *Pseudomonas* increased in alfalfa after three weeks of cold storage at 4 °C, while *Pantoea* dominated in mung bean sprouts after 2 weeks of storage at the same temperature. While previous studies provide important insights to the microbial population composition and potential changes during the shelf life of fresh produce, most of their sampling ended on the "Use By" dates (UBD). No information is available about bacterial populations present in fresh produce after the UBD. However, such information is very much needed for the industry as well as consumers to be better informed about what happens on the microbial quality after the UBD.

To fill the above knowledge gap, the objectives of this study were to investigate the bacterial populations of three brands of bagged chopped Romaine lettuce (RL) purchased during different seasons (early season vs. late season) and then evaluate the microbial quality of RL on the UBD and 5 days after the "Use By" dates (UBD5), using both culture-dependent and culture-independent methods.

Materials and Methods

Romaine lettuce samples

Three brands (Brands A, B, and C) of commercial chopped RL products were purchased from local grocery stores during the early (September to October) and late (March to April) harvest seasons (21). Three batches of RL for each season were purchased for the triplicate experiment. Upon arrival at the laboratory, all samples were stored at 4 °C and sampled on their labeled UBD and UBD5. Photos were also taken for these samples on each sampling day (**Figure S1**). As shown in **Figure S1**, no significant visual differences were observed between UBD and UBD5.

Culture-based analysis of the microbial quality of Romaine lettuce

At each sampling point presenting different shelf lives (UBD or UBD5), 80 grams of RL were taken from each bag and added in a 55 oz Whirl-Pak filter bag (Nasco, Fort Atkinson, WI, USA) with 320 ml of phosphate buffered saline (PBS, pH 7.4). The filter bag with RL samples were homogenized by using the Smasher™ Lab Blender (AES-Chemunix, Bruz, France) for 120 s at the fast speed (620 strokes/min). One milliliter of the homogenate was serially diluted in 9 ml of PBS in 15 ml Falcon tubes (VWR, Atlanta, GA), and four 100 µl of each dilution were plated in duplicate onto plate count agar (PCA, BD Biosciences, Sparks, MD) and anaerobic agar (AA, BD Biosciences). PCA plates were incubated at 37 °C for 48 h before colony enumeration. Anaerobic agar count medium supplemented with 2 mg/L of Methylene blue as the indicator of oxygen levels (light blue color for the presence of oxygen and yellowish color for the absence of oxygen) was used for culturing and enumerating anaerobes. Plated AA agar was incubated in Mitsubishi Gas Chemical (MGC) AnaeroPack-Jars with two MGC AnaeroPack-Anaero packs (Mitsubishi Gas Chemical Co., Tokyo, Japan) in each jar at 37 °C for 48 h before enumeration (22). The generated total aerobic plate counts (APC) and total anaerobic plate counts (AnPC) were calculated and expressed as mean ± standard deviation Log CFU/g. The limit of detection of this plating method was 1.7 Log CFU/g of lettuce.

DNA extraction and 16S rRNA gene sequencing

From each homogenate prepared above at each sampling point, 10 ml of the homogenate was transferred into a 15 ml Falcon tube (VWR, Atlanta, GA) and centrifuged at 3,000 × g for 10 min at 4 °C by using the Eppendorf Centrifuge 5810 R (Eppendorf, Hamburg, Germany) to collect the cell pellets. Each cell pellet was then washed twice by using 10 ml of PBS via centrifugation. Each washed pellet was then re-suspended with 1 mL of PBS and transferred into a 1.5 mL micro-centrifuge tube (VWR, Atlanta, GA). DNA was extracted from these pellets by using the DNeasy Powersoil kit (Qiagen, Gaithersburg, MD) following the manufacturer's instructions. A total of 36 DNA samples were extracted and stored at -80 °C for subsequent 16S rRNA sequencing process. The library construction was carried out based on the amplification of V3-V4 region of 16S rRNA gene (341F: 5'-CCTACGGGNGGCWGCAG-3' and

785R: 5'-GACTACHVGGGTATCTAATCC-3'). Every PCR amplicon was tagged with forward and reverse barcodes (7 bp) (22, 23). The sequencing was performed using the Illumina® MiSeq instrument with the MiSeq Reagent Kit v3 (Illumina, CA) to produce 2 × 300 bp paired end reads. Library preparation and the sequencing were conducted at the HudsonAlpha Genomic Service Laboratory (Huntsville, AL).

Microbiota data analysis

The 16S rRNA gene sequencing data was processed using the Quantitative Insights Into Microbial Ecology (QIIME 2 version 2021.4) pipeline (24). De-multiplexed sequences were obtained from the Illumina BaseSpace platform by assigning reads to each sample based on sample-specific barcodes. For quality control, barcodes and primers were trimmed from the raw sequences using a q2-cutadapt plugin (25), followed by the denoising process to filter out the noisy reads, remove chimeric and singleton sequences, join denoised pair-end sequences, and cluster sequences using q2-dada2 plugin (26). The principle of the DADA2 plugin in the QIIME 2 pipeline (q2-dada2 plugin) was based on the interactive quality plots of the Phred score as a function of each base in paired end sequences (**Figure S2**). The base position after which the median Phred quality score of bases dropping below 30 was applied as the cut-off point for truncation of sequences in the same length (27). With parameters "--p-trunc-len-f" (truncating length for forward reads), "--p-trunc-len-r" (truncating length for reverse reads), "--p-max-ee-f" (the number of maximum expected errors for forward reads), and "--p-max-ee-r" (the number of maximum expected errors for reverse reads) set as 260, 185, 2, and 4 respectively, approximately 62% of the sequences with median Pred quality score greater than 30 were kept for the downstream sequence analysis (**Table S1**).

In this step, the amplicon sequence variants (ASVs) were clustered based on 100% of sequence similarity (28), and the feature table of ASVs with frequency and feature data of representative ASVs was generated. A phylogenetic tree was constructed based on the feature data by using the align-to-tree-mafft-fasttree plugin (29). Diversity analyses were conducted by using the core-metrics-phylogenetic plugin

(30) with the parameter of sampling depth set as the minimum library size across all samples (80,938 reads). Alpha diversity was evaluated based on the observed features (ASVs) and Shannon index values (31). The beta diversity was evaluated by using the Jaccard distance matrix and the weighted Unifrac distance matrix (32, 33). Beta diversity was then visualized via the two-dimensional principal coordinate analysis (PCoA) plots. Taxonomic analysis was conducted at the genus level using the classify-sklearn plugin (34, 35), which employed the pre-trained Naïve Bayes classifier based on SILVA 138 small subunit rRNA (https://www.arb-silva.de/) database (34, 36).

Statistical analyses

All experiment trials were carried out in three independent replicates. Analysis of variance (ANOVA) and Tukey's Honestly Significant Difference (HSD) test (37) were applied to analyze the mean differences of bacterial amounts (APC or AnPC) obtained from RL of different brands (Brands A, B, and C) purchased during different seasons (early vs. late season), as well as at different points in their shelf lives (UBD vs. UBD5). The non-parametric Wilcoxon rank sum test and the Kruskal Wallis test were applied to test the difference between alpha diversities among groups (38). The permutational multivariate analysis of variance (PERMANOVA) (39) was used for analyzing the difference of beta diversities and the impact of shelf life, brand, and harvest season on beta diversity. Analysis of composition of microbiomes (ANCOM) (40) was applied to identify bacteria that had significantly differential abundance between groups when samples were grouped and compared by shelf life, brands or harvest seasons. The identification of these bacteria was based on the calculation of the centered log ratio (clr) F statistic and the W statistic. The clr F statistic measured the differences in effect size between groups based on the clr transformed bacterial abundance data, while the W statistic represented the number of times the log-ratio of a taxon with every other taxon being tested was identified to be significantly different across groups (41). All the above analyses were performed using R scripts (version 4.1.0). Differences were considered statistically significant when the probability (*P*) value was less than 0.05.

Results

Changes in APC and AnPC in Romaine lettuce on the UBD and UBD5

As shown in **Figure 1**, the APC of RL samples ranged from 5.71 ± 0.74 to 7.27 ± 0.20 Log CFU/g when purchased and sampled in the early season (**Figure 1A**) and ranged from 6.61 ± 0.79 to 7.89 ± 0.10 Log CFU/g in late season (**Figure 1B**). At each harvest season, no difference in APC was detected between samples plated on the UBD and UBD5. When comparing the APC among three brands, brand C (6.98 ± 0.46 Log CFU/g in early season and 7.57 ± 0.56 Log CFU/g in late season) had statistically higher APC counts than brand B (5.75 ± 0.61 Log CFU/g in early season and 6.65 ± 0.74 Log CFU/g in late season). The AnPC of RL ranged from 1.75 ± 0.08 to 6.05 ± 0.44 Log CFU/g in samples from the early season (**Figure 1C**) and 2.26 ± 0.97 to 7.32 ± 0.61 Log CFU/g in samples from the late season (**Figure 1D**). No significant difference in RL AnPC was detected between UBD and UBD5 except for brand B when samples were tested in the late season. The AnPC of brand B was 2.26 ± 0.97 Log CFU/g on UBD and was 4.34 ± 0.84 Log CFU/g on UBD5 when sampled in the late season. For samples purchased and analyzed in the early season, regardless of shelf-life, brand C had the highest AnPC among all three brands (5.98 ± 0.57 Log CFU/g), while brand B showed the lowest AnPC (1.97 ± 0.36 Log CFU/g). For samples from the late harvest season, brand B had the lowest AnPC (3.30 ± 1.40 Log CFU/g) among all three brands.

ASV identification and the diversity analyses of bacterial communities

A total of 36 DNA samples were sequenced using 16S rRNA sequencing. With the denoising process in QIIME 2 pipeline, 1,179 ASVs were identified from the total 6,051,758 sequences (**Table S1**). The minimum sequence frequency, 80,938, was chosen for rarefying the reads across all samples to avoid false positive detection (42) (**Figure S3**). The impacts of shelf life (UBD and UBD5), brand (brands A, B, and C), and harvest season (early and late season) on the alpha and beta diversity of bacterial communities were analyzed by using the QIIME 2 pipeline.

The alpha diversity of bacterial communities was evaluated by measuring observed features (ASVs) and the Shannon index values (**Figure 2**). As shown in **Figure 2A**, the numbers of the observed features (left) and the Shannon index values (right) showed no difference between UBD and UBD5 ($P > 0.05$). When comparing alpha diversity among the three brands, brand C had 131 ± 18 observed features (**Figure 2B** left), which was remarkably higher than the observed features from brands A and B ($P = 0.043$ when comparing brand A and brand C; $P = 0.0072$ when comparing brands B and C based on the Wilcoxon rank sum test). No difference of observed features was detected between brands A and C. Brand C samples also had the highest Shannon index value of 3.95 ± 0.80, compared with brand A (3.09 ± 0.82, $P = 0.01$) and brand B (2.19 ± 0.71, $P = 5 \times 10^{-5}$) (**Figure 2B** right). Brand A had a higher Shannon index value than brand B ($P = 0.012$). The Krusal-Wallis test showed that the factor of "brand" pronouncedly impacted the observed features ($P = 0.017$) and Shannon index values ($P = 1.7 \times 10^{-4}$) of RL bacterial communities. When focusing on the impact of harvest seasons, early-harvest-season RL had higher observed features (129 ± 17) than late-season RL (104 ± 21) ($P = 6.3 \times 10^{-4}$, **Figure 2C** left). However, such impact was not detected when focusing on the Shannon index values (**Figure 2C** right).

Changes of the alpha diversity of bacterial communities in each brand of RL between UBD and UBD5 were also studied (**Figure S4**). Brand A RL had a significant increase of observed feature numbers from UBD to UBD5, while no difference of Shannon index values was observed between RL on the UBD and UBD5. Brand B RL had no changes in neither observed feature number nor Shannon index value from UBD to UBD5. For brand C RL, Shannon index value had remarkable increase from UBD to UBD5, while no difference of observed feature numbers was detected between UBD and UBD5.

The beta diversity of RL bacterial community was calculated based on the Jaccard distance matrix and the weighted Unifrac distance matrix. PCoA was used to visualize the beta diversity of bacterial communities in RL samples grouped by shelf life, brands, or seasons. The PERMANOVA test was employed to statistically evaluate the impacts of shelf life, brands, or seasons on the beta diversity. In general, PCoA plots based on the Jaccard distance matrix (**Figure 3** left) showed that the principal coordinates 1 and 2 explained 14.88% and 11.25% of variances respectively, and PCoA plots based on

the weighted Unifrac distance (**Figure 3** right) showed that the PCo1 and PCo2 explained 84.76% and 6.76% of variances respectively, suggesting that the weighted Unifrac distance matrix captured more commensal bacteria variances than the Jaccard distance matrix, as illustrated on the two-dimensional PCoA plots. No obvious visual separation was observed between UBD and UBD5 samples. This result was further confirmed by using the PERMANOVA test with *P* values of 0.767 and 0.077 for the Jaccard distance-based comparison and the weighted Unifrac-based comparison respectively. The factor of "brand," on the other hand, more significantly impacted the beta diversity of RL bacterial community, as the PCoA plots based on the Jaccard distance matrix and the weighted Unifrac distance both showed clear visual separation among three brands. The PERMANOVA test confirmed the significant impact of the factor "brand" on the beta diversity of RL bacterial community by generating *P* values of 0.001 for both the Jaccard distance-based comparison and the weighted Unifrac distance-based comparison. When analyzing the impact of harvest season on beta diversity, analyses based on the Jaccard distance were different from the results obtained based on the weighted Unifrac. The PCoA plot presented clear visual separation between early and late season samples when using the Jaccard distance-based comparison (*P* value = 0.01 in PEMANOVA test), while no separation was observed from the PCoA plot based on the weighted Unifrac distance matrix (*P* value = 0.55 in PEMANOVA test).

Taxonomic analysis of commensal bacterial in RL

The taxonomic analysis of commensal bacteria identified 128 genera across all RL samples based on the SILVA database. The top five genera were *Pseudomonas* (with relative abundances ranging from 9.95 to 94.73%), *Weissella* (0-42.42%), *Serratia* (0-52.35%), *Leuconostoc* (0-31.56%), and *Lactococcus* (0-24.30%) (**Figure 4**). When focusing on shelf life, the top five genera identified from the UBD samples were *Pseudomonas* (16.47-92.72%), *Serratia* (0-52.35%), *Weissella* (0-42.42%), *Pantoea* (0.17-21.33%), and *Lactococcus* (0-24.30%), while the top five genera identified on UBD5 were *Pseudomonas* (9.95-94.73%), *Pantoea* (0-31.56%), *Leuconostoc* (0.23-28.50%), *Serratia* (0-25.66%), and *Weissella* (0.10-39.62%).

When focusing on different brands, the top five genera identified in brand A samples were *Pseudomonas* (15.32-91.62%), *Leuconostoc* (0.01-31.56%), *Serratia* (0-52.35%), *Pantoea* (0.60-19.48%), and *Lactococcus* (0-2-.42%); the top five genera identified in brand B were *Pseudomonas* (25.46-94.73%), *Pantoea* (0.17-28.50%), *Janthinobacterium* (0.47-24.01%), *Xanthomonas* (0-30.26%), and *Serratia* (0.05-16.77%); while the top five genera identified in brand C were *Pseudomonas* (9.95-56.41%), *Weissella* (9.21-42.42%), *Serratia* (0.28-46.06%), *Lactococcus* (0.94-24.30%), and *Leuconostoc* (1.14-22.52%). When grouping the samples by harvest seasons, the top five genera identified in the early season were *Pseudomonas* (9.95-91.62%), *Pantoea* (1.64-28.50%), *Serratia* (0-46.06%), *Weissella* (0-42.42%), and *Leuconostoc* (0-31.56%), while the top five genera identified in the late season were *Pseudomonas* (15.33-94.73%), *Weissella* (0-34.69%), *Leuconostoc* (0-26.41%), *Serratia* (0.05-52.35%), and *Lactococcus* (0-24.30%).

Differential abundance analysis of commensal bacteria in RL based on shelf life, brand, and harvest season

To identify the critical features (indicators) associated with samples when grouped by shelf life, brand, and harvest season, the ANCOM analysis was employed for the differential abundance analysis of commensal bacteria in RL. Results of ANCOM analysis are shown in **Figure 5A**. No bacteria were identified to be significantly different in abundance between UBD and UBD5, while six genera were identified to have significantly differential abundance among three brands; these were *Weissella* (W statistics = 144, clr F statistics = 67.7), *Leuconostoc* (W statistics = 144, clr F statistics = 35.3), *Lactococcus* (W statistics = 144, clr F statistics = 33.9), *Pseudarcobacter* (W statistics = 137, clr F statistics = 19.1), *Yersinia* (W statistics = 134, clr F statistics = 26.3), and *Massilia* (W statistics = 132, clr F statistics = 20.3). *Leuconostoc* (RA 0.011-31.56%) had the highest RA in brand A, while *Massilia* (0-2.12%) had the highest abundance in brand B. *Weissella* (9.21-42.42%), *Lactococcus* (RA 0.94-24.30%), *Pseudarcobacter* (0-4.51%), and *Yersinia* (0-0.50%) had the highest RA in brand C. *Lactococcus*, *Leuconostoc*, *Weissella*, *Pseudarcobacter*, and *Yersinia* had the lowest RA in brand B (**Figure 5B**). When

grouping RL samples by harvest seasons, only one unannotated genus (UAG) of Oxalobacteraceae (W statistics = 137, clr F statistics = 3.1) had significantly higher RA in the samples from the late harvest season (0-5.24%) than in the samples from the early harvest season (0-0.083%) (**Figure 5B**). Based on the ANCOM test, no bacteria in the RL of brands A, B, and C individually could be identified as indicators for between UBD and UBD5 samples (**Figure S4**).

Discussion

This study applied culture-dependent and culture-independent methods to investigate the commensal bacteria present in bagged RL on their "Use By" dates and 5 days after the "Use By" dates. The combination of both methods allows us to better characterize the bacterial populations and discover potential changes associated with abundance, diversity, and composition of different bacterial groups. The application of 16S rRNA gene sequencing is advantageous for microbiota profiling and analyzing bacterial community dynamics, as it is culture-independent and relatively unbiased compared to traditional culture methods that rely highly on selective or differentiated media (43).

During the quality control step of analyzing the sequence data in the QIIME 2 pipeline, parameters "--p-trunc-len-f," "--p-trunc-len-r," "--p-max-ee-f," and "--p-max-ee-r" were set at 260, 185, 2, and 4 respectively in the q2-dada2 plugin command. Estaki et al., 2020 recommended setting the read length at which the median Phred quality score began to drop below 30 or 20 if the entire read quality was too low (27). With these settings, the forward reads and reverse reads were truncated at 260 bp and 185 bp as their median Phred quality score of the next base started to drop below 30 as shown in the interactive quality plots (**Figure S2**). In addition, the forward reads with more than two erroneous bases and reverse reads with more than four erroneous bases were discarded, as the reverse reads had lower Phred quality scores than the forward reads (**Figure S2**). Similar observations have been reported by D'Amore et al. (2016) (44) and Gerasimidis et al. (2016) (45).

As shown in **Figure 1**, culture-dependent results showed no difference in APC or AnPC of RL between UBD and UBD5, except the AnPC of late-season brand B. The population density of culturable

commensal bacteria in the RL on the UBD and UBD5 remained at the same level for aerobic bacteria and anaerobic bacteria, indicating that the short 5-day storage after the UBD had no significant impact on the total culturable bacterial abundance. It can be explained by that bacterial populations on UBD and UBD5 had reached stationary levels under this storage condition. The APC of bagged RL ranged from 5.71 ± 0.74 to 7.89 ± 0.10 Log CFU/g and the AnPC ranged from 1.75 ± 0.08 to 7.32 ± 0.61 Log CFU/g. RL samples obtained from the late season had higher APC ($P = 3.08 \times 10^{-5}$) and AnPC ($P = 1.06 \times 10^{-4}$) than RL samples obtained from the early season. Culturable bacteria counts obtained from this study were consistent with results from previous studies. Aycicek et al. (2006) reported that the outer leaves of Romaine lettuce and iceberg-lettuce samples harbored 3.3-7.4 Log CFU/g of aerobic bacteria (46). Korir et al. (2016) reported that the mean abundance of total aerobic bacteria was 7.76 Log CFU/g for lettuce (47). Similarly, Liao and Wang (2021) reported that the total aerobic bacteria in Spring Mix salad samples were at the level of approximately 6.6 Log CFU/g (22). The differences observed between different seasons were also reported by Williams et al. (2013) (21), in which the abundance of aerobic bacteria in phyllosphere was found to be lower on RL planted in the early season (June) (approximately 3.8-5.5 Log CFU/g) than on RL planted in the late season (August and October) (approximately 5.0-6.2 Log CFU/g).

The diversity analysis of the 16S rRNA gene sequencing detected no significant difference of alpha diversity in RL between UBD and UBD5, indicating that neither the richness nor the evenness of the RL bacterial communities had changed. Beta diversity based on the Jaccard distance matrix and the weighted Unifrac distance matrix also showed no difference between UBD and UBD5, as visualized on the PCoA plots. The results were further confirmed by the PERMANOVA test. These results indicate that the additional 5-day storage after the UBD has no impact on the diversity of RL bacterial communities. Similar results have been reported by our previous Spring Mix study, in which no significant difference of the beta diversity was observed among Spring Mix samples collected at different storage times during 15 days of cold storage at 4 ^{o}C (22).

The top six identified genera, *Pseudomonas*, *Serratia*, *Weissella*, *Pantoea*, *Lactococcus*, and *Leuconostoc*, were the same across RL samples, but in different orders. The exception was that the top six

genera in brand B samples included *Janthinobacterium*, *Xanthomonas*, and *Flavobacterium* instead of *Weissella*, *Lactococcus*, and *Leuconostoc*. Among the top six genera, *Pseudomonas* consistently dominated the RL bacterial communities with RA ranging from 9.95-94.73%. The previous study carried out by Gu et al. (2018) reported that the *Pseudomonas* species consistently dominated microbial communities across all spinach samples during week-long storage at 4 ^{o}C (up to 34.20%) (48). Keshri et al. (2019) reported that *Pseudomonas* dominated the bacterial communities of alfalfa sprouts throughout three weeks of cold storage at 4 ^{o}C, with an RA of greater than 60% (20). The *Pseudomonas* genus contains a group of species associated with fresh produce spoilage, such as *Pseudomonas fluorescents*, *P. marginalis*, *P. viridiflava*, and *P. chloroaphis*, and is able to outcompete other bacteria in food matrixes (49, 50). Another potential spoilage genus is *Pantoea* (20, 51). The study from Bae et al. (2014) and Keshri et al. (2019) reported the *Pantoea* RA was up to 50.40% in mung bean sprouts after 2 weeks of storage at 4 ^{o}C (20, 51). A previous study also reported that *Pantoea* were highly abundant in pepper (11.5%), spinach (32.5%), and sprouts (57.5%) analyzed right after purchase (19).

Weissella, *Lactococcus*, and *Leuconostoc* have been reported as psychrotrophic lactic acid bacteria (LAB) potentially linked to the spoilage of fresh produce (20, 52). Pothakos et al. (2014) reported that *Leuconostoc* spp. was the most dominant population in ready-to-eat salads at the end of their shelf-life (7 days) at 4 ^{o}C, which caused the early spoilage of vegetables before the end of shelf-life (53). Kesheri et al. (2019) found that the abundance of *Leuconostoc* increased after 2 weeks of cold storage at 4 ^{o}C in mung bean sprouts, the bacterium is considered to play a role in the spoilage of sprouts (20). In addition, greater numbers of *Lactococcus* and *Weissella* were reported in Spring Mix salad when stored at 4 ^{o}C for 15 days (22). *Serratia* is a plant-associated genus that is a non-pathogenic symbiont that has also been reported to be present at high levels in organic green leafy lettuce with up to 66% RA (54). *Yersinia* (0-0.50%) and *Bacillus* (0-3.0%), two genera containing human pathogens, e.g. *Yersinia enterocolitica* (55) and *Bacillus cereus* (56), were also identified in RL. No significant change in their relative abundance was detected from UBD to UBD5 based on the ANCOM test.

In this study, ANCOM analysis was applied to analyze and identify indicators (bacterial genera) with differential RA in RL samples when grouped by different shelf lives (UBD vs. UBD5), brands (brands A, B, and C), and harvest seasons (early vs. late). For RL on the UBD and UBD5, the ANCOM test showed no indicator with significantly differential RA between them, suggesting that the composition of bacteria in the RL had no change 5 days past the UBD. When comparing the microbial populations among three brands, the ANCOM test identified that five genera had significantly different RAs among the three brands. Among them, *Leuconostoc*, *Weissella*, *and Lactococcus* contain species associated with food spoilage, such as *Leuconostoc citreum*, *Leuconostoc mesenteroides*, and *Weissella confuse* (52). For the RL from early and late seasons, ANCOM identified only one genus, Oxalobacteraceae_UAG (up to 0.80%), was identified as higher RA in the RL from late season compared to that from early season. Other than this genus in a low RA, RL from the two seasons had no difference in composition of bacterial communities, which was also reflected in the above Shannon diversity of the two-seasonal RL. Williams et al. (2013) illustrated that the season-of-planting factor more strongly impacted RL bacterial communities after four weeks to eight weeks of planting in the field. Specifically, *Pantoea*, *Pseudomonas*, *Erwinia*, and members of Enterobacteriaceae were identified as the most abundant genera in late season plants (21). *Leuconostoc*, *Lactococcus*, *Bacillus*, and *Exiguobacterium* were predominant in early season plants. By comparison, the weaker impact of the seasonal factor on the diversity and composition of RL bacterial communities in our study could be explained by postharvest processing (e.g. washing, packaging, and cold storage), which might alter the RL bacterial communities and mitigate the seasonal impact.

In the United States, except for infant formula, the U.S. Food and Drug Administration (FDA) does not require manufactures to display specific open date labeling on their food products (8). Although a few poultry, meat, and egg products under the jurisdiction of USDA need data documentation, no strict guidelines for using the food date terminology on food products have been set by the USDA (5, 57). The current open-date labels used in food markets are not well regulated, as the “Use By” date, “Best By,” and “Sell By” are interchangeably applied by manufacturers (8). This ambiguous food date labeling has

fostered confusion about food product safety and quality among consumers, which subsequently causes unnecessary levels of food waste (5, 58). Evidence showed that consumers waste food products when they are near the open dates perceived for food quality reasons or food safety reasons (5, 8, 9, 59).

In January 2017, the Grocery Manufacturers Association (GMA) and Food Marketing Institute (FMI) recommended the use of two introductory phrases for food date labeling, including "Best If Used By" and "Use By". They recommended that "Best If Used By" is applied to "indicate to the consumer that, after a specified date, the product may not taste or perform as expected but is safe to be used or consumed" and "Use By" is used to "applies to perishable products that should be consumed by date on the package and discarded after that date" (60). The U.S. FDA strongly supported the industry's voluntary use of "Best If Use By" introductory phrase for quality-based food date labels but was not addressing the use of a "Use By" label for safety reasons (61). Other than the United States, most developed countries required food date labeling for most food products. The EU mandates the use of food date labeling for food products. Australia and New Zealand and the EU clearly regulate that the "Use By" dates related to food safety and "Best Before" related to food quality (5).

To mitigate food waste, stakeholders need to well clarify and regulate the food date labeling, and customers need more information about fresh produce quality and about safety after the open date. The U.S. FDA (2010) reported that the expected shelf-life for bagged fresh-cut leafy green is approximately 12-16 days after production at 4 °C and the sensory quality of fresh leafy green can last at least a week after the open date labels (62). However, limited study on microbial food quality after "open dates" has been reported. As fresh fruits and vegetables comprise the largest part of food waste in the food system (63), this study presented bacterial RL quality on and after the UBD based on culture-dependent and culture-independent methods. The results showed that RL bacterial communities had no change in abundance, diversity, or composition from the UBD to UBD5. RL on the UBD5 had no increase of microbial quality or safety risk compared to RL on the UBD.

Conclusion

In summary, this study determined there was no difference in bacterial abundance, diversity, and composition of bacterial communities in three brands of RL on the UBD5 compared with RL on the UBD, suggesting that the microbial quality of RL remained the same at two storage time points. Factors of "brand" and "harvesting season" played more significant roles in shaping the bacterial communities present in RL samples. The study provides first-hand information about the microbial quality of fresh produce on the UBD and a few days after the UBD.

Data availability statement

The raw 16S rRNA gene sequencing data and metadata from this study have been submitted to GenBank Sequence Read Archive under the BioProject ID PRJNA792031 in NCBI.

Acknowledgements

The authors appreciate the financial support provided by the Alabama Agricultural Experiment Station, the USDA HATCH Multistate Research Project S1077 "Enhancing Microbial Food Safety by Risk Analysis", and the startup funding from the University of California Davis.

References

1. Mijares V, Alcivar J, Palacios C. 2021. Food waste and tts association with diet quality of foods purchased in south florida. Nutrients 13.
2. Buzby JC, Farah-Wells H, Hyman J. 2014. The estimated amount, value, and calories of postharvest food losses at the retail and consumer levels in the United States. SSRN Journal.
3. Environmental Protection Agency. 2015. United States 2030 Food Loss and Waste Reduction Goal. https://www.epa.gov/sustainable-management-food/united-states-2030-food-loss-and-waste-reduction-goal [Accessed February 20, 2022].
4. United States Department of Agriculture. 2015. Food Loss and Waste 2030 Champions. https://www.usda.gov/foodlossandwaste/champions [Accessed February 20, 2022].
5. Newsome R, Balestrini CG, Baum MD, Corby J, Fisher W, Goodburn K, Labuza TP, Prince G, Thesmar HS, Yiannas F. 2014. Applications and perceptions of date labeling of food. Comp Rev Food Sci Food Safety 13:745–769.
6. Buzby J, Bentley J, Padera B, Ammon C, Campuzano J. 2015. Estimated fresh produce shrink and food loss in U.S. supermarkets. Agric 5:626–648.
7. Leib EB, Rice C, Neff R, Spiker M, Schklair A. 2016. Consumer perceptions of date labels: National survey. Saf 23:19
8. Wilson NLW, Rickard BJ, Saputo R, Ho S-T. 2017. Food waste: The role of date labels, package size, and product category. Food Qual Prefer 55:35–44.
9. Tsiros M, Heilman CM. 2005. The effect of expiration dates and perceived risk on purchasing behavior in grocery store perishable categories. J Mark 69:114–129.
10. European Commission. 2011. Regulation (EU) No 1169/2011 of the European Parliament and of the Council of 25 October 2011 on the provision of food information to consumers. http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:304:0018:0063:EN:PDF [Accessed February 20, 2022].

11. EFSA Panel on Biological Hazards (BIOHAZ), Koutsoumanis K, Allende A, Alvarez-Ordóñez A, Bolton D, Bover-Cid S, Chemaly M, Davies R, De Cesare A, Herman L, Nauta M, Peixe L, Ru G, Simmons M, Skandamis P, Suffredini E, Jacxsens L, Skjerdal T, Da Silva Felicio MT, Hempen M, Messens W, Lindqvist R. 2020. Guidance on date marking and related food information: part 1 (date marking). EFSA Journal 18.
12. Australian Government. 2012. Australia New Zealand food standards code - standard 1.2.5 - date marking of packaged food. Available online: http://www.comlaw.gov.au/Details/F2012C00762 [Accessed February 20, 2022].
13. Food Standards Australia New Zealand. 2013. Date marking user guide to standard 1.2.5: date marking of packaged food. Food Standards Australia, New Zealand. http://www.foodstandards.gov.au/code/userguide/Documents/Guide%20to%20StandardA [Accessed February 20, 2022].
14. Lorenzo JM, Munekata PE, Dominguez R, Pateiro M, Saraiva JA, Franco D. 2018. Main groups of microorganisms of relevance for food safety and stability, p. 53–107. *In* Innovative technologies for food preservation. Elsevier.
15. Jeddi MZ, Yunesian M, Gorji ME, Noori N, Pourmand MR, Khaniki GRJ. 2014. Microbial evaluation of fresh, minimally processed vegetables and bagged sprouts from chain supermarkets. J Health Popul Nutr 32:391–399.
16. Berthold-Pluta A, Garbowska M, Stefańska I, Pluta A. 2017. Microbiological quality of selected ready-to-eat leaf vegetables, sprouts and non-pasteurized fresh fruit-vegetable juices including the presence of Cronobacter spp. Food Microbiol 65:221–230.
17. Cruz MRG da, Leite YJB de S, Marques J de L, Pavelquesi SLS, Oliveira LR de A, Silva ICR da, Orsi DC. 2019. Microbiological quality of minimally processed vegetables commercialized in Brasilia, DF, Brazil. fst 39:498–503.
18. Arienzo A, Murgia L, Fraudentali I, Gallo V, Angelini R, Antonini G. 2020. Microbiological quality of ready-to-eat leafy green salads during shelf-life and home-refrigeration. Foods 9.

19. Leff JW, Fierer N. 2013. Bacterial communities associated with the surfaces of fresh fruits and vegetables. PLoS One 8:e59310.

20. Keshri J, Krouptiski Y, Abu-Fani L, Achmon Y, Bauer TS, Zarka O, Maler I, Pinto R, Sela Saldinger S. 2019. Dynamics of bacterial communities in alfalfa and mung bean sprouts during refrigerated conditions. Food Microbiol 84:103261.

21. 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.

22. Liao C, Wang L. 2021. Evaluation of the bacterial populations present in Spring Mix salad and their impact on the behavior of *Escherichia coli* O157:H7. Food Control 107865.

23. Sinclair L, Osman OA, Bertilsson S, Eiler A. 2015. Microbial community composition and diversity via 16S rRNA gene amplicons: evaluating the illumina platform. PLoS One 10:e0116955.

24. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu Y-X, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R,

Caporaso JG. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37:852–857.

25. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet j 17:10.
26. 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.
27. Estaki M, Jiang L, Bokulich NA, McDonald D, González A, Kosciolek T, Martino C, Zhu Q, Birmingham A, Vázquez-Baeza Y, Dillon MR, Bolyen E, Caporaso JG, Knight R. 2020. QIIME 2 enables comprehensive end-to-end analysis of diverse microbiome Ddta and comparative studies with publicly available data. Curr Protoc Bioinformatics 70:e100.
28. 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.
29. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780.
30. Caporaso JG, 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. Nat Methods 7:335–336.
31. Kõiv V, Arbo K, Maiväli Ü, Kisand V, Roosaare M, Remm M, Tenson T. 2019. Endophytic bacterial communities in peels and pulp of five root vegetables. PLoS One 14:e0210542.
32. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. 2011. UniFrac: an effective distance metric for microbial community comparison. ISME J 5:169–172.
33. Edgar RC. 2017. Accuracy of microbial community diversity estimated by closed- and open-reference OTUs. PeerJ 5:e3889.

34. Bokulich NA, Dillon MR, Bolyen E, Kaehler BD, Huttley GA, Caporaso JG. 2018. q2-sample-classifier: machine-learning tools for microbiome classification and regression. J Open Res Softw 3.

35. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA, Gregory Caporaso J. 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 6:90.

36. 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.

37. Abdi H, Williams LJ. 2010. Tukey's honestly significant difference (HSD) test. Encyclopedia of research design 3:1-5.

38. Xia Y, Sun J. 2017. Hypothesis testing and statistical analysis of microbiome. Genes Dis 4:138–148.

39. Anderson MJ. 2014. Permutational multivariate analysis of variance (PERMANOVA), p. 1–15. *In* Balakrishnan, N, Colton, T, Everitt, B, Piegorsch, W, Ruggeri, F, Teugels, JL (ed), Wiley statsref: statistics reference online. John Wiley & Sons, Ltd, Chichester, UK.

40. Mandal S, Van Treuren W, White RA, Eggesbø M, Knight R, Peddada SD. 2015. Analysis of composition of microbiomes: a novel method for studying microbial composition. Microb Ecol Health Dis 26:27663.

41. Stevens AJ, Purcell RV, Darling KA, Eggleston MJF, Kennedy MA, Rucklidge JJ. 2019. Human gut microbiome changes during a 10-week Randomised Control Trial for micronutrient supplementation in children with attention deficit hyperactivity disorder. Sci Rep 9:10128.

42. Saary P, Forslund K, Bork P, Hildebrand F. 2017. RTK: efficient rarefaction analysis of large datasets. Bioinformatics 33:2594–2595.

43. Grim CJ, Daquigan N, Lusk Pfefer TS, Ottesen AR, White JR, Jarvis KG. 2017. High-resolution microbiome profiling for detection and tracking of *Salmonella enterica*. Front Microbiol 8:1587.

44. D'Amore R, Ijaz UZ, Schirmer M, Kenny JG, Gregory R, Darby AC, Shakya M, Podar M, Quince C, Hall N. 2016. A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. BMC Genomics 17:55.

45. Gerasimidis K, Bertz M, Quince C, Brunner K, Bruce A, Combet E, Calus S, Loman N, Ijaz UZ. 2016. The effect of DNA extraction methodology on gut microbiota research applications. BMC Res Notes 9:365.

46. Aycicek H, Oguz U, Karci K. 2006. Determination of total aerobic and indicator bacteria on some raw eaten vegetables from wholesalers in Ankara, Turkey. Int J Hyg Environ Health 209:197–201.

47. Korir RC, Parveen S, Hashem F, Bowers J. 2016. Microbiological quality of fresh produce obtained from retail stores on the Eastern Shore of Maryland, United States of America. Food Microbiol 56:29–34.

48. Gu G, Ottesen A, Bolten S, Ramachandran P, Reed E, Rideout S, Luo Y, Patel J, Brown E, Nou X. 2018. Shifts in spinach microbial communities after chlorine washing and storage at compliant and abusive temperatures. Food Microbiol 73:73–84.

49. Hibbing ME, Fuqua C, Parsek MR, Peterson SB. 2010. Bacterial competition: surviving and thriving in the microbial jungle. Nat Rev Microbiol 8:15–25.

50. Langsrud S, Moen B, Møretrø T, Løype M, Heir E. 2016. Microbial dynamics in mixed culture biofilms of bacteria surviving sanitation of conveyor belts in salmon-processing plants. J Appl Microbiol 120:366–378.

51. Bae Y-M, Zheng L, Hyun J-E, Jung K-S, Heu S, Lee S-Y. 2014. Growth characteristics and biofilm formation of various spoilage bacteria isolated from fresh produce. J Food Sci 79:M2072–80.

52. Säde E, Lassila E, Björkroth J. 2016. Lactic acid bacteria in dried vegetables and spices. Food Microbiol 53:110–114.

53. Pothakos V, Snauwaert C, De Vos P, Huys G, Devlieghere F. 2014. Monitoring psychrotrophic lactic acid bacteria contamination in a ready-to-eat vegetable salad production environment. Int J Food Microbiol 185:7–16.

54. 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.

55. Cristiano D, Peruzy MF, Aponte M, Mancusi A, Proroga YTR, Capuano F, Murru N. 2021. Comparison of droplet digital PCR vs real-time PCR for *Yersinia enterocolitica* detection in vegetables. Int J Food Microbiol 354:109321.

56. Chon J-W, Seo K-H. 2021. Evaluation of Ceftazidime as an antibiotic supplement in mannitol-yolk-polymyxin B agar used for enumeration of *Bacillus cereus* in ready-to-eat vegetables. J Food Prot 84:1698–1703.

57. Leib EB, Gunders D, Ferro J, Nielsen A. 2013. The dating game: How confusing food date labels lead to food waste in America.

58. Wansink B, Wright AO. 2006. best if used by …?" how freshness dating influences food acceptance. J Food Sci 71:S354–S357.

59. Theotokis A, Pramatari K, Tsiros M. 2012. Effects of expiration date-based pricing on brand image perceptions. J Retail 88:72–87.

60. Food Marketing Institute (FMI). 2019. FDA Support of the Product Code Date Label Initiative. https://www.fmi.org/blog/view/fmi-blog/2019/06/10/fda-support-of-the-product-code-date-label-initiative [Accessed February 20, 2022].

61. Food and Drug Administration. 2019. Confused by Date Labels on Packaged Foods? https://www.fda.gov/consumers/consumer-updates/confused-date-labels-packaged-foods [Accessed February 20, 2022].

62. Food and Drug Administration. 2010. Program Information Manual Retail Food Protection: Recommendations for the Temperature Control of Cut Leafy Greens during Storage and Display in Retail Food Establishments. https://www.fda.gov/media/78982/download [Accessed February 20, 2022].63. Augustin MA, Sanguansri L, Fox EM, Cobiac L, Cole MB. 2020. Recovery of wasted fruit and vegetables for improving sustainable diets. Trends Food Sci Technol 95:75–85.

Table S1. Statistical summary of actual sequence variants (ASVs) before and after denoising process during the QIIME 2 analysis.

	Number of samples	Total frequency	Minimum frequency	Median frequency	Maximum frequency
Sequences before denoising	36	9,713,248	172,148	278,255	375,736
Sequences after denoising	36	6,051,758	80,938	172,073	250,847

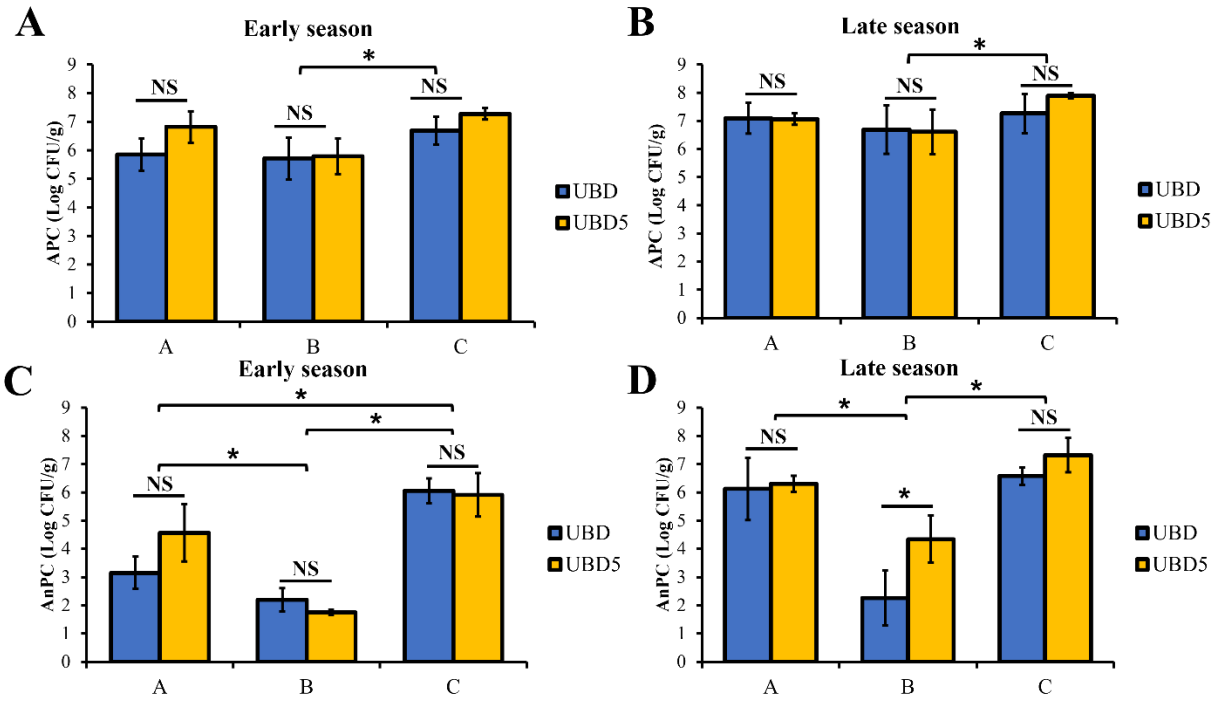

Figure 1. Total aerobic (APC, A and B) and anaerobic bacterial counts (AnPC, C and D) of three brands of Romaine lettuce purchased in the early and late seasons and analyzed on their "Use By" dates (UBD) and 5 days after the "Use By" dates (UBD5) stored at 4 °C. * represents significant differences of APC or AnPC between RL samples of three brands or between UBD and UBD5 ($P < 0.05$).

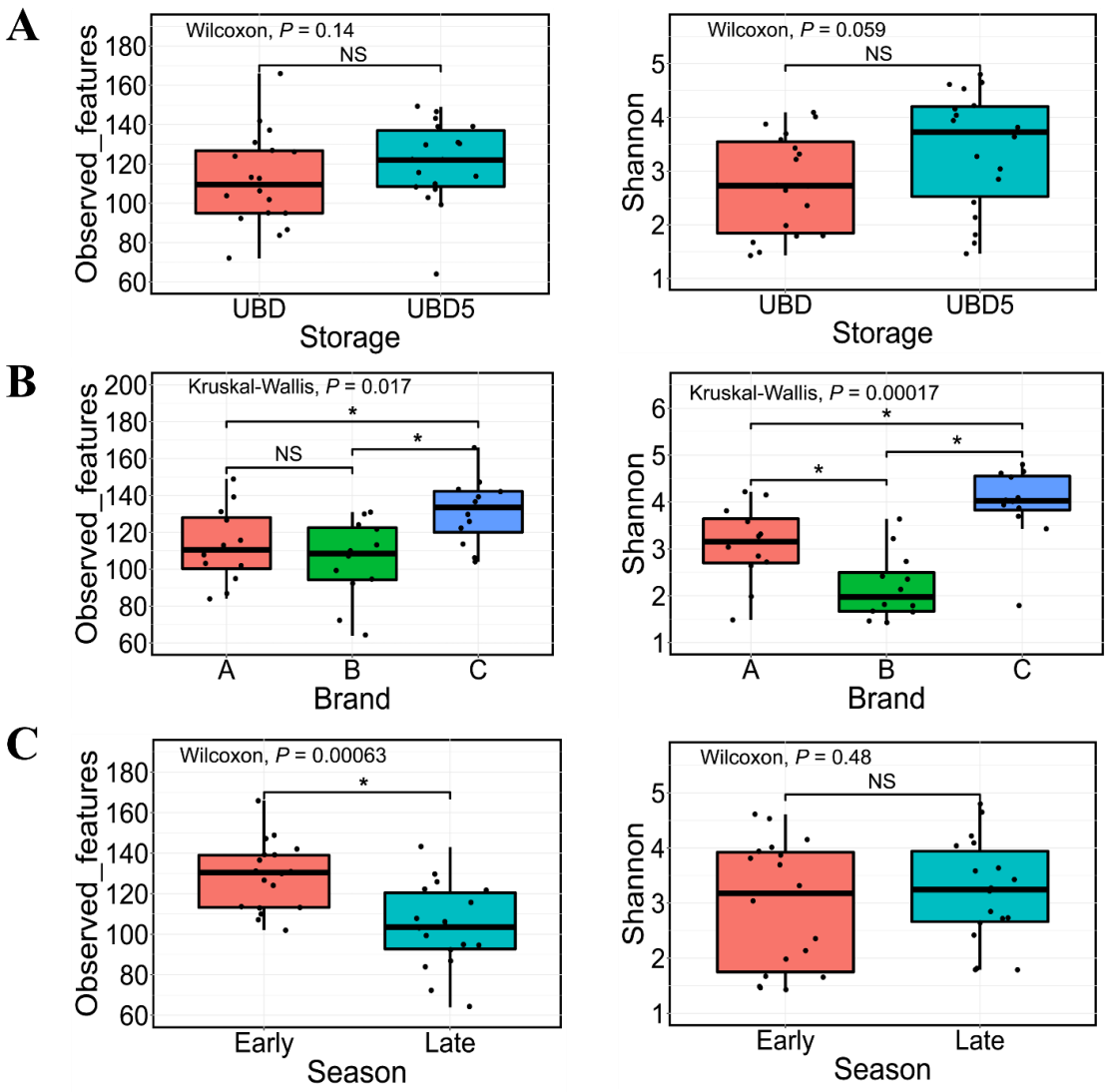

Figure 2. Comparison of the alpha diversity based on the observed features (left) or the Shannon index values (right) of microbial populations present in RL between UBD and UBD5 (A), among three brands (B), and between two seasons (C). The Wilcoxon rank test was used for pairwise comparison while the Kruskal-Wallis test was applied for comparisons among three groups. The * represents significant differences between groups ($P < 0.05$). NS means no significant difference.

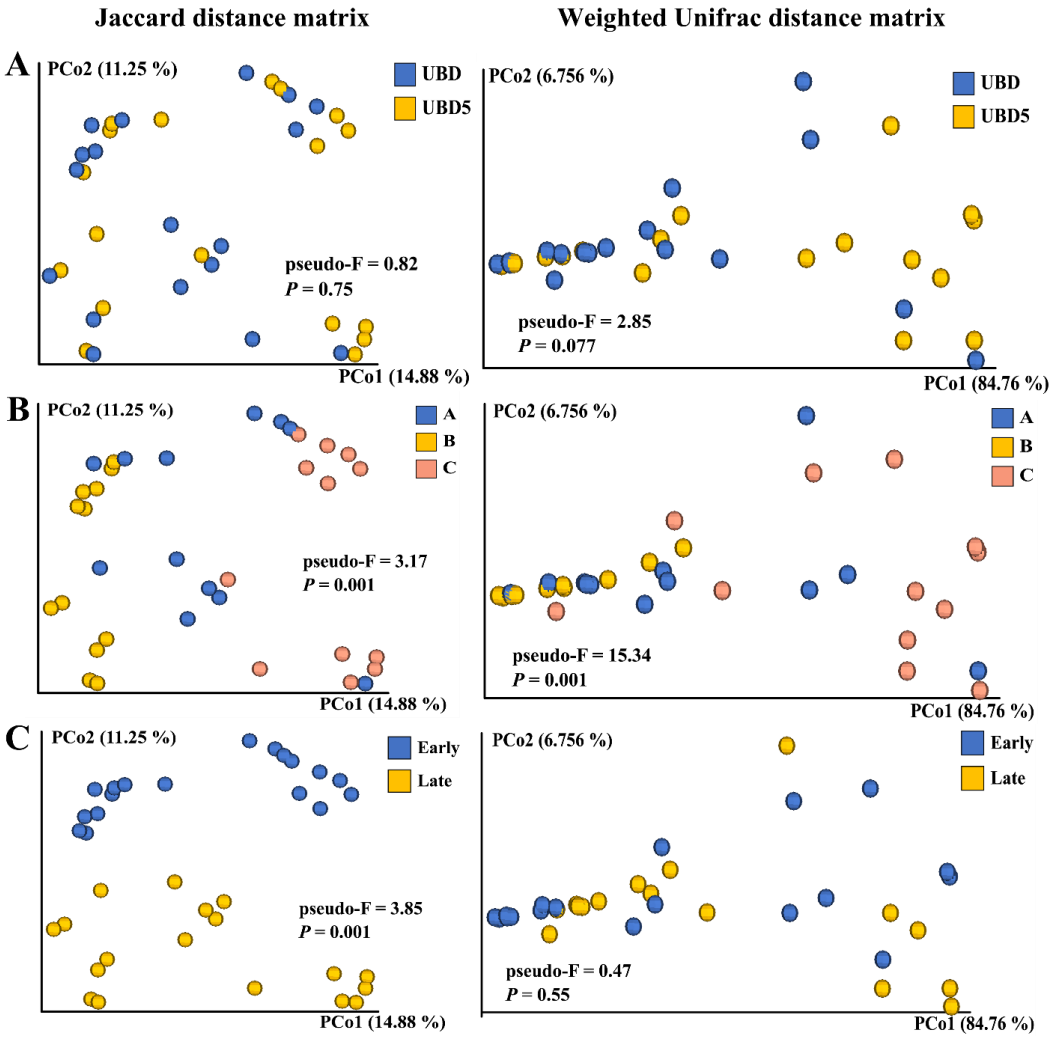

Figure 3. Comparison of the beta diversity based on the Jaccard distance matrix (left) and the weight Unifrac distance matrix (right) of microbial communities present in RL between UBD and UBD5 (A), among three brands (B), or between different seasons (C). The PERMANOVA test was applied to analyze the differences in each group when samples were grouped by shelf life, brand, or seasons. "pseudo-F" values are measures of the effect size, indicating the differences within each group of RL. *P* values of less than 0.05 indicate that the factor of shelf life, brand, or season significantly impacts the beta diversity.

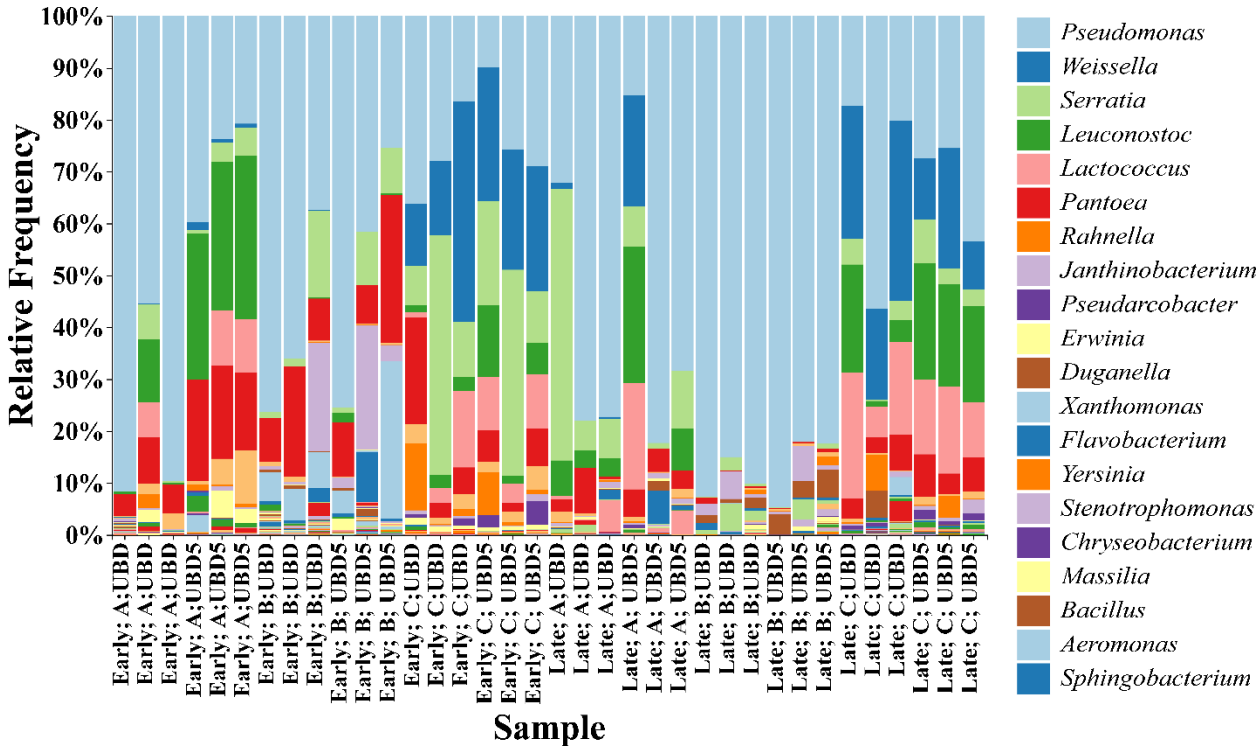

Figure 4. Relative abundance of the top 20 bacteria genera identified from all three brands of RL purchased from two harvest seasons at UBD and UBD5. "UBD" and "UBD5" mean "Use By" dates and five days after the UBD. "A", "B", and "C" stand for three brands and "Early" and "Late" represent the early and late harvest season.

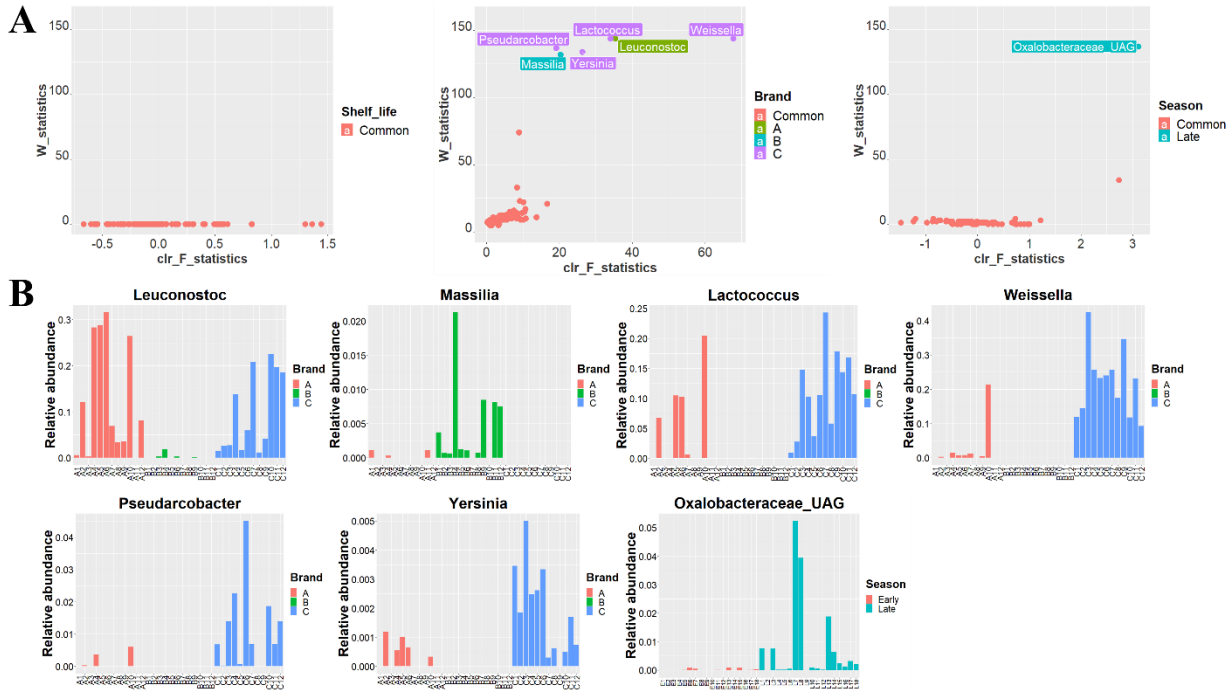

Figure 5. (A) Volcano plots of bacteria, at the genus level, identified to have different abundances when RL are grouped and compared based on their shelf-life (left), brand (middle), and harvest season (right) with the ANCOM test. The abundance differences of bacteria were evaluated by the clr F statistic and W statistic. Bacteria with no different abundance when RL samples were grouped and compared based on shelf life, brand, and season factors are marked in red. Bacteria with significantly higher abundances in the brands A, B, and C RL are marked in green, blue, and purple colors respectively when comparing the RL of the three brands. Bacteria with significantly higher abundances in the late season RL are marked in blue when comparing RL of different seasons. (B) Barplots of the relative abundance changes of *Leuconostoc*, *Massilia*, *Weissella*, *Lactococcus*, *Pseudarcobacter*, *Yersinia* identified as the indicators for the three brands of RL and Oxalobacteraceae identified as the indicator for the late season RL.

Figure S1. Photos of three brands (A, B, and C) of RL on UBD (left), UBD5 (middle), and UBD7 (right) stored at 4 °C. The UBD, UBD5, and UBD7 are abbreviations for the "Use By" date, 5 days after UBD, and 7 days after UBD, respectively.

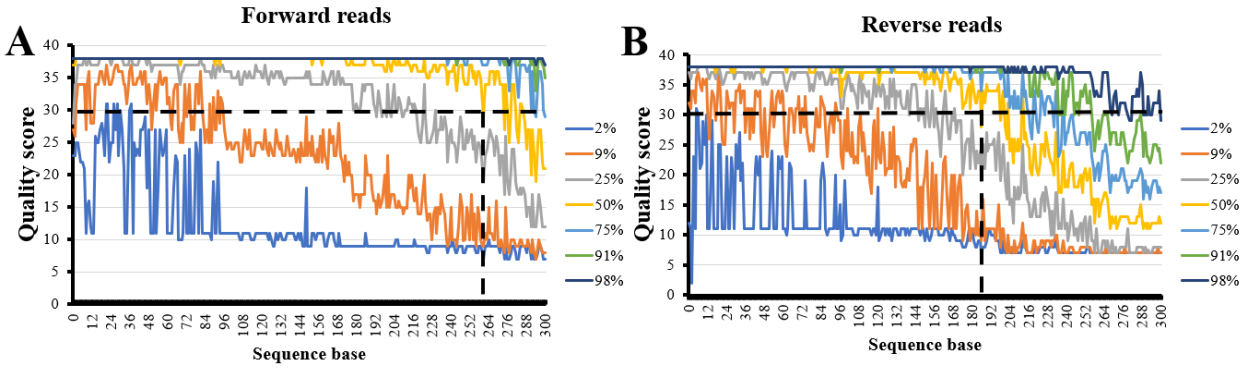

Figure S2. Interactive quality plots of forward and reverse reads with 300 bases each. (A) Interactive quality plots of the forward read with 300 bases for seven percentiles. (B) Interactive quality plots of the reverse read with 300 bases for seven percentiles, including 2%, 9%, 25%, 50%, 75%, 91%, and 98%. The horizontal dash line represents the threshold of Phred quality score as 30 (Q30). The vertical dash line stands for the intersection points between Q30 dash line and the 50% interactive quality curve.

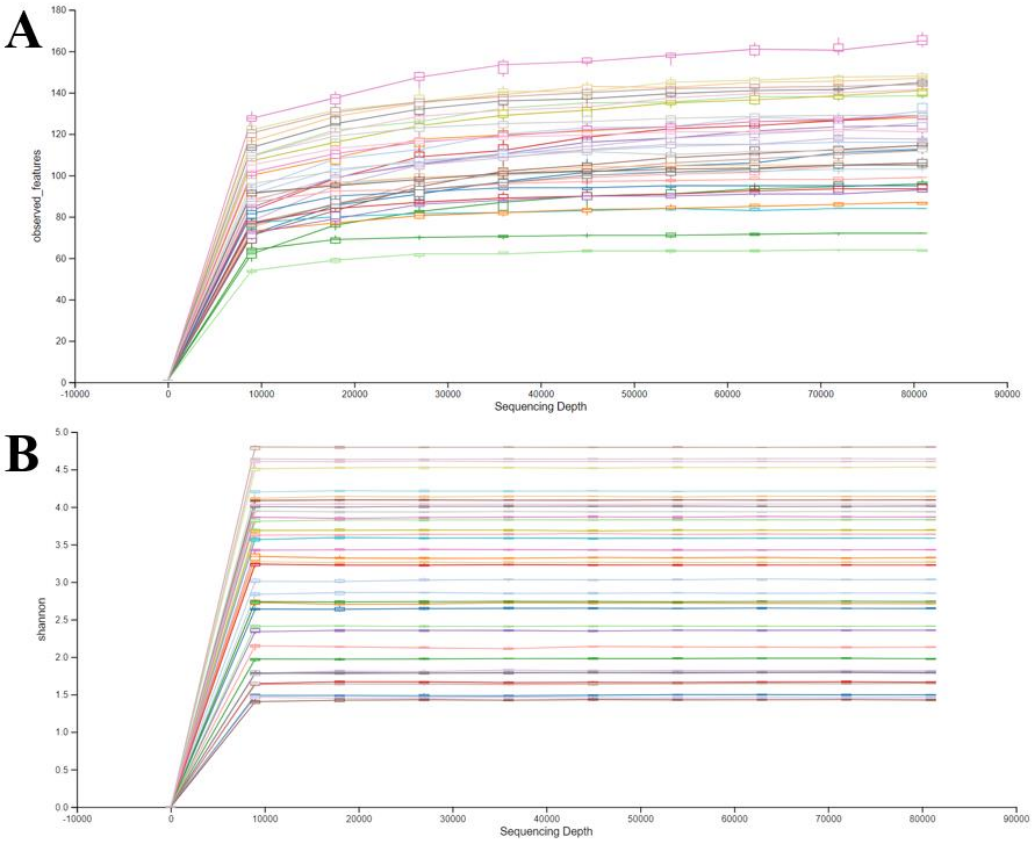

Figure S3. Rarefication curves of alpha diversity as a function of sequencing depth (80,938 reads) of 36 samples. (A) Rarefication curves based on the observed features. (B) Rarefication curves based on the Shannon index.

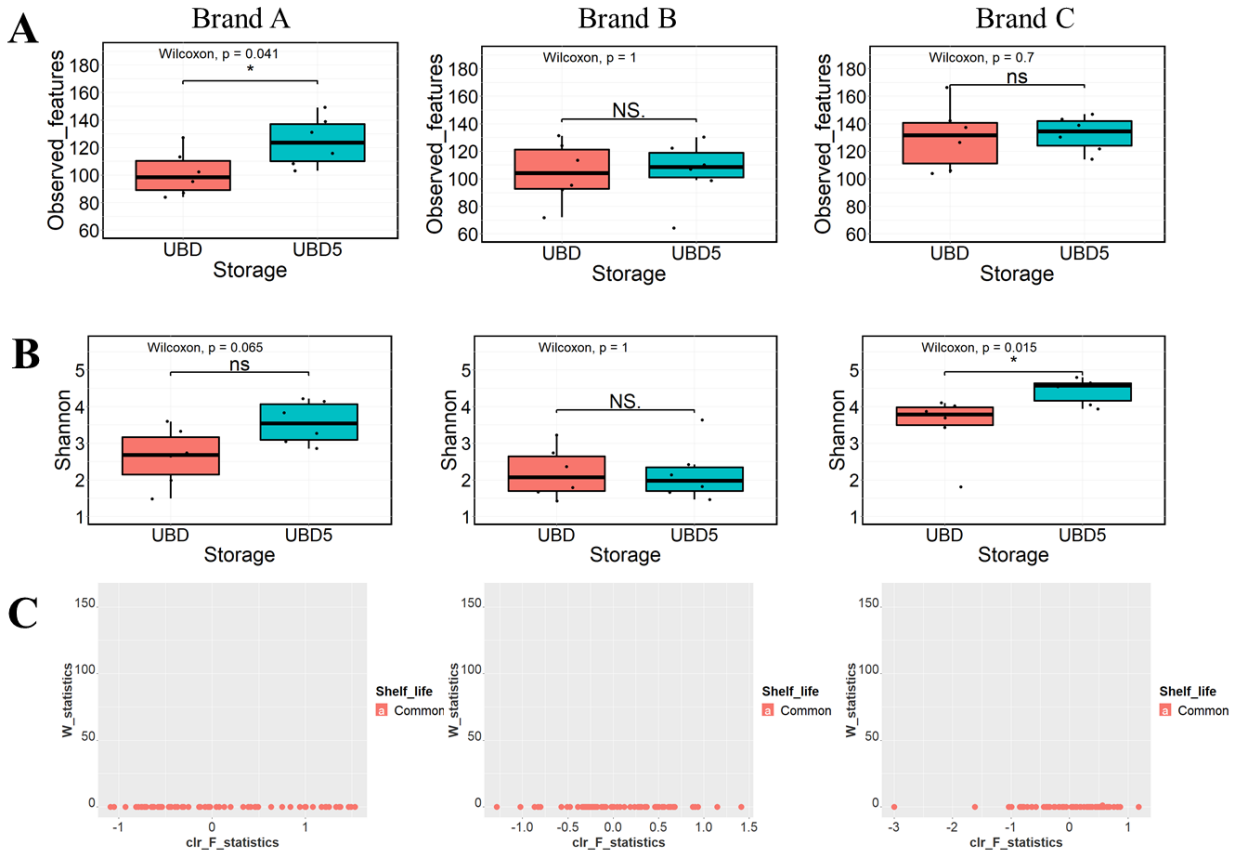

Figure S4. The Alpha diversity variation based on the observed feature numbers and Shannon index values of bacterial communities in three brands (A, B, and C) on UBD and UBD5. (A) boxplots of observed feature numbers; (B) boxplots of Shannon index values; (C) Volcano plot of bacteria indicators for RL on UBD and UBD5 identified by the ANCOM test. UBD and UBD5 mean the "Use By" date and five days after the UBD. The Wilcoxon rank test was used for the pairwise comparison. The Kruskal-Wallis test was applied for the overall comparisons among three groups. "*" stands for $P < 0.05$. "NS" means no significance was observed between two groups. "Common" represents bacterium with no differential abundance between UBD and UBD5.

Chapter 3. Impact of the bacterial communities present in chopped Romaine lettuce on *Listeria monocytogenes*

Running title: Interactions between *Listeria monocytogenes* and Romaine lettuce microbiota

Chao Liao [a] and Luxin Wang [a*]

[a] Department of Food Science and Technology, University of California Davis, Davis, CA 95616, USA

*Corresponding author: Luxin Wang, Associate Professor, University of California Davis.

Email address: lxwang@ucdavis.edu. Institutional address: Department of Food Science and Technology, University of California Davis, Davis, CA 95616

Abstract

This study evaluated the impact of native microbial populations present in Romaine lettuce (RL) on the behavior of *Listeria monocytogenes* (LM) during 5 days of storage at 4 °C. RL (harvested either from the early season or the late season) was inoculated with low (~ 3 Log CFU/g) or high levels of LM (~ 6 Log CFU/g). The levels of LM did not change in early-season RL but decreased in late-season RL. Through the analysis of composition of microbiomes (ANCOM) test and the random forests (RF) feature selection method, results showed the RF identified *Pseudomonas*, *Chryseobacterium*, *Duganella*, and *Listeria* as indicators of high-level LM contamination of RL, while the ANCOM identified *Listeria* for RL inoculated with higher LM levels. For the RL inoculated with a low level of LM, the RF identified *Serratia*, *Pedobacter*, *Janthinobacterium*, *Flavobacterium*, and *Weissella* as indicators of low-level LM contamination, while the ANCOM identified no indicator. Through the spot-on-lawn method, 32 out of 47 isolates from RL were confirmed as antagonistic bacteria for LM. Three antagonistic species were then identified based on whole genome sequencing. They were *Bacillus filamentosus*, *Carnobacterium maltaromaticum*, *Lactococcus lactis*, and. The distribution of the three antagonisticspecies in RL was quantified by qPCR, which showed all three species were present across all the RL. *B. filamentosus* was significantly higher in late-season RL than in early-season RL, while *C. maltaromaticum* and *Lc. lactis* showed no differences. Additional research is needed to better confirm the antagonist function of *B. filamentosus* and to better identify its anti-LM activity.

Keywords: Romaine lettuce, *Listeria monoyctogenes*, bacterial communities, competitive exclusion, analysis of composition of microbiomes, random forests feature selection.

Introduction

In the United States, approximately 46% of foodborne outbreaks are related to fresh produce, accounting for about $39 billion in annual economic losses (1, 2). Fifteen multistate outbreaks due to *Listeria monocytogenes* (LM) infection were traced back to fresh vegetables and fresh fruits, causing a total of 352 illnesses, 326 hospitalizations, and 65 deaths from 2010 to 2022 (3). This fact has caused increased public concern about fresh produce safety. Among the varieties of fresh produce reported as contaminated, Romaine lettuce (RL) was the produce third most frequently involved.

Foodborne pathogen contamination of fresh produce may occur at any stage from farm to fork (4, 5). Although washing has been generally applied as an essential step during fresh produce processing, studies have shown that current washing strategies do not adequately reduce surface bacteria; water is treated with antimicrobials to maintain water sanitation and to prevent cross-contamination (6, 7). Gu et al. (2018), Nou and Luo (2010), and Singh et al. (2018) have also shown that different washing strategies will generate different impacts on the native microbial populations or foodborne pathogens on fresh produce (6, 8, 9). Gu et al. (2018) showed that chlorine wash significantly reduced populations of numerous bacterial species, with f-*Moraxellaceae* sp., *Paracoccus* sp., and *Mycoplana* sp. being the top three most reduced bacteria. Chlorine wash had a limited effect on *Stenotrophomonas* spp. and *Erwinia* spp., as it has been reported that the bacterial populations returned to their original levels after washing (6). In the same Gu et al. (2018) study, for example, the total mesophilic aerobic plate count increased by 1.77 to 3.24 Log after storage at 4, 10, and 15 °C for 7 days (6). Likewise, among the different bacterial species present in spinach, *Stenotrophomonas* sp. and *Erwinia* sp., showed higher tolerance to chlorine treatment. Similarly, Nou and Luo (2010) showed that washing fresh-cut RL leaves in chlorinated water (70 ppm) for 60 s resulted in reducing aerobic plate counts from 6.3 to 5.2 Log CFU/g and the *Escherichia coli* O157:H7 population from 3.6 to 2.2 Log CFU/g (8). A study conducted by Singh et al. (2018) illustrated that washing RL with peracetic acid at 100 ppm for 5 min achieved a reduction by 2.2 Log CFU/g of *E. coli* O157:H7 (with an initial load at 6.9 Log CFU/g) and by 2.4 Log CFU/g of LM (with an initial load at 7.7 Log CFU/g) (9).

Understanding the composition and diversity of bacterial communities present in washed RL and their interactions with LM during postharvest not only better informs the washing efficacy of different washing strategies, but also can facilitate the development of new biocontrol strategies to mitigate LM contamination. A few studies have investigated the structures of bacterial communities fresh produce inoculated with *E. coli* O157:H7 (4, 10, 11), LM (4, 12), and *Salmonella* Infantis (13). Each of these studies have shown an association between pathogen contamination and its impact on native microbial populations. For example, Zhang et al. (2018) illustrated that irrigation with water containing *S.* Infantis decreased the diversity of the lettuce rhizosphere (13). Similarly, in our previous study (10), contamination by *E. coli* O157:H7 shifted the microbial composition of Spring mix salad. With the differential abundance analysis (LEfSe) with Pearson correlation analysis, *Brochothrix*, *Dyadobacter*, *Shewanella*, *Yersinia*, and *Methylotenera* were found to be positively correlated with the growth of *E. coli* O157:H7 during storage at 4 °C for 5 days. On the other hand, *Fusobacterium*, *Carnobacterium*, *Aeromonas*, and *Planomicrobium* were negatively correlated with behaviors of *E. coli* O157:H7 during the storage at 4 °C for 5 days.

Bacterial antagonism, also known as "bacterial interference" or "competitive exclusion," is an approach employing nonpathogenic bacteria to mitigate colonization and inhibit pathogenic bacteria in environments ranging from animal guts to plants or food products (14, 15). A few studies have reported that antagonistic bacteria exhibit consistent antimicrobial activity in fresh produce against various foodborne pathogens, including *E. coli* O157:H7, LM, and *S. enterica* (16–18). Lopez-Velasco et al. (2012) isolated antagonistic bacteria that inhibited *E. coli* O157:H7 on fresh spinach, including *Bacillus* spp., *Bacillus pumilus*, *Erwinia perscina*, *Paenibacillus polymyxa*, *Pantoea agglomerans*, and *Pseudomonas* spp. (17). They found that application of ~ 4 Log CFU/g of each antagonistic strain could reduce *E. coli* O157:H7 on spinach leaves stored at 25 °C after 24 h by 0.3-1.3 Log CFU/g, compared to the control (17). Russo et al. (2015) applied ~ 8 Log CFU/g of *Lactobacillus plantarum* B2 and *Lactobacillus fermentum* PBCC11.5 to reduce ~ 4 Log CFU/g of LM by 3 and 2 Log CFU/g, respectively, in fresh-cut cantaloupe at 4 °C after 9 days compared to the control (18). In addition, a study conducted

by Kim et al. (2022) showed *Erwinia persicina* EUS78 isolated from vegetable seeds at the levels of ~ 5-8 Log CFU/g inhibited *S. enterica* ATCC4931r by up to 5 Log CFU/g compared to the control on alfalfa sprouts at 25 °C after 7 days (16).

The objectives of this study were to evaluate the impacts of native microbial populations in RL on the behavior of LM, identify bacteria with their populations significantly influenced by LM contamination and behaviors, and confirm the levels of potential antagonistic species by using quantitative PCR.

Materials and methods

Commercial RL samples

A total of 48 bags of commercial chopped RL (284 g/bag) were purchased from local grocery stores in two harvest seasons, including the early season (September to October) and the late season (March to April), which were denoted based on the early planting season (June) and the late planting season (August and October) addressed in Williams et al. (2013) (11, 19). Three batches of RL were purchased for triplicate experiments. Once delivered to the laboratory, the samples were stored at 4 °C and sampled on their "Use by" dates (UBD) and 5 days after the UBD (UBD5).

Analysis of microbial communities present in RL

For each season, 24 bags of purchased RL samples were randomly assigned to the control (non-inoculation) group and the inoculation group, in which samples were artificially inoculated with LM. Two inoculation levels were applied, with the low inoculation level being ~ 3 Log CFU/g and the high inoculation level being ~ 6 Log CFU/g. Samples were sampled at two sampling times (UBD and UBD5) in triplicates. At each sampling time, 80 grams of chopped RL were added into 320 ml of phosphate-buffered saline (PBS, pH 7.4) in a 55 oz Whirl-Pak filter bag (Nasco, Fort Atkinson, WI, USA) and homogenized by using the Smasher™ Lab Blender (AES-Chemunix, Bruz, France) for 120 s at the fast speed (620 strokes/min). RL homogenates (100 µl) and corresponding serial 10-fold dilutions were plated onto plate count agar (PCA, BD Biosciences, Sparks, MD, USA) and anaerobic agar (AA, BD

Biosciences, Sparks, MD) to obtain the total aerobic plate counts (APC) and anaerobic plate counts (AnPC), respectively. PCA plates were incubated at 37 °C for 48 h and AA plates were incubated in Mitsubishi Gas Chemical (MGC) AnaeroPack-Jars with two MGC AnaeroPack®-Anaero sachets in each of them (Mitsubishi Gas Chemical Co., Tokyo, Japan) at 37 °C for 48 h (10). After incubation, colonies were enumerated, and results were reported as mean ± standard deviation with common logarithm transformation in the unit of CFU/g. The limit of detection of the direct plating method was 1.7 Log CFU/g of bacteria in RL. In addition, 10 ml of each homogenate was transferred into a 15 ml Falcon tube and stored at -80 °C for subsequent DNA extraction.

Culture preparation and LM inoculation

Three strains of LM were used in this study: LJH1422 (associated with raw dice of yellow onions), LJH512 (related to raw cabbage outbreak), and LJH552 (associated with tomatoes). These strains were kindly provided by Dr. Linda J. Harris at the University of California Davis. To activate the culture, frozen culture stocks were retrieved from the -80°C freezer and thawed on ice followed by streaking a loop of thawed cultures onto Trypticase soy agar (TSA) and incubating at 37 °C for 24 h. After incubation, a single colony was picked from each plate and transferred into 10 ml of fresh Brain Heart Infusion (BHI) broth and incubated at 37 °C for 24 h. A loop of fresh broth culture was transferred again into 10 ml of fresh BHI and incubated at 37 °C overnight (18 h) before use. Each fresh overnight culture was washed with 10 ml of PBS (pH = 7.4) twice via centrifugation. The OD600 of each washed culture was adjusted to the value of 1.60 by using PBS solution via an SPECTRONIC™ 200 Spectrophotometer (Thermo Fisher Scientific, Vacaville, CA, USA), corresponding to ~ 9 Log CFU/ml of LM.

To prepare the three-strain cocktail, 1 ml of each washed LM culture was transferred and mixed in a 15 ml Falcon tube. The final concentration of the cocktail was 9.5 ± 0.6 Log CFU/ml, which was used as the high-level inoculum. The cocktail was diluted 1,000-fold before being used for preparing the samples inoculated with low LM levels. RL samples were inoculated on the UBD (Day 0). To do so, 1 ml of LM cocktail was injected into each bag of commercial RL with a 1 ml BD syringe with a 26 G × 3/8”

detachable needle (Becton Dickinson and Company, Franklin Lakes, NJ). The injection hole was immediately sealed with tape to maintain the same package atmosphere. After inoculation, the inoculated salad bag was vigorously shaken for 1 min to distribute the inoculum evenly (20). Inoculated and non-inoculated RL samples were both stored at 4 ℃ for 5 days.

Analysis of the LM behaviors in inoculated RL

At each sampling time, 80 grams of LM-inoculated RL from a bag were added into a 1.6 L Whirl-Pak filter bag (Nasco, Fort Atkinson, WI, USA) with 320 ml of PBS (pH 7.4) and homogenized by using the Smasher™ Lab Blender (AES-Chemunix, Bruz, France) for 120 s at the fast speed (620 strokes/min). One hundred microliters of LM-inoculated RL homogenates and serial 10-fold dilutions were plated on PCA, AA, and modified Oxford agar (MOX, BD Biosciences, Sparks, MD, USA) for enumerating aerobic bacteria, anaerobic bacteria, and LM cells. The PCA and MOX plates were directly incubated at 37 °C for 48 h, and AA plates were placed in MGC AnaeroPack-Jars with two MGC AnaeroPack®-Anaero sachets for each jar at 37 °C for 48 h. An aliquot of each sample homogenate (10 ml) was stored in a -80 °C freezer for subsequent DNA extraction.

DNA extraction and 16S rRNA gene sequencing

Frozen homogenates of inoculated or non-inoculated RL were thawed on ice for 30 min and washed twice with 10 ml of PBS via centrifugation at 3,000 g for 10 min by using the Eppendorf Centrifuge 5810 R (Eppendorf, Hamburg, Germany). The washed pellets were then suspended in 1 mL of PBS and transferred into 1.5 mL micro-centrifuge tubes (VWR International, Visalia, CA, USA). DNA from 24 non-inoculated and 24 LM-inoculated (12 low inoculation level and 12 high inoculation level) RL samples from two seasons was extracted by using the DNeasy PowerSoil Pro kit (Qiagen, Germantown, MD, USA) following the manufacturer's instruction. The DNA samples were stored in a -80 °C freezer for the following library preparation for the 16S rRNA gene sequencing with the Illumina® MiSeq (Illumina, San Diego, CA, USA).

The library preparation was carried out by using the QIAseq 16S/ITS Region Panel (Qiagen, Germantown, MD, USA), including QIAseq 16S region PCR reaction and QIAseq 16S region panel sample index PCR reaction. Specifically, the QIAseq 16S PCR reaction was carried out based on the amplification of the V3-V4 region of the 16S rRNA gene (341F: 5'-CCTACGGGNGGCWGCAG-3' and 785R: 5'-GACTACHVGGGTATCTAATCC-3'). Every PCR amplicon was tagged with forward and reverse indexes using the QIAseq 16S 96-Index I (p5-RS2-ID# and p7-FS2-ID#) (Qiagen, Germantown, MD, USA). The sequencing was performed using the Illumina® MiSeq instrument with the MiSeq Reagent Kit v3 (Illumina, San Diego, CA, USA) to produce 2 × 300 bp paired end reads. The sequencing library preparation was carried out in our laboratory, and the sequencing process was conducted at the Genome Center at UC Davis.

Microbiota data processing and analysis

After 16S rRNA gene sequencing, demultiplexed sequencing data was obtained from the Genome Center at UC Davis and processed using the Quantitative Insights Into Microbial Ecology 2 (QIIME 2 version 2022.02) (21). The raw demultiplexed sequences were imported into QIIME 2 as qza files by using the "qiime tools import" command. Indexes and primers were removed from raw sequences by using "qiime cutadapt trim-paired" command (22). To control the sequence quality, the Divisive Amplicon Denoising Algorithm 2 (DADA2) plugin was applied to reduce the noisy reads, remove chimeric and singleton sequences, join denoised pair-end sequences, dereplicate sequences, and cluster sequences by using the "qiime dada2 denoise-paired" command (23). To ensure that the median Phred quality score of the sequence base is greater than 30, sequences were truncated at the cut-off points where the quality scores were dropping below 30 for truncating sequences (24). The cut-off points were identified from the interactive quality plots of the Phred score as a function of the sequence base in the paired-end sequences (**Figure S1**). To do that, flags "--p-trunc-len-f" (truncating length for forward reads) and "--p-trunc-len-r" (truncating length for reverse reads) were set as 0 and 204, respectively in the

"qiime dada2 denoise-paired" command. After the denoising step, an average of 66.53% of sequences across all samples were left (**Table S1**).

After the denoising step, amplicon sequence variants (ASVs) based on 100% of sequence similarity were clustered (25), which included a feature table of ASVs with frequency and feature data of representative ASVs. The diversity analysis was conducted by using the core-metrics-phylogenetic plugin (26) with the parameter of sampling depth set as the minimum library size (31,329 reads) to rarefy the total frequency of each sample (**Figure S2**). The alpha diversity was evaluated based on the observed features and the Shannon index. Taxonomic analysis was conducted at the genus level by using the classify-sklearn plugin (27) with a Naïve Bayes-based classifier pre-trained with the SILVA database (https://www.arb-silva.de/) (28).

Antagonistic bacteria isolation and confirmation

Since the LM decrease occurred in the late-season RL on UBD and UBD5, the double-layer agar method (29) was applied to isolate colonies from 48 frozen RL homogenates that showed antagonistic activity against LM. Briefly, for each RL homogenate, its serial 10-fold dilutions were plated on TSA plates. Five replicate media were plated for each dilution. After incubation at 37 °C for 24 h, plates with less than 50 colonies were poured with 6 ml of 2% soft TSA containing 6 Log CFU/ml of the three-strain LM cocktail (29). After incubation at 37 °C for 24 h, LM lawns formed on the TSA plates. The colonies from RL with LM inhibition zones on the LM lawns, considered potential antagonistic colonies, were picked up by using sterilized toothpicks to break the second layer of soft agar. If the number of potential antagonistic colonies on a plate was less than three, all these colonies were picked up. If the number was at three or more, three potential antagonistic colonies, each with different morphologies if possible, were selected. These colonies were streaked onto TSA plates and incubated at 37 °C for 24 h. A single isolated colony from each plate was then transferred into 10 ml of TSB and incubated at 37 °C for 24 h.

The spot-on-lawn method was applied to confirm the antagonistic bacteria with antagonistic activity against LM (17). Specifically, a single colony of each potential antagonist was picked and spotted

on a new TSA plate with a sterile toothpick, and 6 ml of 2% soft TSA containing 6 Log CFU/ml of the three-strain LM cocktail was poured onto the spotted agar. After incubation at 37 ℃ for 24 h, the diameter of the LM inhibition zone of the potential antagonist on the spotted TSA was measured. In the end, 32 antagonistic bacteria were confirmed with obvious antagonistic property against LM, with the diameter of the LM inhibition zones larger than 5 mm (17).

Sanger sequencing and whole genome sequencing

Thirty-two frozen antagonistic cultures were activated following the same protocol for LM cultures as mentioned above, using TSB instead of BHI broth. The activated cultures were centrifuged at 3,000 × g for 10 min to obtain antagonistic pellets. Afterwards, each antagonistic pellet was washed twice with 1 ml of PBS and processed for DNA extraction by using the QIAamp DNA Mini Kit (Qiagen, Germantown, MD, USA). The concentration and the purity of the genomic DNA samples were measured by using the NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, Vacaville, CA, USA). The extracted DNA was stored in a -80 °C freezer for the following qPCR, Sanger sequencing based on the 16S rRNA gene, and the whole genome sequencing (WGS).

In each PCR reaction (50 µl), bacterial universal primers, 27F (5′-AGATTTGATCMTGGCTCAG-3′) and 1492R (5′-TACGGYTACCTTGTTACGAC TT-3′), were applied to amplify the 16S rRNA gene (30). Each PCR reaction contained 25 µl of Dream Taq green PCR 2× master mix (Thermo Fisher Scientific, Vacaville, CA, USA), 2 µl of genomic DNA template, 5 µl of primer 27F (25 pmol), 5 µl of primer 1492R (25 pmol), and 13 µl of nuclease-free water. The PCR program used for amplification of the 16S rRNA gene included an initial denaturation at 95 ℃ for 5 min, 35 cycles of 0.5 min 95 °C denaturation, 1 min 50 ℃ annealing, 1 min 72 ℃ extension, and a final extension at 72 ℃ for 5 min. The PCR products were purified using the QIAquick PCR Purification Kit (Qiagen, Germantown, MD, USA). Purified PCR products were checked by gel electrophoresis with 1% (w/v) agarose gel. The concentration and the purity of PCR products were measured using NanoDrop 2000c spectrophotometers (Thermo Fisher Scientific, Vacaville, CA, USA). The PCR products were then

sequenced using the Applied Biosystems 3730 Capillary Electrophoresis Genetic Analyzer (Thermo Fisher Scientific, Vacaville, CA, USA) with the ABI BigDye® Terminator v3.1 Cycle Sequencing Kit and Gel Company Better Buffer (Thermo Fisher Scientific, Vacaville, CA, USA). The Sanger sequencing was conducted at the UCDNA sequencing facility and analyzed using Chromas 2.6.6 and NCBI Blast. To further identify and characterize antagonistic isolates at the species level, DNA samples of 32 confirmed antagonists were sequenced using Illumina NovaSeq 6000 (Illumina, San Diego, CA, USA) at the Genome center at UC Davis. The WGS data analysis was carried out using Neisseria pipeline 1.2 (31) on the Galaxy platform (https://galaxy.sciensano.be/).

Quantitative PCR for the confirmation of antagonistic culture concentrations in RL

The sequencing-identified antagonistic bacteria were validated and quantified using qPCR. QuantStudio™ 3 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). Primers used for *Bacillus filamentosus, Carnobacterium maltaromaticum Lactococcus lactis* and were listed in **Table 1**. In each 20 µl of reaction system, 10 µl of Maxima SYBR Green/ROX qPCR Master Mix (2×) (Thermo Fisher Scientific, Vacaville, CA, USA), 2 µl of each primer, 2 µl of the DNA template, and 4 µl of nuclease-free H_2O were added. The qPCR was performed by using the following conditions: An initial 10 min of 95 °C for denaturation and 40 cycles of 15s of 95 °C and 1 min of 60 °C. The melting curve was performed by gradually heating the PCR products from 60-95 °C and continuously measuring the fluorescence to verify primer specificity. The qPCR for each DNA sample was carried out in technique duplicates and biological triplicates.

Statistical analyses

The differences in APC, AnPC, and LM counts between RL samples on two sampling times (UBD and UBD5) or from two harvest seasons (early- and late-season) or in contamination status (non-contaminated and contaminated) in two inoculation levels (low inoculation level and high inoculation level) were analyzed by using analysis of variance (ANOVA) followed by Tukey's Honestly Significant

Difference (HSD) (32). To investigate the differences in alpha diversity between two groups of samples or more than two groups, the non-parametric Wilcoxon rank sum test and the Kruskal Wallis test were applied, respectively (33). For the differential abundance analysis, the analysis of composition of microbiomes (ANCOM) method (34) was employed to identify critical bacteria/indicators associated with RL groups based on storage, season, LM contamination, or inoculation level. In the ANCOM test, the W statistic and the centered log ratio (clr) F statistic were calculated to identify the indicators. The W statistic is the number of times that the log-ratio of a taxon with every other taxon being tested was identified to be significantly different between groups (19, 35). The clr F statistic evaluated the effect size of bacteria based on the differences in the clr transformed bacterial abundance data between groups (19, 35). In addition, the random forests (RF) feature selection method (36) was applied to rank and select features (ASVs or bacteria) that positively contributed to the distinction between RL groups based on the calculated mean decrease in accuracy (MDA). The R package "randomForest" (37) was used to execute the RF classification and feature selection (36). All the above statistical analyses were carried out with R programming (version 4.2.0). Differences between groups were considered significant when *P* was less than 0.05.

Results

Culturable aerobic and anaerobic bacteria populations in non-contaminated and LM-contaminated RL

Based on culture-dependent methods, changes in populations of culturable aerobic bacteria (APC) and anaerobic bacteria (AnPC) in RL stored at 4 °C from UBD to UBD5 were evaluated using PCA and AA. **Figures 1A** and **1B** show that no significant variations in APC and AnPC RL were observed between different bags of RL and RL on UBD and UBD5. Significant differences in APC ($P = 8.8 \times 10^{-6}$) and AnPC ($P = 1.0 \times 10^{-7}$) existed between early-season RL and late-season RL. The early-season RL contained 7.3 ± 0.2 Log CFU/g of APC, which was higher than the late-season RL APC (6.8 ± 0.2 Log

CFU/g). In contrast, the early-season RL contained a lower concentration of AnPC (3.2 ± 0.3 Log CFU/g) than the late-season RL (4.2 ± 0.2 Log CFU/g).

In LM-inoculated RL samples, no difference in APC and AnPC between RL on UBD and UBD5 was observed, except the late-season RL inoculated with a high level of LM in which a significant decrease of AnPC between RL on UBD and UBD5 was observed ($P = 0.03$). Compared to the differences seen in APC between non-inoculated RL from two seasons, early-season LM-inoculated RL contained the same level of APC as the late-season LM-inoculation RL (7.2 ± 0.3 Log CFU/g versus 7.1 ± 0.2 Log CFU/g) (**Figure 1C**). However, the late-season RL had significantly higher concentrations of AnPC than early-season RL with 4.9 ± 0.1 Log CFU/g vs. 4.1 ± 0.2 Log CFU/g ($P = 2.2 \times 10^{-5}$) for RL inoculated with the low level of LM and 6.4 ± 0.3 Log CFU/g vs. 6.1 ± 0.1 Log CFU/g ($P = 0.02$) for RL inoculated with the high level of LM (**Figure 1D**).

LM behaviors in contaminated RL harvested from two seasons

For the early-season samples, both LM-inoculated RL with the low inoculation level (3.5 ± 0.1 Log CFU/g) and the high inoculation level (6.5 ± 0.2 Log CFU/g) had no differences in LM concentration between UBD and UBD5 (**Figure 2A**). For the late season samples, the LM-inoculated RL exhibited a statistically significant decrease in LM concentrations (0.65 and 0.56 Log CFU/g decreases, respectively, for both RL inoculated with the low level of LM ($P = 0.002$) and RL inoculated with the high level of LM ($P = 0.008$) from UBD to UBD5 (**Figure 2B**). Results suggest that the population reduction of LM in RL harvested from the late season but not in the early season could be due to the differences among microorganisms between the RL samples from the two seasons.

Impact of storage time, season, LM contamination, and inoculation level on bacterial communities in RL

The culture-dependent results suggested that 5-day storage at 4 °C caused no impact on APC and AnPC across all RL samples, except AnPC in the late-season RL inoculated with the high level of LM.

On the other hand, the harvest season and LM-contamination, as well as the inoculation levels, played critical roles in changing APC and AnPC in RL. To further evaluate the impact of the storage time, season, LM contamination, and inoculation level on RL microbiota, 16S rRNA gene amplicon DNA sequencing was employed to take a closer look at the changes in diversity and composition of bacterial communities in RL and identify critical taxa associated with the four factors. For the diversity of bacterial communities, **Figure S3** shows no difference in alpha diversity based on observed features or the Shannon index of bacterial communities in RL was observed between groups of RL based on storage time, LM contamination, and inoculation level, except for season. The number of observed features of bacterial communities in early-season RL (50 ± 7) is more than that in late-season RL (31 ± 10) ($P = 5.3 \times 10^{-7}$). The Shannon indexes also illustrated ($P = 2.7 \times 10^{-7}$) with 3.6 ± 0.3 significantly higher values (3.6 ± 0.3) in early-season RL than the values (2.5 ± 0.7) in late-season RL ($P = 2.7 \times 10^{-7}$).

Differential abundance analyses, including the ANCOM test and RF feature selection method, were applied for investigating critical bacteria that were significantly impacted by the four factors. Results showed that there were no significant differences in bacterial diversity and composition between RL on UBD and UBD5. However, indicators were identified when RL samples were grouped by seasons and LM contamination. The ANCOM test (**Figure 3A**) showed that *Lactococcus* and *Erwinia* were the top two bacterial genera with the highest W values (57 and 55) with corresponding negative clr F statistics of -2.3 and -2.6, respectively. All these statistics identified *Lactococcus* and *Erwinia* as indicators for the late-season RL. Similarly, *Janthinobacterium* (W=54; clr F=2.3), *Pseudomonas* (W=49; clr F=1.5), and *Serratia* (W=48; clr F=1.4) were identified as indicators for early-season RL.

A total of 408 ASVs were identified after the 16S rRNA gene sequencing data process. RF feature selection analysis ranked all ASVs based on their MDAs. Among the ASVs, 97, 306, and 5 ASVs had positive, zero, and negative MDAs, respectively, indicating that 24% of ASVs made a positive contribution to RL classification based on seasons, while most ASVs (75%) made no contribution to the classification and 1% ASVs negatively contributed to the classification. After taxonomic analysis, 97 ASVs with positive MDAs corresponded with 22 genera (**Figure 3B**). Among the selected features, 17

genera had higher relative abundance in the early-season RL, including *Pseudomonas*, *Serratia*, *Flavobacterium*, *Janthinobacterium*, *Pantoea*, *Pseudarcobacter*, *Cetobacteirum*, *Mycoplasma*, *Weissella*, *Listeria*, *Carnobacterium*, *Rahnella*, *Exiguobacterium*, *Yersinia*, *Lactobacillus*, *Stenotrophomonas*, and *Arthrobacter*. The other five genera were considered indicators for the late-season RL, including *Leuconostoc*, *Lactococcus*, *Pedobacter*, *Duganella*, and *Erwinia* (**Figure 3C**).

To investigate bacteria in RL impacted by LM contamination, the ANCOM test showed that only *Listeria* (W = 60; clr F = -3.2) was an identified indicator for LM-contaminated RL (**Figure 4A**). The RF feature selection method calculated and ranked MDAs of 408 ASVs for the classification of contaminated and non-contaminated RL. Among these features, 38, 277, and 93 ASVs had positive, zero, and negative MDAs, respectively (**Figure 4B**). The results indicated that 9% of ASVs made a positive contribution to the RL contamination classification, 68% of ASVs had no contribution, and 23% ASVs even made a negative contribution to the RL contamination classification. After the taxonomic analysis of ASVs with positive MDA, 17 genera were identified. Nine genera were indicators for non-contaminated RL; they were *Leuconostoc*, *Yersinia*, *Arthrobacter*, *Pantoea*, *Stenotrophomonas*, *Lactococcus*, *Flavobacterium*, *Mycoplasma*, and *Exiguobacterium*. Eight genera were indicators for contaminated RL. The genera were *Pseudomonas*, *Serratia*, *Pedobacter*, *Jathinobacterium*, *Chryseobacterium*, *Duganella*, *Listeria*, and *Weissella* (**Figure 4C**).

To investigate the bacteria in RL impacted by the inoculation level of LM, the ANCOM test identified *Listeria* (W = 60; clr F = 580) as the indicator for RL with a high inoculation level of LM. No indicator was identified for RL inoculated with the low level of LM (**Figure 4D**). However, the RF feature selection method identified five genera as indicators for RL inoculated with the low level of LM. The genera were *Serratia*, *Pedobacter*, *Janthinobacterium*, *Flavobacterium*, and *Weissella*. Another four genera were identified as indicators for RL inoculated with the high level of LM, with these genera being *Pseudomonas*, *Chryseobacterium*, *Duganella*, and *Listeria* (**Figure 4E**).

Identification and quantification of identified competitive exclusion species in RL

Based on the double-layer method, 47 colonies with inhibitory activity against LM on TSA media were isolated from the frozen RL homogenate samples. After the antimicrobial activity test using the spot-on-lawn approach, 32 isolates were confirmed as antagonistic bacteria to be antagonistic to LM. Afterward, Sanger sequencing based on the 16S rRNA gene and WGS methods were applied to identify these isolates. The Sanger sequencing data process and analysis illustrated that the 32 antagonistic isolates belonged to three genera, with these genera being *Carnobacterium* (n = 13), *Lactococcus* (n = 3), and *Bacillus* (n =16). The WGS data process and analysis further identified these antagonistic genera to be *B. filamentous*, *C. maltaromaticum*, *L. lactis*, and (**Table S2**). To understand the distribution of the three antagonistic species across the RL samples, qPCR was applied to quantify their populations.

Standard curves of the qPCR were established based on the linear relationship between cell concentration of serial 10-fold dilution of pure cultures and corresponding CT values. The equations of standard curves for *C. maltaromaticum*, *Lc. lactis*, and *B. filamentosus* quantification are y = -3.4x + 42.7 with R^2 is 0.99 (**Figure 5A**), y = -3.7x + 42.1 with R^2 is 0.99 (**Figure 5B**), and y = -3.6x + 37.9 with R^2 is 0.99 (**Figure 5C**), respectively. Based on these standard curves, the population of three antagonistic species were quantified across all 48 RL samples.

C. maltaromaticum, *Lc. lactis*, and *B. filamentosus* were present in all the early- and late-season RL. In the early-season RL, the concentrations of the three species were 5.3 ± 0.4, 3.6 ± 1.5, and 1.7 ± 0.5 Log CFU/g, respectively. In late-season RL, their concentrations were 5.0 ± 0.6, 2.9 ± 0.5, and 2.4 ± 0.3 Log CFU/g, respectively. When every antagonistic species between the two seasons were compared, no significant differences in *C. maltaromaticum* concentration and *Lc. lactis* concentration between the early-season RL and the late-season RL were observed ($P > 0.05$) (**Figures 5D and 5E**). However, a significant distinction was observed in *B. filamentosus* populations between the RL harvested from the two seasons ($P < 0.05$) (**Figure 5F**).

Discussion

Numerous foodborne outbreaks of LM infection linked to fresh vegetables have occurred over the past decade (3), which suggests that current sanitation or intervention strategies are limited in their ability to control LM. A washing process with sanitizers, such as chlorine or peroxyacetic acid, has been commonly used in fresh produce processing to reduce the overall bacterial population but not necessarily target pathogens (6, 12). Since these processes are normally used for a one-time treatment during fresh produce processing, the bacteria that are reduced as a result of the treatment can grow back to the before-treatment levels over time under proper conditions (6, 7). Therefore, strategies that can target foodborne pathogens and provide consistent antimicrobial effects are needed. A few studies have reported that antagonistic bacteria exhibit consistent antimicrobial activity against various foodborne pathogens, including *E. coli* O157:H7, LM, and *S. enterica*, in fresh produce (16–18). However, studies on the investigation of antagonistic isolated from RL are limited. Understanding the structures of bacterial communities in RL and investigating the impact of LM contamination on RL microbiota is of primary importance and necessary. In this study, we first evaluated the influence of storage, season, LM contamination, and inoculation level on the diversity and composition of the bacterial communities in RL; second, we identified the specific bacteria significantly impacted by these factors; and third, we isolated and identified antagonistic bacteria that inhibited LM from RL.

RL samples used in this study were purchased from local grocery stores during two harvest seasons (early- and late-season). Results from culture-dependent methods (PCA and AA media) and culture-independent methods (16S rRNA gene sequencing) illustrated that the early-season RL had significantly higher APC but low AnPC, as well as higher alpha diversity both in observed features and the Shannon index, indicating the critical role played by harvesting season on bacterial abundance, diversity, and composition of bacterial communities in RL. Our previous studies (19) also showed that the alpha diversity in observed features of bacterial communities in early-season RL presented a higher number than that in late-season RL, but no differences in the Shannon index were observed. The differences in the Shannon index between the two studies indicate that differences in bacterial abundances between RL from two seasons in this study were larger than that in our previous study (19). Similar

observations were made by Williams et al. (2013) (11). They also showed that season played an important role in affecting the abundance, diversity, and composition of bacterial communities in the phyllosphere in RL during planting. Their results also showed the abundance of culturable aerobic bacteria was lower in RL planted in the early season (June) than RL planted in the late season (August and October). *Leuconostoc*, *Lactococcus*, *Bacillus*, and *Exiguobacterium* were identified as the most dominant genera in early-season plants, and *Pantoea*, *Pseudomonas*, *Erwinia*, and Enterobacteriaceae were determined as the predominant bacteria in late-season plants.

In addition to bacterial abundance and diversity, the ANCOM test and the RF feature selection method identified specific bacteria significantly impacted by the season, which were regarded as indicators for early-season and late-season RL. Based on the ANCOM test, *Janithonbacterium*, *Pseudomonas*, and *Serratia* were identified as indicators for the early-season RL. *Lactococcus* and *Erwinia* were identified as indicators for the late-season RL. Based on the RF feature selection method, 22 bacteria were identified as associated with seasons, including 17 indicators for early-season RL and five indicators for late-season RL. When comparing the indicators identified by the two methods, it was found that ANCOM-based indicators were covered by RF-based indicators for both the early-season RL and the late-season RL. This suggested that the RF feature selection method is more sensitive than the ANCOM method to catch the effect of season on bacteria between groups of samples. Sheh et al. (2022) reported a similar result, noting that the ANCOM test identified two indicators while the RF method selected nine indicators associated with the development of strictures (7).

The sensitivity differences in the identification of indicators could be caused by their algorithms. The ANCOM test is a statistical framework used for the comparison of the composition of two or more microbiota groups to identify features with significantly different abundances (34). The ANCOM test's identification of features was carried out by the calculation of the centered log ratio (clr) F statistic and the W statistic. The clr F statistic is measured to evaluate the effect size of features between groups, in which the microbial abundance data was clr-transformed. The W statistic is the number of times that the log-ratio of a taxon with every other taxon being tested was identified to be significantly different across

groups (19). RF is a supervised learning method for classification and regression, which constructs a number of decision trees during the training time and outputs different classes (classification) or mean predictions (regression) of individual decision trees (36). The RF feature selection method is based on the calculation and rank of the MDA, which is evaluated by measuring the loss in classification accuracy of the model when randomly permuting the values of the feature (38).

The five indicators for the late-season RL contained *Leuconostoc*, *Lactococcus*, *Pedobacter*, *Duganella*, and *Erwinia*. Among them, *Leuconostoc* and *Lactococcus* are the core genera of lactic acid bacteria (LAB) (39), which have been reported to be inhibitory against LM in various kinds of fresh produce (12, 40). *Leuconstoc* is a genus of LAB. All species included in this genus are heterofermentative, can producing multiple metabolites. The metabolites include lactic acid, ethanol, acetic acid, and carbon dioxide (41). These metabolites can decrease the environmental pH, create an anaerobic environment, or denature proteins of the microbial cells to inactivate the microbes, including pathogens (42). *Lactococcus,* a genus in the LAB group, can produce lactic acid and/or bacteriocin, e.g., nisin, to inhibit other bacteria or pathogens, e.g., *Lc. lactis* (43). Strains with antagonistic activities against LM in fresh produce have been reported (12). In addition, *Erwinia* is a genus containing mostly species that are pathogenic to plants, e.g. *Erwinia carotovora,* causing bacterial soft rot (44). *Erwinia* also has been reported to have an inhibitory capacity against LM in fresh produce (16, 17). In addition, *Erwinia* was also reported by Williams et al. (2013) to inhibit an attenuated strain of *E. coli* O157:H7 on the leaf of Romaine lettuce planted in the field (11). These LAB indicators identified for late-season RL could be a potential reason for the decrease in LM abundance from UBD to UBD5 in late-season RL but not early-season RL.

This study also investigated the effect of LM contamination on the structures of bacterial communities. Based on the culture-dependent methods, no changes in APC were observed between RL before and after LM contamination, which indicates that LM contamination had no impact on APC in RL. However, significant increases in AnPC were seen after RL was inoculated with LM. The reason could be that LM is a facultative anaerobic bacterium (45). Studies showed that oxygen concentration does not

impact most strains of LM (45, 46). Thus, LM could grow on the aerobic agar in anaerobic chambers with oxygen removed by using anaerobic packs. Based on the 16S rRNA gene sequencing approach, no differences in alpha diversity based on observed features and Shannon index ($P > 0.05$) of bacterial communities were apparent between RL before and after LM contamination. These results indicate that LM contamination did not impact bacterial diversity in RL. However, we used the ANCOM test and the RF feature selection method to identify bacteria that were significantly impacted by contamination. The ANCOM test detected only *Listeria* as having a significantly higher relative abundance in contaminated RL than in non-contaminated RL (**Figure 4A**), which was identified as a indicator for contaminated RL. While using the RF feature selection method, eight indicators were identified for contaminated RL. These indicators were *Pseudomonas*, *Serratia*, *Pedobacter*, *Jantinobacterium*, *Chryseobacterium*, *Duganella*, *Listeria*, and *Weissella* (**Figure 4C**).

The impact of inoculation levels of LM (3 and 6 Log CFU/g) on the diversity and composition of bacterial communities in RL was also evaluated. Similar to LM contamination, inoculation levels of LM have no influence on the bacterial alpha diversity in RL. When used to identify bacterial indicators for the RL inoculated at different inoculation levels, the ANCOM test identified only *Listeria* as an indicator for RL with a high inoculation level. No indicator was determined for RL with a low LM inoculation level (**Figure 4D**). However, the RF feature selection method identified five indicators for RL with a low inoculation level, with these indicators being *Serratia*, *Pedobacter*, *Janthinobacterium*, *Flavobacterium*, and *Weissella.* There were four indicators for RL with a high inoculation level, *Pseudomonas*, *Chryseobacterium*, *Duganella*, and *Listeria*. (**Figure 4E**). When comparing the ANCOM test and the RF feature selection method for indicator identification, both methods identified *Listeria* as an indicator for contaminated RL, indicating the inoculated LM can be successfully detected by 16S rRNA gene sequencing with these two approaches. LM at the high level (6 Log CFU/g) was identified while LM at the low level (3 Log CFU/g) was not detected by either the ANCOM test or the RF feature selection method. The other seven indicators detected by the RF feature selection suggest that the identification sensitivity of indicators is higher than for the ANCOM test.

To further explore the reason that the LM decreased in concentration in late-season RL during the 5-day storage at 4 °C, the double layer method and spot-on-lawn method were applied to randomly isolate and confirm 32 CE bacteria with antagonistic activity against LM from frozen RL samples. Subsequently, Sanger sequencing and WGS were applied to identify 32 antagonistic isolates as belonging to three species; the species were *C. maltaromaticum*, *Lc. lactis*, and *B. filamentosus* (**Table S2**). It was noted that antagonistic bacteria could not be isolated from fresh RL but were isolated from the frozen RL homogenates stored at -80 °C for 1 month. This result indicates that first, no antagonistic was isolated from the most abundant culturable bacteria in fresh RL; and second, the frozen temperature may cause certain bacteria to enter the VBNC state (47, 48) to reduce the magnitude of the bacteria population when plated on TSA agar media. Frozen conditions can be a strategy to reduce the culturable bacterial level and then allow us to explore the culturable bacteria that are not dominant or the low-temperature-tolerant bacteria in the produce microbiota.

The qPCR was employed to validate and quantify the three antagonistic species identified by sequencing analyses existing in 48 RL samples. Although no significant differences were apparent in *C. maltaromaticum* and *Lc. lactis* concentrations between early-season and late-season RL, *B. filamentosus* had a significantly higher concentration in late-season RL than in early-season RL (**Figure 5F**), suggesting that *B. filamentosus* could be one of the reasons for the reduction in LM abundance in late-season RL but not early-season RL. To compare the abundances of the three antagonists in early-season and late-season RL, *C. maltaromaticum* had the highest magnitude of average level magnitude (~ 5 Log CFU/g), followed by *L. lactis* (~ 3 Log CFU/g). *B. filamentosus* had the lowest magnitude of average level (~ 2 Log CFU/g). *C. maltaromaticum* is a species of psychrotropic LAB, which is known for its ability to produce bacteriocin (49). *C. maltaromaticum* is frequently isolated from dairy, meat, fish, and shrimp (50, 51), but has not been reported to be isolated from RL. *C. maltaromaticum* has been extensively studied to show inhibition of LM in fish and meat products (51, 52). Pedrozo et al. (2019) reported that *C. maltaromaticum* H-17 reduced the LM population by ~ 4 Log CFU/ml compared to the LM control when co-cultured with *C. maltaromaticum* H-17 in sterile raw fish extract at 5 ± 2°C after 50

h (52). *Lc. lactis* is another species of LAB that has been extensively used in the production of buttermilk and cheese. Dong et al. (2021) applied a fermentate consisting of 8 Log CFU/g *Lc. lactis* 537 cells and produced nisin to reduce 3-4 log CFU/g LM to the level below the limit of detection immediately on fresh cut iceberg stored at 4 °C (43). Gu et al. (2022) isolated multiple strains of *Lc. lactis* from Romaine lettuce and showed inhibitory activity against LM when co-cultured in Romaine lettuce juice at 30 °C after two days compared to the LM control (12). *B. filamentosus* is an endospore forming, Gram positive bacterium that was first isolated from marine sediment and identified as a new species of *Bacillus* (*Bacillus filamentosus* sp. nov. SGD-14^{T}) by Sonalkar et al. (2015) (53). Abiala et al. (2022) isolated a strain of *B. filamentosus* (C8) from cowpea and found that C8 inoculation could increase the biomass of cowpea root (54). In addition, Gu et al. (2020) isolated sanitizer-resistant *B. filamentosus* from the rinse water of spinach treated with 50 mg/L peracetic acid. Our study is the first to report the isolation of *B. filamentosus* from Romaine lettuce and its antagonistic activity against LM (55). Due to the sanitizer-resistant property, *B. filamentosus* has the potential to be applied as a biocontrol strategy collaborating with the sanitation washing process to improve microbial produce safety.

Conclusion

In summary, this study evaluated the impacts of storage, season, LM contamination and inoculation level on the diversity and composition of bacterial communities in RL. The factor of harvest season had the most significant impact on the diversity and composition of bacterial communities in RL. To compare bacterial communities between contaminated and non-contaminated samples, the ANCOM test identified *Listeria* as the indicator for LM-contaminated RL, while the RF feature selection method selected 17 bacteria as indicators, including eight for contaminated samples and nine for non-contaminated samples. For samples inoculated with a low level of LM, the ANCOM test identified no indicators, while the RF method identified *Serratia, Pedobacter*, *Janthinobacterium, Flavobacterium,* and *Weissella* as indicators. For samples inoculated with a high level of LM, the ANCOM test identified *Listeria* as the indicator, while the RF method identified *Pseudomonas*, *Chryseobacterium*, *Duganella*,

and *Listeria* as indicators. Thirty-two antagonistic bacteria that inhibited LM were isolated from the frozen RL homogenates and confirmed by using the double-layer agar method and spot-on-lawn method. Three antagonistic species were identified from the 32 antagonistic isolates based on the Sanger sequencing and WGS. These three species were *C. maltaromaticum* (n = 13), *Lc. lactis* (n = 3), and *B. filamentosus* (n = 16). The qPCR results confirmed the significant differences in levels of *B. filamentosus* between early-season RL and late-season RL, indicating that *B. filamentosus* could be the key antagonist in late-season RL. Future studies are needed to better confirm the antagonistic function of *B. filamentosus* by conducting co-culturing experiments.

References

1. 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. Emerging Infect Dis 19:407–415.
2. Scharff RL. 2010. Health-related costs from foodborne illness in the United States. https://www.pewtrusts.org/~/media/legacy/uploadedfiles/phg/content_level_pages/reports/pspscharff20v9pdf.pdf [Accessed November 1, 2022].
3. CDC National Outbreak Reporting System (NORS) Dashboard. 2022. https://wwwn.cdc.gov/norsdashboard [Accessed November 1, 2022].
4. Söderqvist K, Ahmed Osman O, Wolff C, Bertilsson S, Vågsholm I, Boqvist S. 2017. Emerging microbiota during cold storage and temperature abuse of ready-to-eat salad. Infect Ecol Epidemiol 7:1328963.
5. Telias A, White JR, Pahl DM, Ottesen AR, Walsh CS. 2011. Bacterial community diversity and variation in spray water sources and the tomato fruit surface. BMC Microbiol 11:81.
6. Gu G, Ottesen A, Bolten S, Ramachandran P, Reed E, Rideout S, Luo Y, Patel J, Brown E, Nou X. 2018. Shifts in spinach microbial communities after chlorine washing and storage at compliant and abusive temperatures. Food Microbiol 73:73–84.
7. Shen Z, Mustapha A, Lin M, Zheng G. 2017. Biocontrol of the internalization of Salmonella enterica and Enterohaemorrhagic *Escherichia coli* in mung bean sprouts with an endophytic Bacillus subtilis. Int J Food Microbiol 250:37–44.
8. Nou X, Luo Y. 2010. Whole-leaf wash improves chlorine efficacy for microbial reduction and prevents pathogen cross-contamination during fresh-cut lettuce processing. J Food Sci 75:M283–90.
9. Singh P, Hung Y-C, Qi H. 2018. Efficacy of peracetic acid in inactivating foodborne pathogens on fresh produce surface. J Food Sci 83:432–439.

10. Liao C, Wang L. 2021. Evaluation of the bacterial populations present in Spring Mix salad and their impact on the behavior of *Escherichia coli* O157:H7. Food Control 107865.
11. 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.
12. Gu G, Kroft B, Lichtenwald M, Luo Y, Millner P, Patel J, Nou X. 2022. Dynamics of *Listeria monocytogenes* and the microbiome on fresh-cut cantaloupe and romaine lettuce during storage at refrigerated and abusive temperatures. Int J Food Microbiol 364:109531.
13. Zhang Y, Jewett C, Gilley J, Bartelt-Hunt SL, Snow DD, Hodges L, Li X. 2018. Microbial communities in the rhizosphere and the root of lettuce as affected by *Salmonella*-contaminated irrigation water. FEMS Microbiol Ecol 94.
14. Callaway TR, Edrington TS, Anderson RC, Harvey RB, Genovese KJ, Kennedy CN, Venn DW, Nisbet DJ. 2008. Probiotics, prebiotics and competitive exclusion for prophylaxis against bacterial disease. Anim Health Res Rev 9:217–225.
15. Nurmi E, Nuotio L, Schneitz C. 1992. The competitive exclusion concept: development and future. Int J Food Microbiol 15:237–240.
16. Kim W-I, Choi SY, Han I, Cho SK, Lee Y, Kim S, Kang B, Choi O, Kim J. 2020. Inhibition of *Salmonella enterica* growth by competitive exclusion during early alfalfa sprout development using a seed-dwelling *Erwinia persicina* strain EUS78. Int J Food Microbiol 312:108374.
17. Lopez-Velasco G, Tydings HA, Boyer RR, Falkinham JO, 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.
18. Russo P, de Chiara MLV, Vernile A, Amodio ML, Arena MP, Capozzi V, Massa S, Spano G. 2014. Fresh-cut pineapple as a new carrier of probiotic lactic acid bacteria. Biomed Res Int 2014:309183.

19. Liao C, Wang L. 2022. The microbial quality of commercial chopped romaine lettuce before and after the “use by” date. Front Microbiol 13:850720.

20. Zeng W, Vorst K, Brown W, Marks BP, Jeong S, Pérez-Rodríguez F, Ryser ET. 2014. Growth of *Escherichia coli* O157:H7 and *Listeria monocytogenes* in packaged fresh-cut romaine mix at fluctuating temperatures during commercial transport, retail storage, and display. J Food Prot 77:197–206.

21. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu Y-X, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37:852–857.

22. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet j 17:10.

23. 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.

24. Estaki M, Jiang L, Bokulich NA, McDonald D, González A, Kosciolek T, Martino C, Zhu Q, Birmingham A, Vázquez-Baeza Y, Dillon MR, Bolyen E, Caporaso JG, Knight R. 2020. QIIME 2 Enables Comprehensive End-to-End Analysis of Diverse Microbiome Data and Comparative Studies with Publicly Available Data. Curr Protoc Bioinformatics 70:e100.

25. 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.

26. Caporaso JG, 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. Nat Methods 7:335–336.

27. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA, Gregory Caporaso J. 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 6:90.

28. 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.

29. Wang H, Jiang X. 2022. Isolation and characterization of competitive exclusion microorganisms from animal wastes-based composts against *Listeria monocytogenes*. J Appl Microbiol 132:4531–4543.

30. Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ. 2008. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. Appl Environ Microbiol 74:2461–2470.

31. Bogaerts B, Winand R, Fu Q, Van Braekel J, Ceyssens P-J, Mattheus W, Bertrand S, De Keersmaecker SCJ, Roosens NHC, Vanneste K. 2019. Validation of a bioinformatics workflow for routine analysis of whole-genome sequencing data and related challenges for pathogen typing in a

european national reference center: Neisseria meningitidis as a Proof-of-Concept. Front Microbiol 10:362.

32. Abdi H, Williams LJ. 2010. Tukey's honestly significant difference (HSD) test. Encyclopedia of research design.

33. Xia Y, Sun J. 2017. Hypothesis testing and statistical analysis of microbiome. Genes Dis 4:138–148.

34. Mandal S, Van Treuren W, White RA, Eggesbø M, Knight R, Peddada SD. 2015. Analysis of composition of microbiomes: a novel method for studying microbial composition. Microb Ecol Health Dis 26:27663.

35. Stevens AJ, Purcell RV, Darling KA, Eggleston MJF, Kennedy MA, Rucklidge JJ. 2019. Human gut microbiome changes during a 10-week Randomised Control Trial for micronutrient supplementation in children with attention deficit hyperactivity disorder. Sci Rep 9:10128.

36. Breiman L. 2001. Random forests. Machine learning 45:5-32.

37. RColorBrewer S, Liaw MA. Package 'randomforest'. University of California, Berkeley: Berkeley, CA, USA.

38. Wang H, Yang F, Luo Z. 2016. An experimental study of the intrinsic stability of random forest variable importance measures. BMC Bioinformatics 17:60.

39. Françoise L. 2010. Occurrence and role of lactic acid bacteria in seafood products. Food Microbiol 27:698–709.

40. Shao X, Fang K, Medina D, Wan J, Lee JL, Hong, SH. 2020. The probiotic, *Leuconostoc mesenteroides*, inhibits *Listeria monocytogenes* biofilm formation. J Food Safety 40.

41. Mokoena MP, Omatola CA, Olaniran AO. 2021. Applications of Lactic Acid Bacteria and Their Bacteriocins against Food Spoilage Microorganisms and Foodborne Pathogens. Molecules 26.

42. Webb L, Ma L, Lu X. 2022. Impact of lactic acid bacteria on the control of *Listeria monocytogenes* in ready-to-eat foods. Food Quality and Safety.

43. Dong A, Malo A, Leong M, Ho VTT, Turner MS. 2021. Control of *Listeria monocytogenes* on ready-to-eat ham and fresh cut iceberg lettuce using a nisin containing *Lactococcus lactis* fermentate. Food Control 119:107420.

44. Zhao Y, Li P, Huang K, Wang Y, Hu H, Sun Y. 2013. Control of postharvest soft rot caused by *Erwinia carotovora* of vegetables by a strain of *Bacillus amyloliquefaciens* and its potential modes of action. World J Microbiol Biotechnol 29:411–420.

45. Roberts BN, Chakravarty D, Gardner JC, Ricke SC, Donaldson JR. 2020. *Listeria monocytogenes* Response to Anaerobic Environments. Pathogens 9.

46. Couvert O, Divanac'h M-L, Lochardet A, Thuault D, Huchet V. 2019. Modelling the effect of oxygen concentration on bacterial growth rates. Food Microbiol 77:21–25.

47. Li Y, Huang T, Bai C, Fu J, Chen L, Liang Y, Wang K, Liu J, Gong X, Liu J. 2020. Reduction, prevention, and control of *Salmonella enterica* viable but non-culturable cells in flour food. Front Microbiol 11:1859.

48. Wei C, Zhao X. 2018. Induction of Viable but Nonculturable *Escherichia coli* O157:H7 by low temperature and its resuscitation. Front Microbiol 9:2728.

49. Martin-Visscher LA, van Belkum MJ, Garneau-Tsodikova S, Whittal RM, Zheng J, McMullen LM, Vederas JC. 2008. Isolation and characterization of carnocyclin a, a novel circular bacteriocin produced by *Carnobacterium maltaromaticum* UAL307. Appl Environ Microbiol 74:4756–4763.

50. Afzal MI, Jacquet T, Delaunay S, Borges F, Millière J-B, Revol-Junelles A-M, Cailliez-Grimal C. 2010. *Carnobacterium maltaromaticum*: identification, isolation tools, ecology and technological aspects in dairy products. Food Microbiol 27:573–579.

51. Leisner JJ, Laursen BG, Prévost H, Drider D, Dalgaard P. 2007. *Carnobacterium*: positive and negative effects in the environment and in foods. FEMS Microbiol Rev 31:592–613.

52. Pedrozo HA, Dallagnol AM, Vignolo GM, Pucciarelli AB, Schvezov CE. 2019. Mechanistically inspired kinetic approach to describe interactions during co-culture growth of *Carnobacterium maltaromaticum* and *Listeria monocytogenes*. J Food Sci 84:2592–2602.

53. Sonalkar VV, Mawlankar R, Venkata Ramana V, Joseph N, Shouche YS, Dastager SG. 2015. *Bacillus filamentosus* sp. nov., isolated from sediment sample. Antonie Van Leeuwenhoek 107:433–441.

54. Abiala MA, Sahoo L. 2022. *Bacillus aryabhattai* enhanced proline content, stabilized membrane and improved growth of cowpea under NaCl-induced salinity stress. J Appl Microbiol 133:1520–1533.

55. Gu G, Ottesen A, Bolten S, Luo Y, Rideout S, Nou X. 2019. Microbiome convergence following sanitizer treatment and identification of sanitizer resistant species from spinach and lettuce rinse water. Int J Food Microbiol 318:108458.

Table 1. Primers used in qPCR for quantifying identified antagonistic bacteria from RL homogenate.

Antagonist	Gene name	Primer name	Primer sequence	Amplicon length (bp)	Reference
Bacillus filamentosus	Superoxide dismutase gene (soda)	BF_sodA_F	TCGTCCCAGTTTACAACATTCC	148	(56)
		BF_sodA_R	CAAGCACTCCAAACCAAGATTC		
Lactococcus lactis	Elongation factor Tu gene (tuf)	LL_tuf_F	TGAAGAATTGATGGAACTCG	126	(57)
		LL_tuf_R	CATTGTGGTTCACCGTTC		
Carnobacterium maltaromaticum	Recombinase A gene (RecA)	CM_RecA_F	CGTTAGGAGTAGGCGGGTATC	111	(58)
		CM_RecA_R	CATGTTTTTGAGCTTCTGCTAC		

RL: Romaine lettuce.

Table S1. Statistical summary of amplicon sequence variants (ASVs) before and after the denoising process during the QIIME 2 analysis.

	Number of samples	Total frequency	Minimum frequency	Median frequency	Maximum frequency
Sequences before denoising	48	5,437,814	45,667	118,458	199,809
Sequences after denoising	48	3,619,180	31,329	79,046	144,009

Table S2. Antagonistic bacteria identified by Sanger sequencing based on the 16S rRNA gene and WGS and their antagonistic capacity based on the LM-inhibition zone.

Sample ID	Sanger sequencing	WGS	Percent identity (%)	Inhibition zone (mm diameter)
CM1	*Carnobacterium*	*Carnobacterium maltaromaticum*	84.2	12.2 ± 2.6
CM2	*Carnobacterium*	*C. maltaromaticum*	83.4	11.8 ± 2.1
CM3	*Carnobacterium*	*C. maltaromaticum*	87.9	10.5 ± 3.1
CM4	*Carnobacterium*	*C. maltaromaticum*	83.1	11.3 ± 0.5
CM5	*Carnobacterium*	*C. maltaromaticum*	86.7	13.0 ± 1.6
CM6	*Carnobacterium*	*C. maltaromaticum*	84.7	12.8 ± 1.5
CM7	*Carnobacterium*	*C. maltaromaticum*	84.6	6.3 ± 1.0
CM8	*Carnobacterium*	*C. maltaromaticum*	83.9	14.0 ± 0.8
CM9	*Carnobacterium*	*C. maltaromaticum*	83.7	13.0 ± 2.4
CM10	*Carnobacterium*	*C. maltaromaticum*	84.2	12.5 ± 1.7
CM11	*Carnobacterium*	*C. maltaromaticum*	84.6	9.8 ± 2.2
CM12	*Carnobacterium*	*C. maltaromaticum*	85.1	13.8 ± 2.2
CM13	*Carnobacterium*	*C. maltaromaticum*	83.1	10.0 ± 2.2
LL1	*Lactococcus*	*Lactococcus lactis*	91.3	8.5 ± 1.9
LL2	*Lactococcus*	*Lc. lactis*	94.8	9.0 ± 1.4
LL3	*Lactococcus*	*Lc. lactis*	93.8	9.0 ± 1.4
BF1	*Bacillus*	*Bacillus filamentosus*	80.6	10.3 ± 0.6
BF2	*Bacillus*	*B. filamentosus*	82.5	10.3 ± 0.6
BF3	*Bacillus*	*B. filamentosus*	82.1	10.3 ± 0.6
BF4	*Bacillus*	*B. filamentosus*	77.9	11.3 ± 0.6
BF5	*Bacillus*	*B. filamentosus*	81.6	9.8 ± 1.0
BF6	*Bacillus*	*B. filamentosus*	81.8	10.5± 1.3
BF7	*Bacillus*	*B. filamentosus*	80.8	11.8 ± 1.7
BF8	*Bacillus*	*B. filamentosus*	74.7	11.8 ± 2.9
BF9	*Bacillus*	*B. filamentosus*	75.6	12.3 ± 2.6
BF10	*Bacillus*	*B. filamentosus*	82.6	8.7 ± 2.1
BF11	*Bacillus*	*B. filamentosus*	81.0	10.0 ± 1.0
BF12	*Bacillus*	*B. filamentosus*	82.0	8.0 ± 0.0
BF13	*Bacillus*	*B. filamentosus*	81.7	10.7 ± 1.2
BF14	*Bacillus*	*B. filamentosus*	80.9	10 ± 0.0
BF15	*Bacillus*	*B. filamentosus*	85.7	10.0 ± 1.7
BF16	*Bacillus*	*B. filamentosus*	79.6	9.3 ± 0.6

WGS: Whole genome sequencing; LM: *Listeria monocytogenes*.

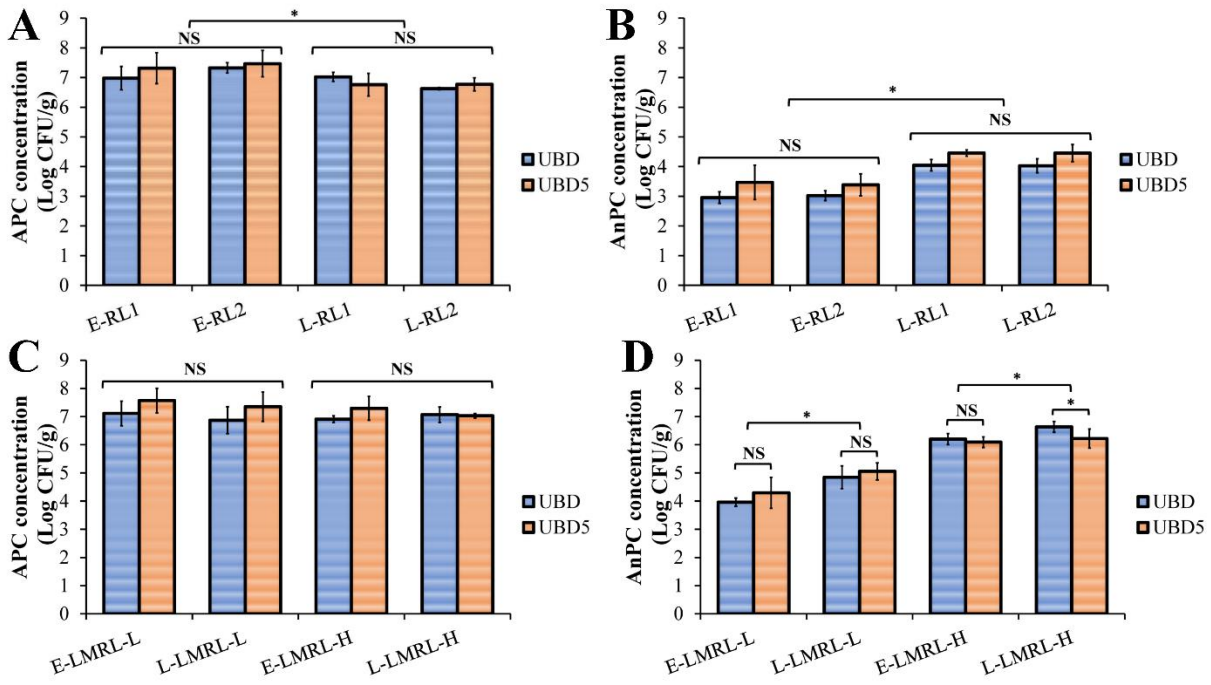

Figure 1. Aerobic plate count (APC) and anaerobic plate count (AnPC) of non-inoculated Romaine lettuce (RL) (A and B) and *Listeria monocytogenes* (LM)-inoculated RL (C and D) harvested from the early season and the late season on the "Use By" date (UBD) and 5 days after the UBD (UBD5) when stored at 4 °C. The letters "E" and "L" in the x-axis labels: early season and late season, respectively; "LMRL:": LM-inoculated RL; The last letters "L" and "H:" low inoculation level of LM and high inoculation level of LM, respectively; "NS:" no significance; "*:" $P < 0.05$.

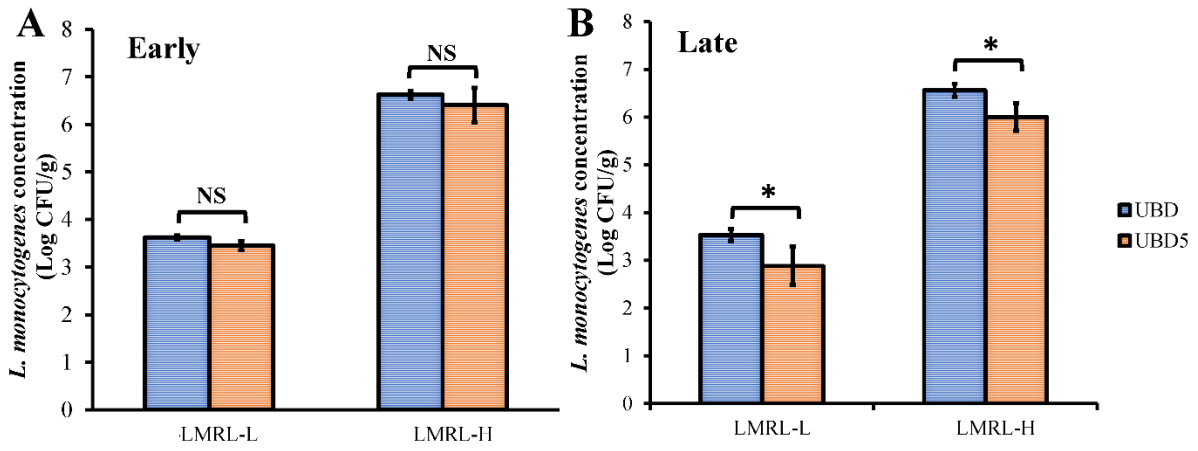

Figure 2. *Listeria monocytogenes* (LM) concentrations in Romaine lettuce (RL) when stored at 4 °C on the "Use By" date (UBD and 5 days after the "Use By" date (UDB5). RL samples, harvested from the early season (A) and the late season (B), were inoculated with LM at a low (~ 3 Log CFU/g) or a high level (~ 6 Log CFU/g). "LMRL:" LM-inoculated RL; "L:" low inoculation level of LM; "H:" high inoculation level of LM; "NS:" no significance; "*:" $P < 0.05$.

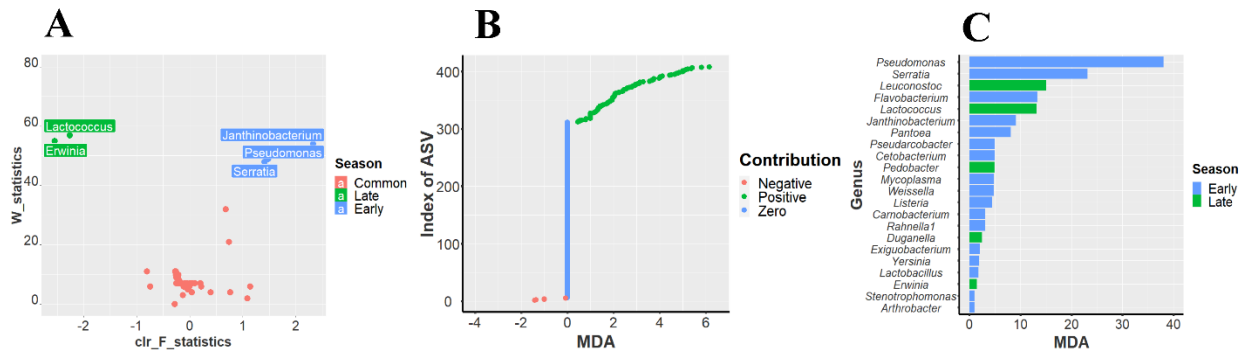

Figure 3. Comparison of indicators at the genus level between early-season and late-season Romaine lettuce (RL) identified by the analysis of composition of microbiomes (ANCOM) test and the random forests (RF) feature selection method. (A) Volcano plot of the genus indicators for the early-season and late-season RL identified by the ANCOM test; (B) Scatter plot of the ranked identified amplicon sequence variant (ASV) features based on the mean decrease in accuracy (MDA) calculated by the RF method; and (C) Horizontal bar plot of ranked taxa at the genus level with a positive contribution to the classification of early-season and late-season RL. "Common" represents the genera with no difference in relative abundance between RL from early and late seasons. "Early" and "Late" represent genera with remarkably higher relative abundance in early-season RL and late-season RL, respectively. "clr" strands for centered log ratio. "Negative," "Positive," and "Zero" represent ASV features with negative, positive, and zero MDA values, respectively.

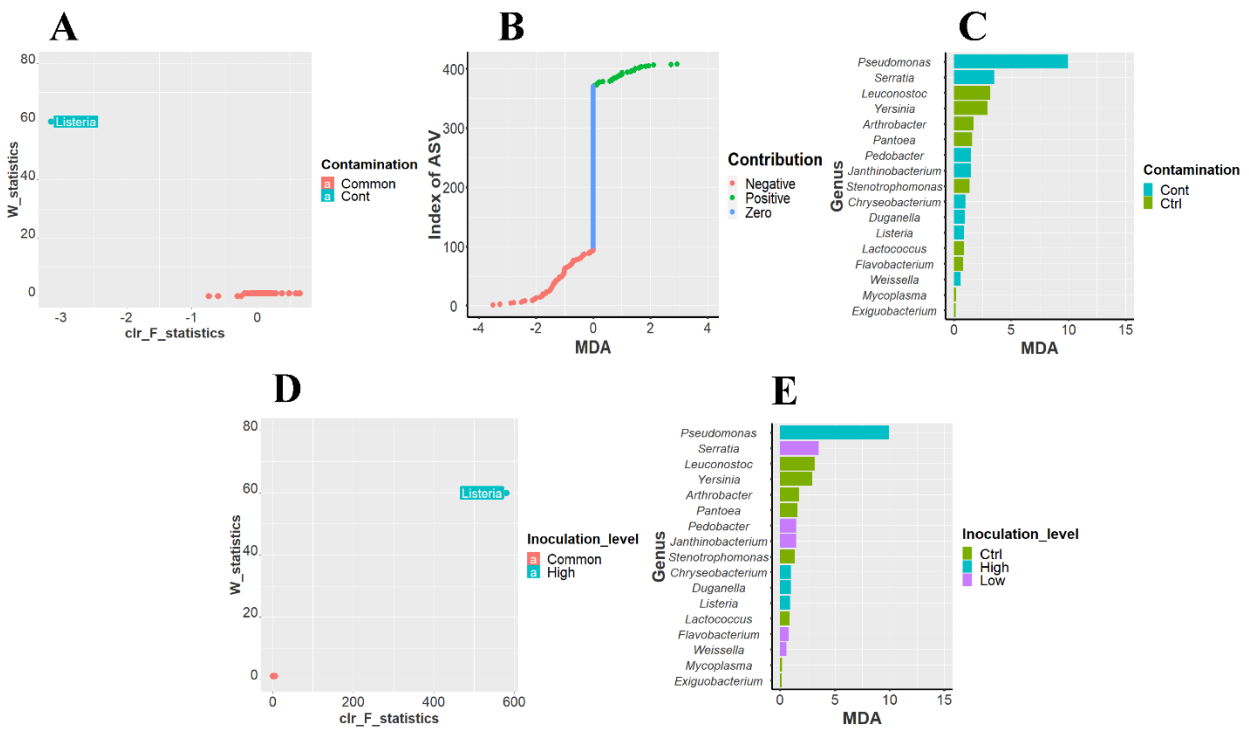

Figure 4. Comparison of indicators at the genus level between *Listeria monocytogenes* (LM)-contaminated and non-contaminated Romaine lettuce (RL) identified by the analysis of composition of microbiomes (ANCOM) test and the random forests (RF) feature selection. (A) Volcano plot of the genus indicators identified for LM-contaminated and non-contaminated RL; (B) Scatter plot of the ranked identified amplicon sequence variant (ASV) features based on the mean decrease in accuracy (MDA) calculated by the RF feature selection method for contaminated and non-contaminated RL; (C) Horizontal bar plot of ranked taxa at genus level with a positive contribution to the classification of contaminated and non-contaminated RL samples; (D) Volcano plot of the genus indicators identified for RL inoculated with low level, high level of LM, and non-inoculated RL; and (E) Horizontal bar plot of ranked taxa at the genus level with a positive contribution to the classification of RL inoculated with low level, high level of LM, and non-inoculated RL. "Common" represents the genera with no difference in relative abundance between contaminated and non-contaminated RL or RL inoculated with the low level, the high level of LM, and non-inoculated RL. "Cont" represents contaminated RL. "clr" strands for centered log ratio. "Negative," "Positive," and "Zero" represent ASV features with negative, positive, and zero MDA values, respectively. "Ctrl" stands for non-contaminated RL. "Low" and "High" stand for RL inoculated with low and high inoculation levels of LM, respectively.

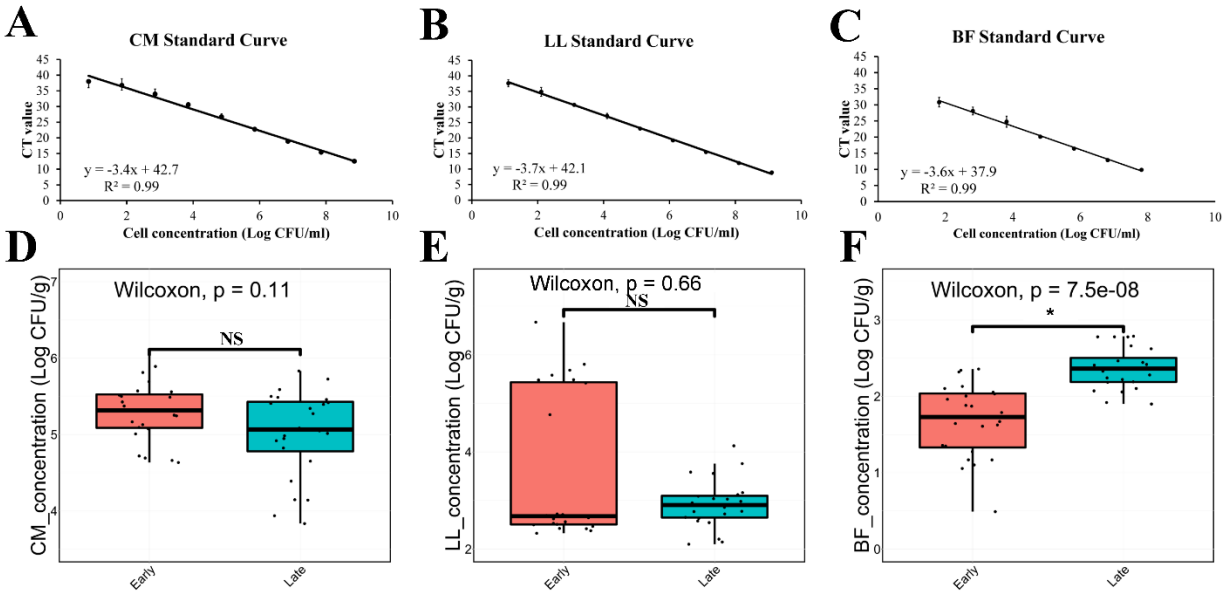

Figure 5. Quantification of three identified antagonistic bacteria, *Carnobacterium maltaromaticum* (CM), *Lactococcus lactis* (LL), and *Bacillus filamentosus* (BF), from all the Romaine lettuce samples by qPCR. (A-C) qPCR standard curves of the three antagonistic species, changes of the cycle threshold (CT) values as a function of the cell concentrations (Log CFU/ml). (D-F) Comparison of cell concentrations (Log CFU/g) of three antagonistic species, CM, LL, and BF, between early season (Early) and late season (Late) Romaine lettuce. "NS:" no significance; "*:" $P < 0.05$.

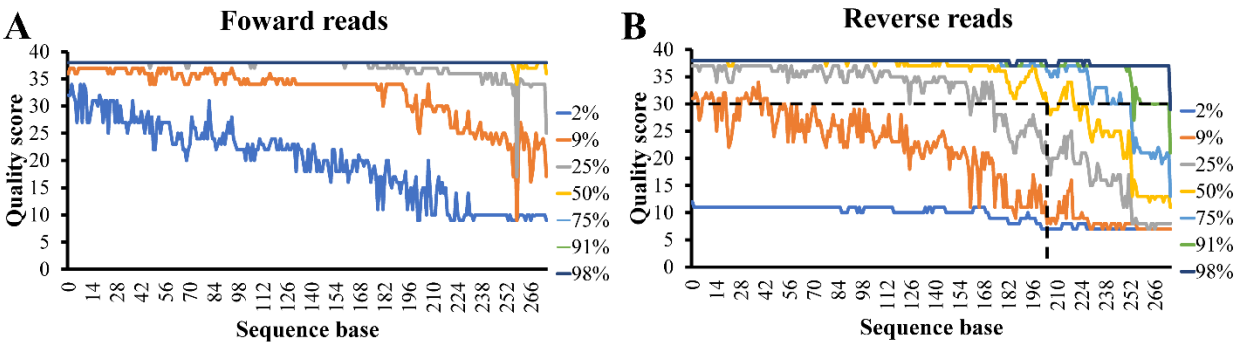

Figure S1. Interactive quality plots of forward and reverse reads with 276 bases each. (A) Interactive quality plots of the forward reads at seven percentiles. (B) Interactive quality plots of the reverse reads at seven percentiles, including 2%, 9%, 25%, 50%, 75%, 91%, and 98%. The horizontal dash line represents the threshold of Phred quality score of 30 (Q30). The vertical dash line stands for the intersection points between the Q30 dash line and the 50% interactive quality curve.

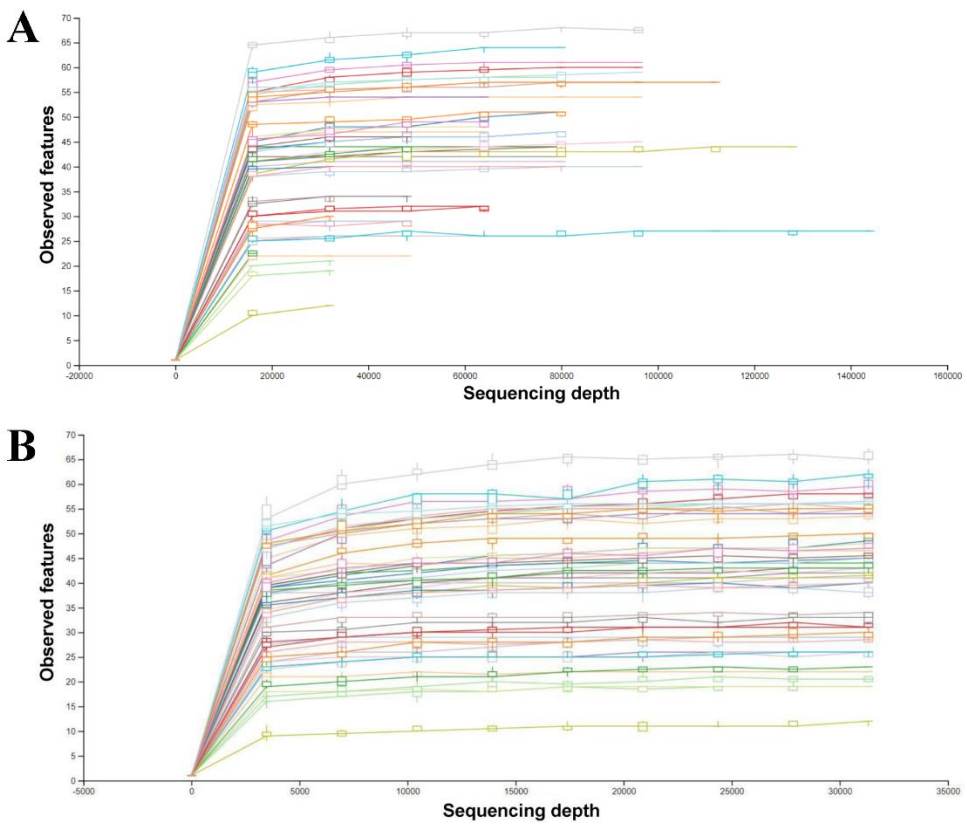

Figure S2. Sequencing depth curves before and after the rarefication over observed features in 48 samples. (A) Sequencing depth curves over observed features before rarefication. (B) Sequencing depth curves over observed features after rarefication at 31,329 reads.

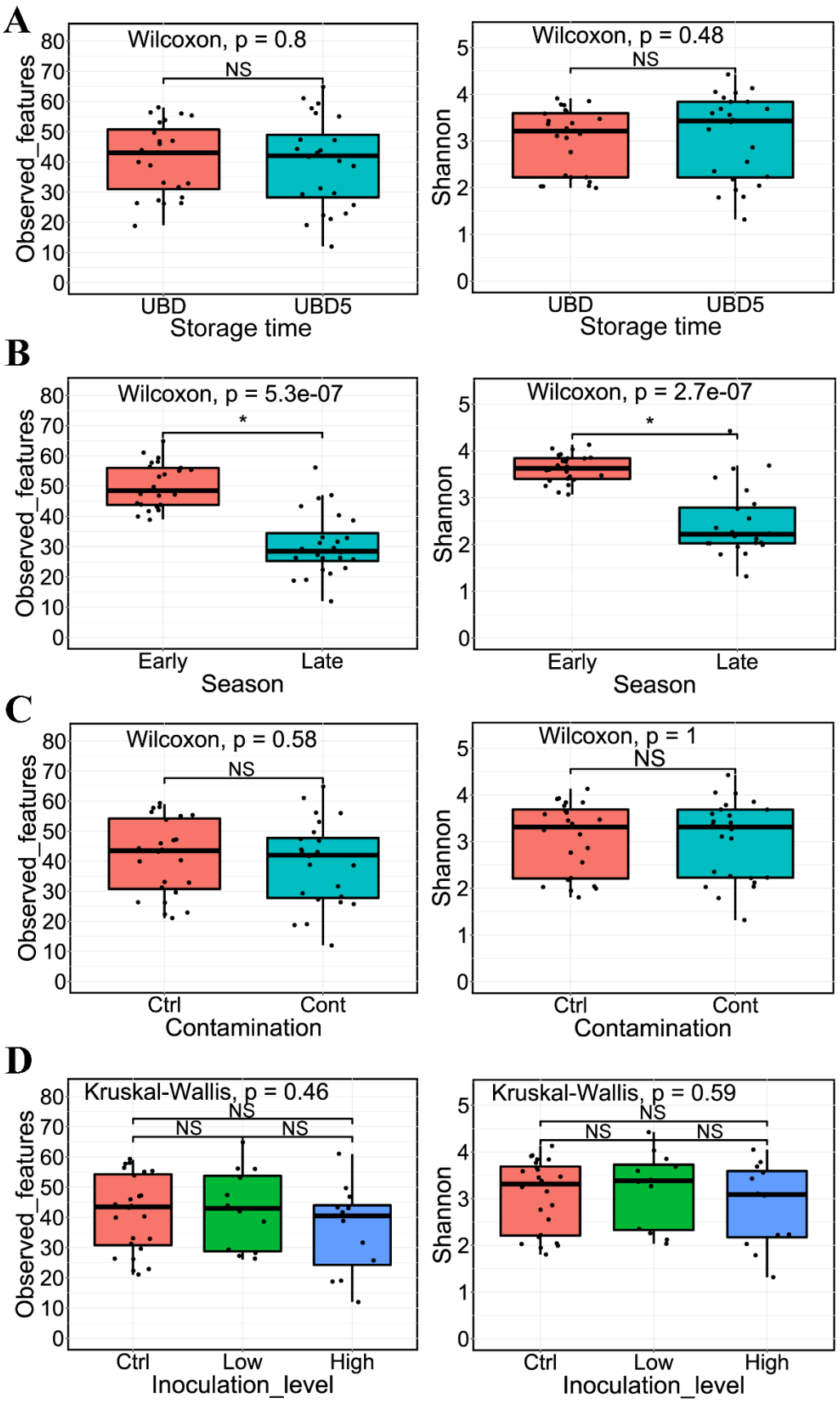
A
Wilcoxon, p = 0.8
NS
Observed_features
UBD
UBD5
Storage time
Wilcoxon, p = 0.48
NS
Shannon
UBD
UBD5
Storage time
B
Wilcoxon, p = 5.3e-07
*
Observed_features
Early
Late
Season
Wilcoxon, p = 2.7e-07
*
Shannon
Early
Late
Season
C
Wilcoxon, p = 0.58
NS
Observed_features
Ctrl
Cont
Contamination
Wilcoxon, p = 1
NS
Shannon
Ctrl
Cont
Contamination
D
Kruskal-Wallis, p = 0.46
NS
NS
NS
Observed_features
Ctrl
Low
High
Inoculation_level
Kruskal-Wallis, p = 0.59
NS
NS
NS
Shannon
Ctrl
Low
High
Inoculation_level

Figure S3. Comparison of the alpha diversity based on observed features (Left) and Shannon index (Right) of bacterial communities present in Romaine lettuce (RL) between UBD and UBD5 (A), between early- and late-season RL (B), between *Listeria monocytogenes* (LM)-contaminated and non-contaminated RL (C), and between RL inoculated with a low inoculation level of LM and RL with a high inoculation level of LM (D). "UBD:" "Use By" date; "UBD5:" 5 days after UBD; "Early:" early-season RL; "Late:" late-season RL; "Ctrl:" non-contaminated RL or non-inoculated RL; "Cont:" LM-contaminated RL; "Low:" RL with low inoculation level of LM; "High:" RL with high inoculation level of LM; "NS:" no significance; "*:" $P < 0.05$.

Chapter 4. Alignment-free microbiota-based classification of fresh produce safety and quality

Chao Liao[a] Luxin Wang[a]* Gerald Quon[b]*

[a] Department of Food Science and Technology, University of California Davis, Davis, CA 95616, USA

[b] Department of Molecular and Cellular Biology, University of California Davis, Davis, CA 95616, USA

*Co-corresponding authors: Luxin Wang, Associate Professor, University of California Davis, Davis, CA 95616, Email: lxwang@ucdavis.edu; Gerald Quon, Assistant Professor, University of California Davis, CA 95616, Email: gquon@ucdavis.edu.

Submitted to Microbiome and posted in BioRxiv: https://doi.org/10.1101/2022.08.25.505309

Abstract

Small samples size and loss of up to 50-70% of sequencing data during the data denoising step of preprocessing can limit the statistical power of fresh produce microbiota analyses and prevent detection of important bacterial species associated with produce contamination or quality reduction. Here, we explored an alignment-free analysis strategy using k-mer hashes to identify DNA signatures predictive of produce safety (contaminated vs. non-contaminated) and produce quality (good-quality vs. decreasing-quality) and compared it against the amplicon sequence variant (ASV) strategy that uses a typical alignment and denoising step. Random forests (RF)-based classifiers were trained on publicly available fresh produce microbiota datasets with data preprocessed using either the k-mer hash or ASV approach. RF-based classifiers for fresh produce safety and quality using 7-mer hash datasets had significantly higher classification accuracy than those using the ASV datasets, supporting the hypothesis that data preprocessing strategies that keep more data (k-mer hash) retain more useful information about bacterial species than approaches that lose data during preprocessing (ASV). Integrating multiple datasets together led to higher classification accuracy compared to those trained with individual datasets. Integrated datasets enabled the identification of more consistent and generalizable indicators (ASV, 7-mer hash, or bacterial taxa) associated with fresh produce safety and quality. The proposed combination of integrating multiple datasets and leveraging an alignment-free 7-mer hash strategy substantially mitigates the loss of sequencing data due to the ASV denoising step and leads to better classification performance for fresh produce safety and quality. Results generated from this study lay the foundation for future studies that wish and need to incorporate and/or compare different microbiota sequencing datasets (generated from different studies or different laboratories) for the application of machine learning in microbial safety and quality of food.

Keywords: Fresh produce, Microbiota, Produce safety and quality, Machine learning, Data integration, k-mer hash.

Introduction

Next generation sequencing approaches for the analysis of microbial communities in fresh produce include amplicon-based sequencing (e.g., 16S rRNA gene sequencing) and metagenomic sequencing (e.g., shotgun sequencing). These technologies help identify microbial populations within individual produce samples, traditionally by aligning sequenced reads to known genomes to quantify the presence of known species in a sample (1). When multiple samples are sequenced, statistical and machine learning approaches can be leveraged to identify species of importance to produce safety (PS) and produce quality (PQ) (2–12). Interactions between pathogenic and/or spoilage microorganisms and other native or background microbiota shine light on new competitive exclusion microorganisms to protect and improve the safety and quality of fresh produce (9, 11–14).

Of particular interest in this study is the use of machine learning to identify indicators: bacterial species whose presence is correlated with produce safety or quality. Indicators can be identified by constructing classification models that identify produce sample features that distinguish samples of different PS or PQ statuses. Both the accuracy of the classifiers, as well as the quality and reproducibility of the indicators identified from them, typically increase with dataset and sample sizes (15). When reviewing published produce microbiota datasets, we found no overlap of PS-related tentative indicators across three PS-related studies (11, 12, 16), and only one indicator consistently observed across three PQ-related studies (9, 11, 16). This poor overlap of indicators suggests an opportunity to identify alternative indicator identification strategies that identify more reproducible indicators across studies.

The poor overlap of indicators identified by different studies may be driven by two reasons. The first reason is due to the limited sample sizes in each individual study. On average, 110 samples were analyzed in previously published studies (1, 4–12, 17), which is much smaller than that suggested given the large number of microbial species being identified in each sample (18). Small sample sizes coupled with the profiling of many bacterial species can lead to more spurious (false) indicators, or correlations between bacterial species occurrence and pathogen contamination or quality decline, and ultimately yield poor reproducibility among studies (11–13). Secondly, this poor overlap might be caused by low effective

sequencing depth of each produce sample. Two key steps of microbiota sequencing data analysis are the alignment of reads to known genomes used in the metagenomic sequencing analysis, and the denoising step used in the amplicon sequencing analysis. During this process, up to 70% of reads for the metagenomic sequencing and 50% of reads for the 16S rRNA gene sequencing can be removed or "wasted" due to their poor matching to the database and nosing sequences (19, 20). This loss of over half of the sequencing reads reduces the power to identify low abundant species or sequence variations (21).

Here we proposed a computational strategy to address both challenges to better identify bacterial indicators broadly correlated with both PS and PQ phenotypes. First, we used an alignment-free approach based on counting short k-mers in sequencing reads (22, 23) to characterize microbiota in individual fresh produce samples (**Figure 1A**). We showed that k-mer hash-based models were significantly better at predicting both PS and PQ compared to the commonly used approach of counting amplicon sequence variants (ASV). Second, we integrated PS and PQ datasets before analysis to boost dataset size and power and show by doing so we are able to identify taxa that are more broadly associated with PS and PQ.

Materials and Methods

Fresh produce microbiota datasets

Four published fresh produce microbiota studies that generated 16S rRNA gene sequencing were selected, as they all have complete metadata information and are associated with PS and/or PQ. These four studies are named Zhang18 (12) (n = 236), Kusstatscher19 (9) (n = 240), LiaoSm21 (11) (n = 72), and LiaoR121 (n = 108) from the project LiaoRl21 (16) (this study) (**Table 1**). Project LiaoRl21 consists of combining published non-contaminated samples (n = 36) (16), with unpublished data from pathogen-contaminated samples (n = 72) that were prepared and sequenced for this paper. The preparation of the contaminated samples, DNA extraction step, as well as library preparation and sequencing followed the protocol published by Liao and Wang (2021) (11). The new data has also been uploaded in the public database with the accession number PRJNA792031. Based on these four studies, we created the PS dataset composed of case-control studies in which fresh produce was artificially inoculated with

Escherichia coli O157:H7 (LiaoSm21 and LiaoRl21), *Listeria monocytogenes* (LiaoRl21), or *Salmonella* Infantis (Zhang18). Produce samples without contamination of pathogens were labeled as non-contaminated, while all other samples were given the uniform label contaminated. Also from these four studies, we created the PQ dataset in which samples were either labeled as good-quality (GQ, samples sequenced before their use-by dates or showing no decaying signs) or decreasing-quality (DQ, samples after their use-by date or decayed) (LiaoSm21, LiaoRl21, and Kusstatscher19).

ASV dataset and k-mer hash dataset preparation

For each dataset, raw 16S rRNA gene sequences were imported into the QIIME 2 (Version 2021.8) pipeline (24) by using the "qiime tools import" command. The barcodes and primers were removed by using "qiime cutadapt trim-paired/single" plugin of cutadapt (25). Reads with median Phred quality scores of less than 30 were removed by truncating reads with a certain length using the "qiime dada2 denoise-paired/single" plugin of DADA2 (26, 27). During this process, raw sequences were filtered, denoised, dereplicated, and clustered into ASV (28). Sequences with barcode and primer sequences removed were then processed for quality control by truncating bases with the median Phred quality scores of less than 30 with the DADA2 package in R without ASV clustering (26, 27). After that, the processed sequences were used to compute k-mer hash signatures (k=3,…,7) for each sample by using the "sourmash sketch dna" command in the sourmash pipeline (version 4.2.3) (22).

Common sum scaling

Since the sequencing depth varies across samples, datasets, and studies, the common sum scaling (COM) method was applied to normalize the sequencing depth among them for both ASV and k-mer hash methods. The ASV and k-mer hash counts were scaled to the minimum depth of each sample with the following equation (29):

$$COM\left(\omega_i{}^{(j)}\right) = \left[\text{round}(\omega_1{}^{(j)}\frac{m^{(min)}}{m^{(j)}}) \ldots, \text{round}(\omega_p{}^{(j)}\frac{m^{(min)}}{m^{(j)}})\right] \epsilon\, \mathbb{R}^{n\times p},$$

where $i = 1, \ldots, p$ is the ASV or k-mer index; $j = 1, \ldots, n$ is the sample index; $\omega_i^{(j)}$ is the read count of ASV or k-mer i in sample j; $m^{(j)}$ is the total ASV or *k*-mer hash count number for sample j, where $m^{(j)} = \sum_{i=1}^{p} \omega_i^j$; $m^{(min)} = min\{m^{(1)}, m^{(2)}, \ldots, m^{(n)}\}$ and round() is an operator rounding the fraction to be the nearest integer.

Data integration and the removal of confounding factors

Data integration by using the Combat function in the 'sva' R package (version 3.42.0) was applied to remove batch effects indicated in the study's metadata, such as sample type, location, and study-of-origin (30). The parameter "par.prior" was set as "false" to use the nonparametric adjustments, and the parameter "mean.only" was set to "true" to adjust the mean of the batch effect across batches (30).

Classification of fresh produce safety and quality samples

Classifiers for PS and PQ using either the processed ASV or k-mer hash datasets were constructed. Three different classification methods, including random forests (RF) (31), *k*-nearest neighbors (*k*-NN) (32), and fully connected, feed-forward neural networks (NN) (33) were evaluated (**Figures S1 and S2**). For RF, the number of decision tree (n_{tree}) was set at 500, and the number of features randomly sampled as candidates at each split (mtry) was set to the square root of the total number of input features. RF-based models were trained by using the "randomForest" R package (34) (version 4.6-14). Predicted class labels were decided based on the majority vote (>50%) by 500 decision trees. The *k*-NN-based classifiers were trained by using the 'caret' R package (version 6.0-90). *k* values ranging from 1 to 10 were tested to identify *k* values with the best classification performance (**Figures S1 and S2**). The feed-forward NN were trained by using the 'nnet' R package (version 7.3-17), with hyperparameters set as the same as Arbajian et al. (2019) (35) and Džal et al. (2021) (36). The 'nnet' fits a

feed-forward NN with a single hidden layer. The number of nodes in the hidden layer was set to 5, the decay parameter was set to 0.1, and the activation function was set to the logistic activation function.

Model validation

For the individual dataset classification experiments, we used 10-fold cross validation to measure the accuracy of RF-, *k*-NN- and NN-based classifiers. Error was measured as the unweighted total classification accuracy across both the case and control phenotypes, as the classification experiments were generally well balanced (**Table S1**). For RF-based models, the out-of-bag (OOB) estimate was also calculated in order to validate the models trained on PS or PQ datasets (31).

Several cross-study classification experiments were also conducted in order to test the generalization performance of PS and PQ classifiers. In the cross-study classification experiments for PS, we constructed pairs of training and test datasets in which one study formed the test dataset and the remaining studies formed the training dataset. For example, for the PS experiments, we constructed a training dataset consisting of the LiaoRl21 and Zhang18 studies, while using the LiaoSm21 study as a test dataset.

Evaluation of feature importance

To evaluate and quantify the importance of individual features identified based on the ASV or the k-mer hash method, features were ranked by their mean decrease in accuracy (MDA) calculated by the RF classification model. The MDA quantifies the importance of a variable by measuring the decrease in prediction accuracy when the variable is randomly permuted in comparison with the original observations (37).

Taxonomic analyses

To profile the bacterial communities within individual produce samples and carry out the differential abundance analysis between phenotypes, taxonomic analyses at the genus level were

conducted by using the QIIME 2 (Version 2021.8) pipeline for the ASV dataset. The QIIME 2 plugin "q2-feature-classifier" (38) was used to assign ASV to the SILVA 138 small subunit rRNA database as the taxonomy reference (39). Chloroplast, mitochondria, and unassigned taxa were filtered by using "qiime taxa filter-seqs" command in QIIME 2 (40). For integrated taxonomic datasets, the COM method was first employed for data normalization. Data integration was carried out by using Combat to regress out study-of-origin, sample types, and storage conditions for PS and PQ datasets. The processed taxonomic datasets were employed for the differential abundance analysis between "contaminated" and "non-contaminated" samples, as well as between GQ and DQ samples.

For taxonomic analysis of 16S rRNA gene sequences based on the k-mer hash datasets, the SILVA 138 database was first downloaded, and the taxonomic sequences were computed for 7-mer hash using the command "sourmash sketch dna" in the sourmash pipeline. The sourmash lowest common ancestor (LCA) taxonomy database was established by using a "sourmash lca index" command on the processed 7-mer hash of taxonomic sequences from the SILVA 138 database, followed by taxonomic classification of query 7-mer hash datasets against the established sourmash LCA taxonomy database carried out by using a command "sourmash lca classify" (22). However, the query k-mer hashes could not be accurately assigned to the sourmash LCA taxonomy database based on SILVA database. This issue may be due to the inappropriateness of the current algorithm of LCA taxonomic classification in the sourmash pipeline for 16S rRNA gene sequences. This is an established issue (https://github.com/sourmash-bio/sourmash/issues/1421). Therefore, the taxonomic analysis here was performed only on the ASV representation of the data.

Data visualization and statistical analysis

The non-parametric Wilcoxon rank sum test was applied to test for significant differences in classification accuracy between pairs of RF-based, *k*-NN-based, and NN-based classifiers. The analysis of compositions of microbiomes with bias correction (ANCOM-BC) test and the RF-based feature selection method mentioned above were employed to identify the bacterial indicators that had significantly

abundance changes in one group over another (41). All the above statistical analyses were conducted in R (version 4.1.0).

Results

Alignment-free analysis leads to more accurate classification of microbiota samples

We first tested the hypothesis that using a k-mer hash strategy to analyze 16S rRNA sequencing data that avoids read alignment and the corresponding data loss in the ASV strategy would improve PS and PQ classification accuracy and indicator identification. **Figure 1A** illustrates the conceptual differences between the two strategies. For each individual PS and PQ study, we constructed machine learning-based classifiers by selecting one of three methods (RF, *k*-NN, and NN), and training them using either the ASV or k-mer hash representations of the study data. Overall, the RF-based classifiers using 7-mer hash (**Figure 1B**) representations showed higher classification accuracy than *k*-NN-based and NN-based for both PS (**Figure S1**) and PQ (**Figure S2**). RF-based classification using the 7-mer hash representation also generally outperformed classification using the ASV representation by 8% on average ($P < 0.05$, Wilcox rank sum test), except for the LiaoSm21_PS and LiaoSm21_PQ studies. In LiaoSm21_PS both the ASV and 7-mer hash versions of the data yielded an accuracy of 100%. We also found the k-mer hash datasets performed better with larger k values within the range of 3 to 7. This is unsurprising because larger k values lead to more features extracted from the sequencing data, which in turn provides more opportunities to distinguish DNA sequence signatures between cases and controls.

Integrated microbiota data analysis leads to more generalizable classification

To increase the effective sample sizes of published microbiota studies and therefore identify taxa that are robustly associated with PS and PQ, a single, integrated produce safety (IPS) dataset and a single, integrated produce quality (IPQ) dataset were established by applying batch correction and merging studies within each category. We computed ASV and 7-mer hash representations of the IPQ and IPS datasets and used them to construct RF-based models whose classification performance were compared

against our previous models constructed on individual studies. Consistent with the analyses conducted on the individual studies, the integrated 7-mer hash representation achieves higher classification accuracy compared to ASV representation by 5% ($P < 0.05$) and 17% ($P < 0.05$) for the IPS and IPQ datasets, respectively (**Figure 2A**), supporting the notion that the 7-mer hash dataset retains more actionable information in sequencing reads. **Figure 2B** illustrates the performance differences separated by component study, and we see that the classification performance is systematically higher in samples from all studies, not just a select study. We also found that for the IPS dataset, when we replaced the binary labels with the pathogen-specific labels (*E. coli* O157, *L. monocytogenes*, and *S.* Infantis), performance similarly was higher for 7-mer representations (85%) compared to ASV (78%) (**Figure S7**). The accuracy of the 7-mer hash IPS classifier was 82%, and when the IPS classifier was used to predict specific pathogens, it rose to 85% (**Figure S7**). Our IPS classifier therefore performs as well as the commonly used culture-dependent approach using selective agars that does not rely on sequencing (42). Given the size and number of PS datasets will increase over time, we expect sequencing-base classifiers to outperform the culture-dependent approach in the future. For the IPQ classifiers that achieved 82% accuracy, there are no benchmarks to compare its performance against as this was the first-time predictive models were used to evaluate the quality of fresh produce using microbiota.

To gain insight into which samples were better classified under the integrated 7-mer hash dataset, we then visualized the label predictions of individual samples generated by the RF-based classifiers**.** **Figures 2C-2F** compare the voting rates based on 7-mer hash and ASV for predicted versus true labels of each sample in IPS and IPQ. For the IPS samples, while the number of correctly predicted contaminated samples is similar (254 and 264 for 7-mer hash and ASV, respectively) (**Figure 2C**), there is a marked increase in accuracy for 7-mer hash predictions of the non-contaminated samples (85 and 52 for 7-mer hash and ASV, respectively) (**Figure 2D**). Similarly, in the analysis of the IPQ samples, the 7-mer hash representation led to more accurate prediction of the decreasing-quality samples (174 and 163 for 7-mer hash and ASV, respectively) (**Figure 2E**), while the number of correctly predicted "good-quality" samples was similar (100 versus 62 for 7-mer hash and ASV, respectively) (**Figure 2F**). Our results

support our hypothesis that the 7-mer hash representation leads to better classification performance of microbiota samples, consistent with our results on individual samples.

We next wondered to what extent integration of microbiota data from multiple studies explicitly led to the construction of more generalizable classifiers, compared to classifiers trained on individual studies. We therefore performed six experiments (three for each of IPS and IPQ), in which we repeatedly removed one study as a test (held out) study, and compared the model performance of RF classifiers when trained on the remaining two studies separately versus combined (**Figure 3**). In principle, classifiers trained on the combined datasets would be encouraged to learn taxa that are more broadly associated with either produce pathogen contamination or produce quality decline, because both the case and control samples in the combined dataset will be more heterogeneous compared to the samples in the individual studies. **Figure 3** shows that across the six experiments, the integrated datasets performed significantly better than the individual datasets alone in three of them (**Figs. 3A**, **3B**, and **3D**), achieving on average 17% higher accuracy than individual datasets. In comparison, for only one experiment, combining data led to significantly worse performance (**Figure 3C**). These results suggest that it can be sensible to combine data from multiple studies together, even if the phenotypes studied do not directly match up and tend to lead to better generalizable performance of the classifiers for both PS and PQ.

Integrated taxonomic analysis identifies generalizable taxa associated with pathogen contamination

The higher classification performance of the integrated datasets suggests that there are taxa that are broadly associated with PS and PQ identified by the classifiers. We therefore first inspected the 7-mer hash features identified as predictive by the classifier, to try to subsequently identify the underlying taxa contributing those predictive DNA signatures. Unfortunately, the current sourmash pipeline that was used to compute the k-mer hash signatures did not support mapping 7-mer hash signatures to the LCA database (see Methods). We therefore instead analyzed the classifiers trained on the ASV representation of the IPS and IPQ datasets for the following taxonomic analysis.

For the IPS taxonomic analysis, 1,131 genera were identified in total. The top 3 most dominant genera in the bacterial communities across fresh produce samples were *Pseudomonas* (with relative abundance ranging from 0.07% to 57.95%), unclassified genera (4.54% to 43.61%), and *Flavobacterium* (1.31% to 41.31%). *Pseudomonas* and *Flavobacterium* as dominant genera in bagged Spring mix salad and lettuce have been reported (11, 43). However, unclassified genera with high relative abundance were not often mentioned in fresh produce microbiota studies. The integration of datasets might bring out more unclassified genera than individual datasets, due to the limited number of taxonomic references available for taxonomic analysis.

The ANCOM-BC test was applied to identify bacteria at the genus level that have significantly different relative abundance between contamination groups (contaminated samples versus non-contaminated samples) or pathogen groups (*E. coli* O157, *L. monocytogenes*, and *S.* Infantis). Five genera were identified as indicators for the contaminated group, including *Escherichia-Shigella*, *Listeria*, *Bacteroides*, *Peredibacter*, and *Faecalibacterium*, and two indicators (*Rheinheimera* and *Pseudomonas*) were identified for the non-contaminated group (**Figure 4A**). Their significance values were listed in **Table S3.** We noticed that the contaminating pathogens, *E. coli* O157 and *L. monocytogenes*, were accurately identified as *Escherichia-Shigella* and *Listeria.* Our previous study (11) reported that no *Escherichia* could be identified in contaminated Spring mix samples when the *E. coli* O157:H7 inoculation level was at 5.5 Log CFU/ml by using 16S rRNA gene sequencing. We? assumed that the concentration of *E. coli* O157 was below the limit of detection of 16S rRNA gene sequencing method. However, the present study showed the *Escherichia-Shigella* was identified for *E. coli* O157 with relative abundance from 0.019% to 0.18%. One explanation for the difference is that the present study conducted the taxonomic classification analysis based on the SILVA 16S rRNA sequence database, while Liao and Wang (2021) applied Greengenes database, which contains less annotation taxonomic references than SILVA database (11).

On the other hand, *S.* Infantis was not identified in the Zhang18 study (12). One explanation is that 10^4 copies/ml of *S.* Infantis located in the root of lettuce was below the limit of detection of 16S

rRNA gene sequencing. In this study, *Peredibacter* was identified as an indicator for *S.* Infantis-contaminated samples although *S.* Infantis could not be detected. *Bacteroides* is an obligate anaerobic bacterium making up a remarkable portion of fecal bacterial communities, which has been suggested as fecal indicator organisms for water samples (44). Our result indicate that *Bacteroides* could be also applied as pathogen contamination indicator for produce. Davi-dov and Jurkevitch (2004) reported that *Peredibacter* is a member of *Bdellovibrio*-and-like organisms (BALO) that are highly motile microbes preying on other Gram-negative bacteria (45). Lu and Cai (2010) reported that *Peredibacter* sp. strain BD2GS significantly retarded the growth of *S.* Typhimurium within 3-12 h through lysing prey cells (46). Based on these, the elevated relative abundance of *Peredibacter* might be triggered by contamination of *S.* Infantis. *Bacteroides* and *Faecalibacterium* have been reported as commensal bacteria of the human gastrointestinal microbiota (47, 48) and classified as fecal indicator bacteria (49). Savichtcheva et al. (2007) reported that 16S rRNA gene marker of *Bacteroides* had a better prediction for the presence of bacterial enteric pathogens than total and fecal coliforms (50).

Integrated taxonomic analysis identifies generalizable taxa associated with produce quality

For the taxonomic analysis at genus level of integrated microbiota datasets related to PQ, 664 genera were identified. The top 3 genera based on the relative abundance were *Pseudomonas* (0.10% to 58.84%), *Flavobacterium* (0.73% to 38.53%), and Unclassified genera (5.49% to 55.52%). Through the ANCOM-BC test, five genera, including *Leuconostoc*, *Gluconobacter*, *Lactobacillus*, *Acetobacter*, and *Clostridium*, were identified as indicators in decreasing-quality samples (**Figure 4B**). Their significance values were listed in **Table S4.** Del Árbol et al. (2016) mentioned that *Gluconobacter* and *Acetobacter* are acetic acid bacteria, which can generate acetic acid to spoil fruits causing bacterial rot and browning (51). *Leuconostoc* and *Lactobacillus* are commonly known as psychrotrophic spoilage lactic acid bacteria that spoil meat products and fresh fruits and vegetables during 4 °C storage (52, 53). *Clostridium* spp. have also been reported to be associated with meat and cheese spoilage (54, 55). In the Kusstatscher et al. (2019) study, except for *Clostridium*, the other four genera identified as indicators in decreasing-quality

samples in the present study were recognized as core microbiota in decaying samples (9). Six genera, including *Sphigomonas*, *Pedobacter*, *Parablastomonas*, *Paracoccus*, *Pir4_lineage*, and *Nocardioides*, were identified as indicators in good-quality samples. These six indicators for good-quality samples were reported to be related to agriculture rhizosphere soil (56–61). Three genera, *Sphigomonas*, *Pedobacter*, and *Nocardioides*, were also identified as core microbiota in healthy samples in the Kusstatscher et al. (2019) study (9).

Most of the indicators identified in the integrated datasets were covered by the indicators identified from individual datasets (**Figures S4** and **S5**), but there were several new indicators identified solely in the integrated dataset. For the PS related datasets, seven indicators identified from the individual datasets (**Figure S4**) were also identified in the integrated dataset. However, seven indicators identified for the contaminated group in the individual datasets were not identified in the integrated dataset, including *Eubacterium*, *Janthinobacterium*, *Serratia*, *Carnobacterium*, *Phascolarctobacterium*, *Brochothrix*, and *Exiguobacterium*. Twelve indicators identified for the non-contaminated group in the individual datasets were not identified in the integrated dataset including *KF.JG30.B3*, *Gemmatimonas*, *Candidatus_Solibacter*, *SWB02*, *Conexibacter*, *Pedomicrobium*, *Bryobacter*, *Reyranella*, *Gaiella*, *Duganella*, *Pedobacter*, and *Pantoea.* (**Figure 4C**).

For PQ datasets, 10 indicators identified in the integrated dataset were covered by the indicators identified in the individual datasets (**Figure S5**). Interestingly, *Paracoccus* was identified as a new indicator for good-quality group only in the integrated dataset. To the best of our knowledge, *Paracoccus* has not been reported to be associated with produce quality. However, this genus contains a number of species that can produce Astaxanthin (62), which has been reported to exhibit antagonism against food spoilage bacteria (63). Fourteen bacteria identified as indicators for the good-quality group in the individual datasets were not identified in the integrated dataset, including *Flavobacterium*, *Ferruginibacter*, *Devosia*, *Skermanella*, *Luteimonas*, *Methylobacterium*, *Novosphingobium*, *Dyadobacter*, *Bacteroides*, *Exiguobacterium*, *Pantoea*, *Eubacterium*, *Escherichia_Shigella*, and *Listeria*. And four

indicators recognized for decreasing-quality group in the individual datasets were not identified in the integrated dataset, including *Brevundimonas*, *Rhizobium*, *Pedobacter*, and *Dyadobacter* (**Figure 4D**).

In addition to identifying critical features associated with PS and PQ using ANCOM-BC test, we ranked the features based on the MDA measures from RF-based PS and PQ classifiers established using individual and integrated ASV or 7-mer hash datasets and classified them into three groups, including negative, zero, and positive contributions to PS and PQ classification (**Figure 5** and **Figure S3**). RF-based classifiers using the 7-mer hash representation utilized on average 62% of the 8192 hash features provided to RF for the integrated datasets, compared to an average of 9% of the ASV features provided to RF for the integrated datasets (**Table S2**). Our interpretation is that the 7-mer hash representation may lead to better classification performance in part because so many features are leveraged in the classification; this may make classification more robust by making individual features weight less towards the label prediction.

Overall, 332 genera were identified from the 2057 ASVs with a positive contribution to the IPS classification. The top 10 most important genera were *Pseudomonas*, *Flavobacterium*, *Rheinheimera*, *Massilia*, *Sphingomonas*, *Enterobacter*, *Serratia*, *Allorhizobium*, *Pantoea*, and *Chryseobacterium* (**Figure 5E**). *Pseudomonas*, *Massilia*, *Enterobacter*, *Serratia*, *Allorhizobium*, and *Pantoea* were considered as the indicators for contaminated samples. Our previous study (11) has reported that *Massilia* was identified as an indicator for *E. coli* O157 contamination of Spring mix salad. Oliveira et al. (2015) have reported that *Enterobacter*, *Serratia*, and *Pantoea* isolated from lettuce presented an inhibitory effect against *Salmonella* and *L. monocytogenes* (64). *Allorhizobium* was not reported previously to be associated with foodborne pathogen contamination of fresh produce, but some species are plant pathogens, e.g. *A. vitis* can infect grapevines (65). Similarly, 138 genera were identified from the 1451 ASVs with positive contributions to the IPQ classification. Among them, the top 10 genera were *Pseudomonas*, *Flavobacterium*, *Serratia*, *Massilia*, *Pantoea*, *Rheinheimera*, *Chryseobacterium*, *Pedobacter*, *Sphingomonas*, and *Leuconostoc* (**Figure 5F**). *Flavobacterium*, *Serratia*, *Chryseobacterium*, and *Leuconostoc* were regarded as indicators for the decreasing-quality samples. *Flavobacterium* has been

reported to be associated with the spoilage of meat, milk, and seafood (66). Hernández et al. (2020) reported *Flavobacterium* and *Chryseobacterium* belong to the class Flavobacteriales were dominant during the ready-to-eat lettuce spoilage but their importance during vegetable spoilage was still unclear (43). Our results filled the knowledge gap to present the importance of *Flavobacterium* and *Chryseobacterium* during the quality reduction of fresh produce. In addition, it has been widely known that *Serratia* and *Leuconostoc* are microorganism that can cause food spoilage (66). In addition to these four indicators, the RF-selected features contained another 41 indicators for decreasing-quality produce.

Discussion

The integrated PS and PQ classifiers using 7-mer hash datasets had significantly higher accuracy than the models using ASV datasets. Two reasons are proposed: first, to obtain the ASV representations, the DADA2 plugin in QIIME 2 was used to process the primer-removed sequences, including quality filtering, denoising, chimera removing, dereplicating, and pair-end sequence joining. Although the parameter "--p-trunc-len" was set as 0 to trim no base due to the median quality score greater than 30 in this study, 1.66% to 40.94% of sequences from the integrated dataset related to PS were still discarded (**Figure S6A**) and 0.50% to 46.98% of overall sequences from integrated dataset related to PQ were discarded (**Figure S6B**). For the k-mer hash representations, the raw sequences were directly used to generate the k-mer hash datasets by using the sourmash pipeline as the all the median phred quality scores of nucleotides in reads were greater than 30 (27). The discarded reads may contain abundant sequence variation information, which contributed to the downstream sample classification. Werner et al. (2012) reported that 43% of total 16S rRNA gene sequences were discarded after the denoising process (21). Second, the 7-mer hash computed from the original reads could improve the detection sensitivity of bases than ASV in a longer size (600 bp). The raw sequences were sub-sequenced into 7-base subsequences (7-mer), then transformed into hash codes, and finally randomly picked for 8,192 of 7-mer hash signatures for each sample (22). The 7-mer hash datasets for establishing classification models can significantly

shorten the computing time and improve the specificity and sensitivity of detection of different nucleotides among sequences compared with ASV.

The 3-mer to 7-mer hash datasets of individual studies related to PS and PQ were computed by using the sourmash pipeline, which were used subsequently to establish PS and PQ classifiers. Generally, the RF-based models using 6-mer and 7-mer hash datasets had similar accuracy but significantly higher accuracy than that of models using 3-mer to 5-mer hash datasets, except for PQ models using 3-mer of LiaoRl21. Sourmash pipeline generates k-mer minhash signatures of DNA sequences, which randomly samples k-mer content to produce small subsets as known as "sketches" (22). The Jaccard similarity of two sketches of sequence data sets remains approximately equal to their true Jaccard similarity (67). The factor "scaled" was set as 1 in 'sourmash sketch dna' command to generate the hash signature without down sampling the number of sketches. Using 'sourmash sketch dna' command with the k-mer size parameter set as 3 to 7 generated 32, 136, 512, 2080, and 8192 k-mer hashes of each DNA sample, respectively. Based on this, 7-mer hashes with a larger size than 3-mer to 6-mer contain more sequence variances, which contributes to the higher accuracy of models (68). In addition, a larger k size of k-mer improves the specificity of different bases among sequences (69). However, when k was set greater than 7, e.g., when k = 8, there were 32,896 of 8-mer hashes computed. The computing time and storage space remarkably increased while the accuracy of models had no obvious improvement (data not shown).

In the previous PS and PQ studies, a few bacterial indicators for produce contamination (11) or decaying produce (9) have been identified. However, these indicators were not consistent across the individual datasets due to the limited data size. In this study, we applied the data integration method to homogenize three PS datasets and three PQ datasets, respectively. The ANCOM-BC test was then employed to identify indicators for pathogen contamination and quality reduction of fresh produce based on individual datasets and integrated datasets. We identified 26 and 28 genus indicators for PS and PQ produce, respectively, from individual datasets. Among them, seven indicators related to PS and 10 indicators related to PQ were validated by ANCOM-BC test using the integrated datasets. These validated indicators can provide a more generalizable and consistent indication or prediction of PS or PQ statuses

(70). Interestingly, we also identified a new indicator, *Paracoccus*, for good-quality produce only from the integrated PQ dataset. The result indicates that more new indicators could be potentially identified with the size increase of integrated datasets. The non-validated indicators from individual datasets may be due to these studies having different types of fresh produce, distinct inoculated pathogens, and/or various storage conditions, which make the composition and diversity of bacterial communities largely different.

In addition to the ANCOM-BC test, the RF feature selection method was also used to identify indicators for fresh produce contamination and quality reduction. Compared to the indicators identified by the ANCOM-BC test, we found RF-based feature selection method can identify many more indicators for PS (332 genera) and PQ (138 genera) with a positive contribution to the PS and PQ classification samples. These features covered all the indicators identified by ANCOM-BC test, indicating the RF-based models are more sensitive and powerful to catch the variation of features between classification groups. RF-based model and ANCOM-BC test can be used together to determine the reliable indicators for indicating PS and PQ. Sheh et al. (2022) reported that RF-based models were the most accurate models and correctly classified strictures indicators for chronic gastrointestinal diseases using 9 ASVs (71). Based on the RF-based model and ANCOM results, *Clostridium perfringens* was identified as a potential causative agent associated with the development of strictures.

Conclusion

In summary, we established and compared RF-based PS and PQ classifiers by using publicly available microbiota datasets in ASV and 7-mer hash representations for predicting the contamination conditions or quality statuses. This study illustrates that the alignment-free 7-mer hash-based approaches are useful for building more accurate classifiers than ASV, but not necessarily for the taxonomic analysis yet. Data integration of multiple datasets leads to greater classification performance of the integrated RF-based models than that using individual datasets, with significantly higher accuracy and more features with positive contribution to PS or PQ classification identified. In addition, we found more consistent and

generalizable microbes were identified as indicators for safety and quality groups of fresh produce through integrated taxonomic analysis, illustrating the benefits of integrating datasets.

Acknowledgements

We would like to acknowledge Dr. C. Titus Brown for the advice on using sourmash pipeline.

Availability of data and materials

Raw reads from this study have been deposited to the National Center for Biotechnology Information under the project accession number PRJNA792031. The sample datasets are available in the GitHub repository (https://github.com/LZC0034/CeDAR-Project). Additional datasets that support the findings of this study are available from the corresponding authors upon request. The codes for data processing are also available in the GitHub repository above.

References

1. Jackson C, Stone B, Tyler H. 2015. Emerging perspectives on the natural microbiome of fresh produce vegetables. Agriculture 5:170–187.
2. Bergholz TM, Moreno Switt AI, Wiedmann M. 2014. Omics approaches in food safety: fulfilling the promise? Trends Microbiol 22:275–281.
3. Ceuppens S, Delbeke S, De Coninck D, Boussemaere J, Boon N, Uyttendaele M. 2015. Characterization of the bacterial community naturally present on commercially grown basil leaves: Evaluation of sample preparation prior to culture-independent techniques. Int J Environ Res Public Health 12:10171–10197.
4. Gu G, Ottesen A, Bolten S, Ramachandran P, Reed E, Rideout S, Luo Y, Patel J, Brown E, Nou X. 2018. Shifts in spinach microbial communities after chlorine washing and storage at compliant and abusive temperatures. Food Microbiol 73:73–84.
5. 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.
6. Jarvis KG, White JR, Grim CJ, Ewing L, Ottesen AR, Beaubrun JJ-G, Pettengill JB, Brown E, Hanes DE. 2015. Cilantro microbiome before and after nonselective pre-enrichment for *Salmonella* using 16S rRNA and metagenomic sequencing. BMC Microbiol 15:160.
7. Jarvis KG, Daquigan N, White JR, Morin PM, Howard LM, Manetas JE, Ottesen A, Ramachandran P, Grim CJ. 2018. Microbiomes associated with foods from plant and animal sources. Front Microbiol 9:2540.
8. Keshri J, Krouptiski Y, Abu-Fani L, Achmon Y, Bauer TS, Zarka O, Maler I, Pinto R, Sela Saldinger S. 2019. Dynamics of bacterial communities in alfalfa and mung bean sprouts during refrigerated conditions. Food Microbiol 84:103261.

9. Kusstatscher P, Zachow C, Harms K, Maier J, Eigner H, Berg G, Cernava T. 2019. Microbiome-driven identification of microbial indicators for postharvest diseases of sugar beets. Microbiome 7:112.
10. Leff JW, Fierer N. 2013. Bacterial communities associated with the surfaces of fresh fruits and vegetables. PLoS One 8:e59310.
11. Liao C, Wang L. 2021. Evaluation of the bacterial populations present in Spring Mix salad and their impact on the behavior of *Escherichia coli* O157:H7. Food Control 107865.
12. Zhang Y, Jewett C, Gilley J, Bartelt-Hunt SL, Snow DD, Hodges L, Li X. 2018. Microbial communities in the rhizosphere and the root of lettuce as affected by *Salmonella*-contaminated irrigation water. FEMS Microbiol Ecol 94.
13. Söderqvist K, Ahmed Osman O, Wolff C, Bertilsson S, Vågsholm I, Boqvist S. 2017. Emerging microbiota during cold storage and temperature abuse of ready-to-eat salad. Infect Ecol Epidemiol 7:1328963.
14. Yurgel SN, Abbey, Lord, Loomer N, Gillis-Madden R, Mammoliti M. 2018. Microbial Communities Associated with Storage Onion. Phytobiomes Journal 2:35–41.
15. Costello Z, Martin HG. 2018. A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data. NPJ Syst Biol Appl 4:19.
16. Liao C, Wang L. 2022. The microbial quality of commercial chopped romaine lettuce before and after the “use by” date. Front Microbiol 13:850720.
17. 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.
18. van der Ploeg T, Austin PC, Steyerberg EW. 2014. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. BMC Med Res Methodol 14:137.
19. Maran MIJ, Davis G DJ. 2022. Benefits of merging paired-end reads before pre-processing environmental metagenomics data. Mar Genomics 61:100914.

20. Prodan A, Tremaroli V, Brolin H, Zwinderman AH, Nieuwdorp M, Levin E. 2020. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. PLoS One 15:e0227434.
21. Werner JJ, Zhou D, Caporaso JG, Knight R, Angenent LT. 2012. Comparison of Illumina paired-end and single-direction sequencing for microbial 16S rRNA gene amplicon surveys. ISME J 6:1273–1276.
22. Pierce NT, Irber L, Reiter T, Brooks P, Brown CT. 2019. Large-scale sequence comparisons with sourmash. [version 1; peer review: 2 approved]. F1000Res 8:1006.
23. Vinje H, Liland KH, Almøy T, Snipen L. 2015. Comparing K-mer based methods for improved classification of 16S sequences. BMC Bioinformatics 16:205.
24. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu Y-X, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37:852–857.

25. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet j 17:10.

26. 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.

27. Estaki M, Jiang L, Bokulich NA, McDonald D, González A, Kosciolek T, Martino C, Zhu Q, Birmingham A, Vázquez-Baeza Y, Dillon MR, Bolyen E, Caporaso JG, Knight R. 2020. QIIME 2 enables comprehensive end-to-end analysis of diverse microbiome data and comparative studies with publicly available data. Curr Protoc Bioinformatics 70:e100.

28. 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.

29. Badri M, Kurtz ZD, Bonneau R, Müller CL. 2020. Shrinkage improves estimation of microbial associations under different normalization methods. NAR Genom Bioinform 2:lqaa100.

30. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 28:882–883.

31. Breiman L. 2001. Random forests. Machine learning 45:5-32.32. Altman NS. 1992. An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat 46:175–185.

33. Suykens JAK, Vandewalle J. 1999. Least squares support vector machine classifiers. Neural process lett 9:293-300.34. Liaw A, Wiener M. 2002. Classification and regression by randomForest. R news 2:18-22.

35. Arbajian P, Hajja A, Raś ZW, Wieczorkowska AA. 2019. Effect of speech segment samples selection in stutter block detection and remediation. J Intell Inf Syst 1–24.

36. Džal D, Kosović IN, Mastelić T, Ivanković D, Puljak T, Jozić S. 2021. Modelling bathing water quality using official monitoring data. Water (Basel) 13:3005.

37. Calle ML, Urrea V. 2011. Letter to the editor: Stability of Random Forest importance measures. Brief Bioinform 12:86–89.

38. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA, Gregory Caporaso J. 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 6:90.

39. 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.

40. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J 6:610–618.

41. Lin H, Peddada SD. 2020. Analysis of compositions of microbiomes with bias correction. Nat Commun 11:3514.

42. Park S-H, Chang P-S, Ryu S, Kang D-H. 2014. Development of a novel selective and differential medium for the isolation of *Listeria monocytogenes*. Appl Environ Microbiol 80:1020–1025.

43. Hernandez I, Alfaro B. 2020. Enhancing high throughput sequencing unveils changes in bacterial communities during ready-to-eat lettuce spoilage. J Hortic Postharvest Res 3:297-310. 44. Tsai K, Hoffmann V, Simiyu S, Cumming O, Borsay G, Baker KK. 2021. Bacteroides microbial source tracking markers perform poorly in predicting Enterobacteriaceae and Enteric pathogen contamination of cow milk products and milk-containing infant food. Front Microbiol 12:778921.

45. Davidov Y, Jurkevitch E. 2004. Diversity and evolution of *Bdellovibrio*-and-like organisms (BALOs), reclassification of *Bacteriovorax starrii* as *Peredibacter starrii* gen. nov., comb. nov., and description of the *Bacteriovorax-Peredibacter* clade as Bacteriovoracaceae fam. nov. Int J Syst Evol Microbiol 54:1439–1452.

46. Lu F, Cai J. 2010. The protective effect of *Bdellovibrio*-and-like organisms (BALO) on tilapia fish fillets against *Salmonella enterica* ssp. *enterica* serovar Typhimurium. Lett Appl Microbiol 51:625–631.

47. Khan MT, Duncan SH, Stams AJM, van Dijl JM, Flint HJ, Harmsen HJM. 2012. The gut anaerobe *Faecalibacterium prausnitzii* uses an extracellular electron shuttle to grow at oxic-anoxic interphases. ISME J 6:1578–1585.

48. Wexler HM. 2007. *Bacteroides*: the good, the bad, and the nitty-gritty. Clin Microbiol Rev 20:593–621.

49. Zhao X-L, Qi Z, Huang H, Tu J, Song X-J, Qi K-Z, Shao Y. 2022. Coexistence of antibiotic resistance genes, fecal bacteria, and potential pathogens in anthropogenically impacted water. Environ Sci Pollut Res Int 29:46977–46990.

50. Savichtcheva O, Okayama N, Okabe S. 2007. Relationships between Bacteroides 16S rRNA genetic markers and presence of bacterial enteric pathogens and conventional fecal indicators. Water Res 41:3615–3628.

51. Toledo Del Árbol J, Pérez Pulido R, La Storia A, Grande Burgos MJ, Lucas R, Ercolini D, Gálvez A. 2016. Microbial diversity in pitted sweet cherries (*Prunus avium* L.) as affected by High-Hydrostatic Pressure treatment. Food Res Int 89:790–796.

52. Andreevskaya M, Jääskeläinen E, Johansson P, Ylinen A, Paulin L, Björkroth J, Auvinen P. 2018. Food spoilage-associated *Leuconostoc*, *Lactococcus*, and *Lactobacillus* Species display different survival strategies in response to competition. Appl Environ Microbiol 84.

53. Barth M, Hankinson TR, Zhuang H, Breidt F. 2009. Microbiological spoilage of fruits and vegetables, p. 135–183. *In* Sperber, WH, Doyle, MP (eds.), Compendium of the microbiological spoilage of foods and beverages. Springer New York, New York, NY.

54. Gómez-Torres N, Dunne M, Garde S, Meijers R, Narbad A, Ávila M, Mayer MJ. 2018. Development of a specific fluorescent phage endolysin for in situ detection of *Clostridium* species associated with cheese spoilage. Microb Biotechnol 11:332–345.

55. Palevich N, Palevich FP, Maclean PH, Altermann E, Gardner A, Burgess S, Mills J, Brightwell G. 2021. Comparative genomics of *Clostridium* species associated with vacuum-packed meat spoilage. Food Microbiol 95:103687.

56. Asaf S, Numan M, Khan AL, Al-Harrasi A. 2020. *Sphingomonas*: from diversity and genomics to functional role in environmental remediation and plant growth. Crit Rev Biotechnol 40:138–152.

57. Fagervold SK, Bessette S, Romano C, Martin D, Plyuscheva M, Le Bris N, Galand PE. 2013. Microbial communities associated with the degradation of oak wood in the Blanes submarine canyon and its adjacent open slope (NW Mediterranean). Prog Oceanogr 118:137–143.

58. Kwon S-W, Son J-A, Kim S-J, Kim Y-S, Park I-C, Bok J-I, Weon H-Y. 2011. *Pedobacter rhizosphaerae* sp. nov. and *Pedobacter soli* sp. nov., isolated from rhizosphere soil of Chinese cabbage (*Brassica campestris*). Int J Syst Evol Microbiol 61:2874–2879.

59. Liu X, Ge W, Zhang X, Chai C, Wu J, Xiang D, Chen X. 2019. Biodegradation of aged polycyclic aromatic hydrocarbons in agricultural soil by *Paracoccus* sp. LXC combined with humic acid and spent mushroom substrate. J Hazard Mater 379:120820.

60. Shi J, Gong X, Rahman MK u, Tian Q, Zhou X, Wu F. 2021. Effects of wheat root exudates on bacterial communities in the rhizosphere of watermelon. Plant Soil Environ 67:721–728.

61. Takagi K, Fujii K, Yamazaki K, Harada N, Iwasaki A. 2012. Biodegradation of melamine and its hydroxy derivatives by a bacterial consortium containing a novel *Nocardioides* species. Appl Microbiol Biotechnol 94:1647–1656.

62. Fang N, Wang C, Liu X, Zhao X, Liu Y, Liu X, Du Y, Zhang Z, Zhang H. 2019. De novo synthesis of astaxanthin: From organisms to genes. Trends Food Sci Technol 92:162–171.

63. Mageswari A, Subramanian P, Srinivasan R, Karthikeyan S, Gothandam KM. 2015. Astaxanthin from psychrotrophic *Sphingomonas faeni* exhibits antagonism against food-spoilage bacteria at low temperatures. Microbiol Res 179:38–44.

64. Oliveira M, Abadias M, Colás-Medà P, Usall J, Viñas I. 2015. Biopreservative methods to control the growth of foodborne pathogens on fresh-cut lettuce. Int J Food Microbiol 214:4–11.

65. Habbadi K, Meyer T, Vial L, Gaillard V, Benkirane R, Benbouazza A, Kerzaon I, Achbani EH, Lavire C. 2018. Essential oils of *Origanum compactum* and *Thymus vulgaris* exert a protective effect against the phytopathogen *Allorhizobium vitis*. Environ Sci Pollut Res Int 25:29943–29952.

66. Odeyemi OA, Alegbeleye OO, Strateva M, Stratev D. 2020. Understanding spoilage microbial community and spoilage mechanisms in foods of animal origin. Comp Rev Food Sci Food Safety 19:311–331.

67. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol 17:132.

68. Martínez-Porchas M, Vargas-Albores F. 2017. Microbial metagenomics in aquaculture: a potential tool for a deeper insight into the activity. Rev Aquacult 9:42–56.

69. Nasko DJ, Koren S, Phillippy AM, Treangen TJ. 2018. RefSeq database growth influences the accuracy of k-mer-based lowest common ancestor species identification. Genome Biol 19:165.

70. Kim Y, Bismeijer T, Zwart W, Wessels LFA, Vis DJ. 2019. Genomic data integration by WON-PARAFAC identifies interpretable factors for predicting drug-sensitivity in vivo. Nat Commun 10:5034.

71. Sheh A, Artim SC, Burns MA, Molina-Mora JA, Lee MA, Dzink-Fox J, Muthupalani S, Fox JG. 2022. Alterations in common marmoset gut microbiome associated with duodenal strictures. Sci Rep 12:5277.

Table 1. Characteristics of fresh produce microbiota datasets

Dataset	Produce types	Sequencing Instrument	Sequencing Region	Factors	Sample size	References	Database (Access ID)
1	Lettuce	Illumina MiSeq	515F/ 806R (V4)	● Contamination treatment ● Sample type ● Soil texture ● Cultivar ● Harvest time	236	(12)	NCBI (PRJNA289142)
2	Spring Mix salad	Illumina MiSeq	314F/785R (V3-V4)	● Contamination treatment ● Quality ● Brand	72	(11)	EMBL-EBI (ERP112563)
3	Romaine Lettuce	Illumina MiSeq	314F/785R (V3-V4)	● Contamination treatment ● Quality ● Brand ● Season	108	(16) and this study	NCBI (PRJNA792031)
4	Sugar beet	Illumina MiSeq	515F/ 926R (V4-V5)	● Quality ● Location ● Clamp type ● Amplicon type	240	(9)	NCBI (PRJEB28964)

NCBI: National Center for Biotechnology Information; EMBL-EBI: European Molecular Biology Laboratory's European Bioinformatics Institute

Table S1. The number of case and control samples in each fresh produce microbiota dataset

Dataset	Class	Total sample number	Case number (Cont/DQ)	Control number (Ctrl/GQ)
Zhang18	PS	236	158	78
LiaoSm21		72	36	36
LiaoRl21		108	72	36
Kusstatscher19	PQ	240	160	80
LiaoSm21		72	45	27
LiaoRl21		108	54	54

PS and PQ represent produce safety and produce quality; Cont and Ctrl stand for contaminated group and non-contaminated group; GQ and DQ mean good-quality group and decreasing-quality group.

Table S2. The number of features with a positive or negative contribution to the PS and PQ classification based on mean decrease accuracy provided by RF-based models.

Dataset	Class	Total number of ASV	Number of positive features (ASV)	Number of negative features (ASV)	Total number of 7-mer hash	Number of positive features (7-mer hash)	Number of negative features (7-mer hash)
Zhang18	PS	49404	2029	1253	8192	2665	1506
LiaoSm21		2510	401	46	8192	349	0
LiaoRl21		2365	228	276	8192	1114	543
IPS		68007	2057	1179	8192	4084	1408
Kusstatscher19	PQ	1164	210	66	8192	2070	603
LiaoSm21		2510	354	222	8192	908	448
LiaoRl21		2365	278	238	8192	1469	654
IPQ		18732	1451	918	8192	3403	1271

PS and PQ represent produce safety and produce quality; IPS and IPQ stand for integrated produce safety and integrated produce quality. ASV means amplicon sequence variant.

Table S3. *P* values of the indicators identified by the ANCOM-BC analysis for fresh produce in different contamination and pathogen groups.

Indicator	**Group**	***P* value**
Rheinheimera	Ctrl (C)	1.19×10^{-5}
Pseudomonas	Ctrl (C)	8.41×10^{-3}
Escherichia-Shigella	Cont	1.98×10^{-4}
Listeria	Cont	1.53×10^{-3}
Bacteroides	Cont	1.57×10^{-3}
Peredibacter	Cont	6.69×10^{-3}
Faecalibacterium	Cont	0.048
Rheinheimera	Ctrl (P)	3.81×10^{-7}
Pedobacter	Ctrl (P)	1.51×10^{-5}
Duganella	Ctrl (P)	5.15×10^{-4}
Pseudomonas	Ctrl (P)	5.16×10^{-3}
Listeria	*L. moncytogenes*	1.39×10^{-10}
Escherichia-Shigella	*E. coli* O157:H7	6.97×10^{-10}
Faecalibacterium	*E. coli* O157:H7	3.56×10^{-6}
Bacteroides	*E. coli* O157:H7	4.34×10^{-5}
Eubacterium	*E. coli* O157:H7	2.57×10^{-3}
Ruminococcus	*E. coli* O157:H7	0.019
Sanguibacter	*E. coli* O157:H7	0.032
Peredibacter	*Salmonella* Infantis	3.37×10^{-3}

Ctrl (C) represents the non-contaminated samples under the contamination groups; Ctrl (P) stands for the non-pathogenic samples under the pathogen groups.

Table S4. *P* values of the indicators identified by the ANCOM-BC analysis for fresh produce in different quality groups.

Indicator	**Group**	***P* value**
Sphigomonas	GQ	4.54×10^{-9}
Pedobacter	GQ	3.16×10^{-4}
Parablastomonas	GQ	2.31×10^{-3}
Paracoccus	GQ	7.09×10^{-3}
Pir4_lineage	GQ	0.010
Nocardioides	GQ	0.048
Leuconostoc	DQ	8.23×10^{-9}
Gluconobacter	DQ	1.38×10^{-7}
Lactobacillus	DQ	1.16×10^{-4}
Acetobacter	DQ	5.8×10^{-4}
Clostridium	DQ	0.015

GQ represents good quality; DQ means decreasing quality.

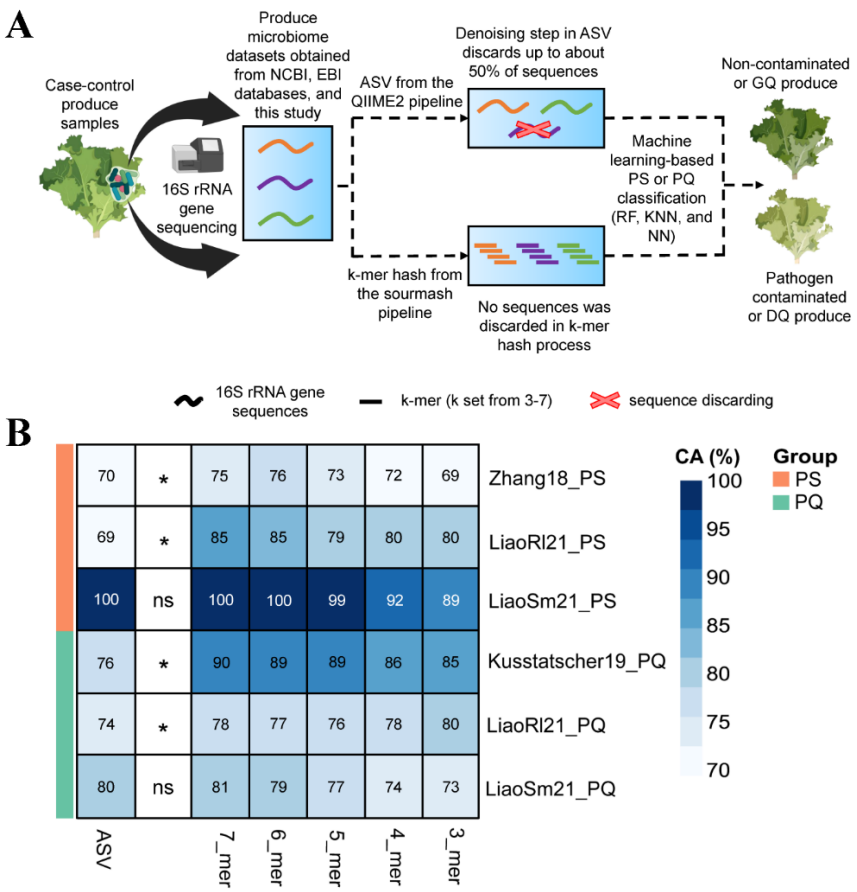

Figure 1. (A) Schematic comparing the preprocessing including denoising of the ASV approach against the alignment-free preprocessing of the k-mer hash approach for constructing produce safety (PS) and produce quality (PQ) classifiers. (B) Heatmap of the accuracies achieved by RF-based models using fresh produce microbiota datasets associated with PS and PQ in ASV format and k-mer hash format (k from 3 to 7). Accuracies are measured using 10-fold cross-validation. GQ and DQ represent good-quality produce and decreasing-quality produce, respectively. Partial icons in (A) were obtained from "BioRender.com".

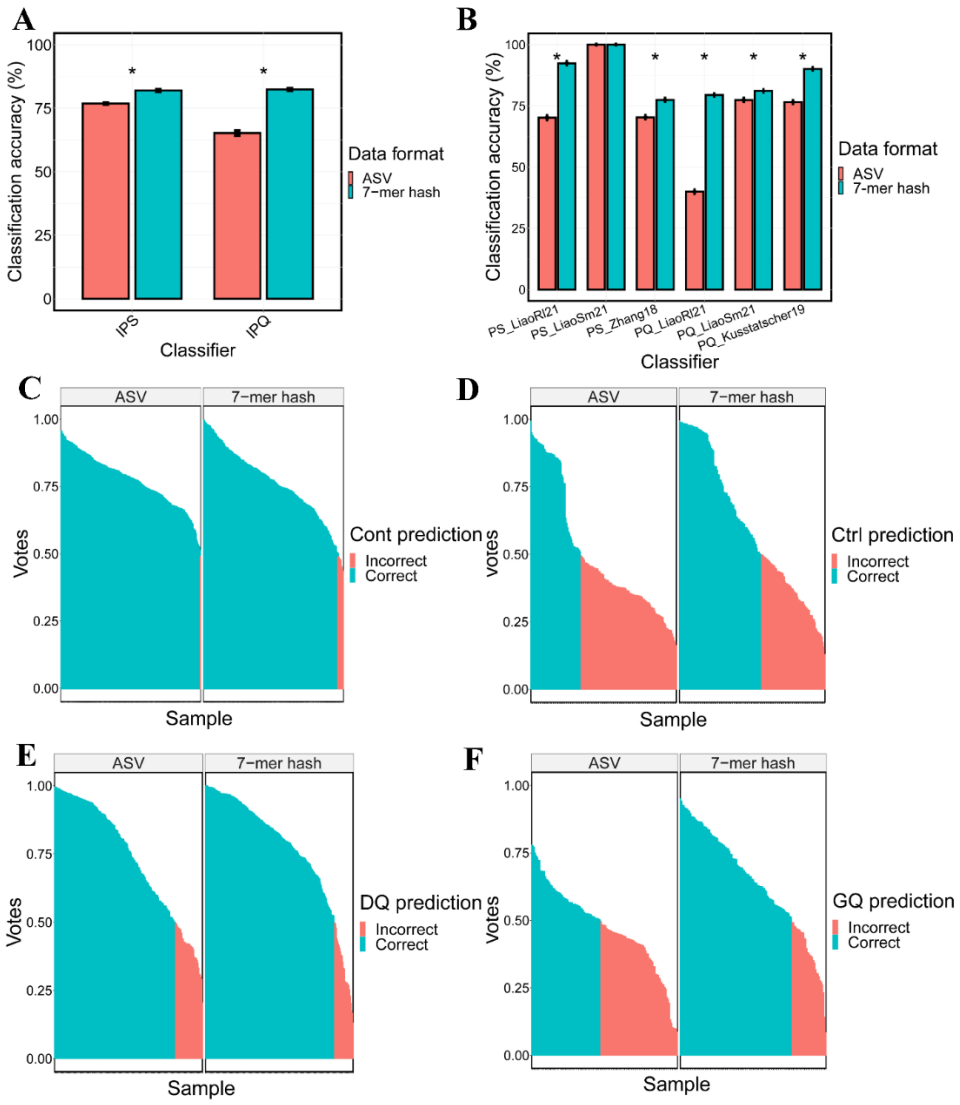

Figure 2. (A) Barplot of classification performance of RF-based models on the integrated PS (IPS) and integrated PQ (IPQ) datasets, represented using either ASV or 7-mer hashes. The Wilcoxon rank sum test was used for the pairwise comparison of the accuracies of the IPS and IPQ classifiers. (B) Barplot of classification performance of classifiers established by using individual datasets from IPS and IPQ datasets in ASV and 7-mer hash. (C) Barplot of the prediction of PS samples with the true label of contamination. The prediction was made by votes of 500 decision trees in RF-based classifiers established by using ASV and 7-mer hash. The cutoff voting rate (50% votes) indicates whether a labeled sample is predicted correctly or not. Ctrl and Cont represent non-contaminated samples and contaminated samples.

The * stands for $P < 0.05$. (D) Same as (C), but barplot of the prediction of PS samples with the true label of control. (E) Same as (C), but barplot of the prediction of PQ samples with the true label of decreasing quality (DQ). (F) Same as (C), but barplot of the prediction of PQ samples with the true label of good quality (GQ).

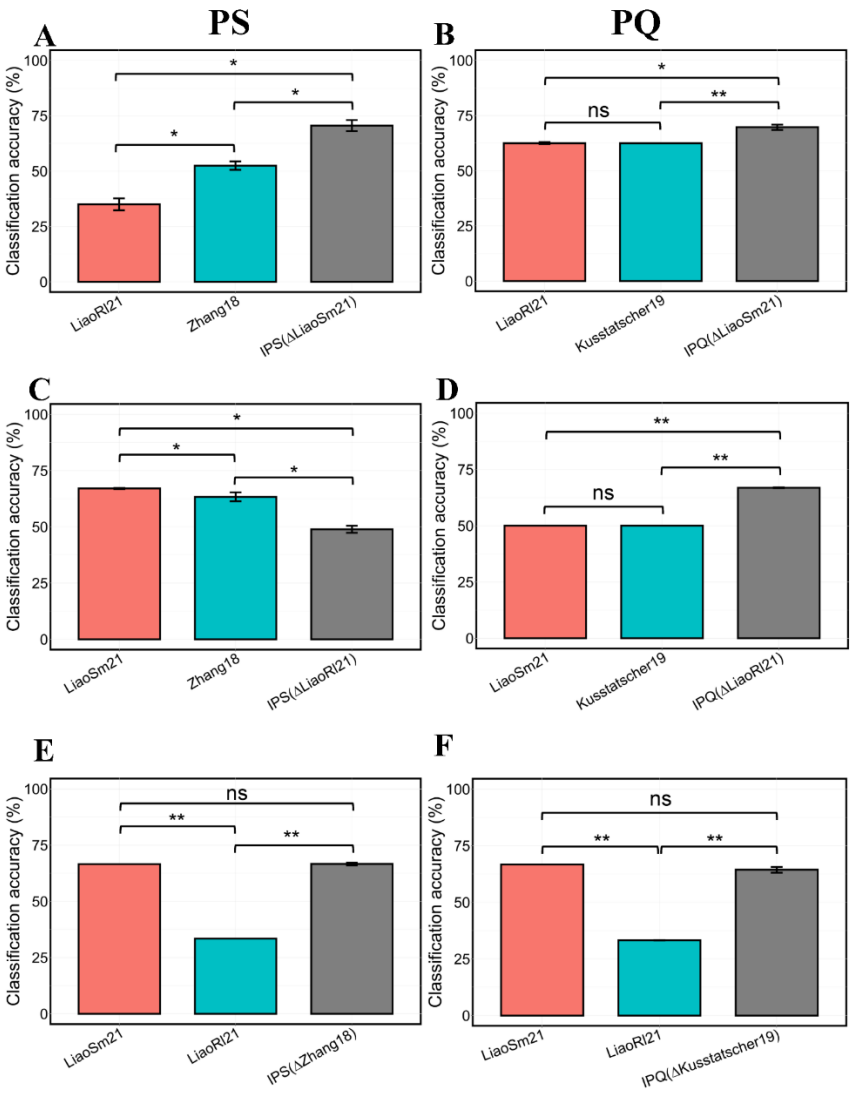

Figure 3. Barplots comparing the classification performance of classifiers trained on individual PS and PQ datasets, and IPS and IPQ datasets. (A) Classifiers were trained on individual LiaoRl21 or Zhang18 datasets and a classifier was trained on an integrated dataset excluding LiaoSm21, denoting IPS (ΔLiaoSm21). The LiaoSm21 was used as a testing set. (B) Classifiers were trained on individual LiaoRl21 and Kusstatscher datasets and an integrated dataset excluding LiaoSm21, denoting as IPQ (ΔLiaoSm21). The LiaoSm21 was used as a testing set. (C) Same as (B), but individual and integrated PS classifiers were tested on LiaoRl21. (D) Same as (B), but individual and integrated PQ classifiers were tested on LiaoRl21. (E) Same as (B), but individual and integrated PS classifiers were tested on Zhang18. (F) Same as (B), but individual and integrated PQ classifiers were tested on Kusstatscher19. Wilcoxon

rank sum tests were applied for testing the significance of the difference in testing accuracy (%) of individual and integrated PS and PQ classifiers. PS and PQ mean produce safety and produce quality, respectively. The * represents $P < 0.05$; The ** stands for $P < 0.01$; The ns represents no significance.

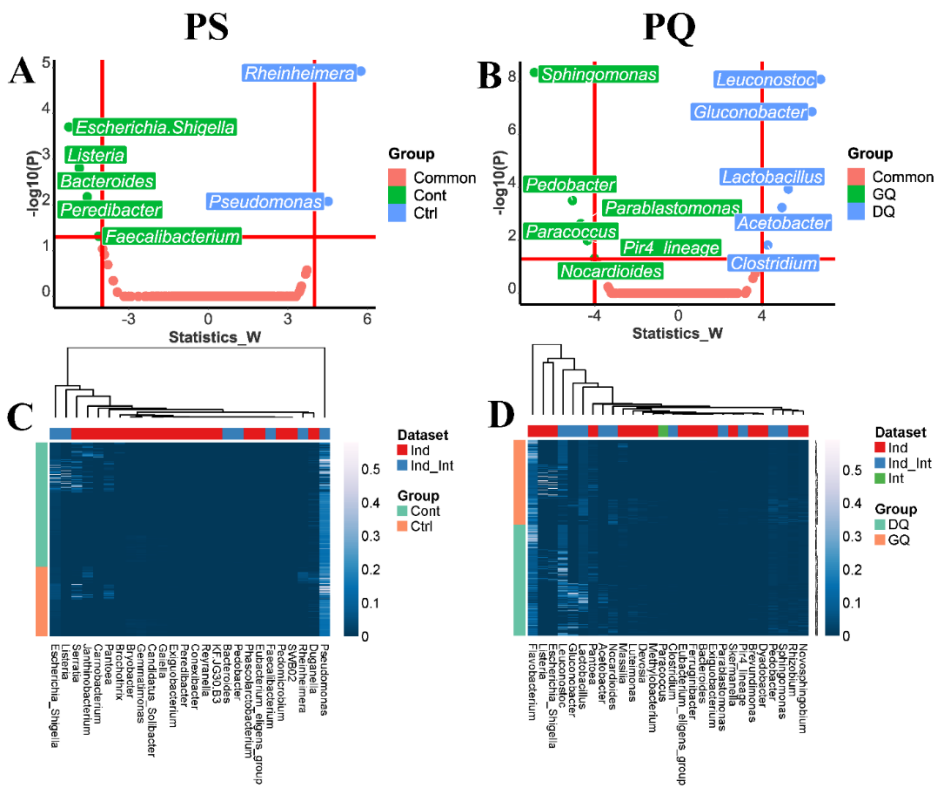

Figure 4. (A) Volcano plots based on the W statistic values and -log10(P) values obtained from the ANCOM-BC test presenting the differential abundances of genera among two contamination groups for PS. (B) Volcano plots based on the W statistic values and -log10(P) values obtained from the ANCOM-BC test presenting the differential abundances of genera among two quality groups for PQ. (C) Heatmap of the relative abundances of bacterial indicators identified from individual and integrated PS datasets. (D) Heatmap of the relative abundances of bacterial indicators identified from individual and integrated PQ datasets. Ind means individual datasets. Int represents the integrated dataset. Cont means contaminated samples. Ctrl represents non-contaminated samples. DQ represents decreasing-quality samples. GQ represents good-quality samples.

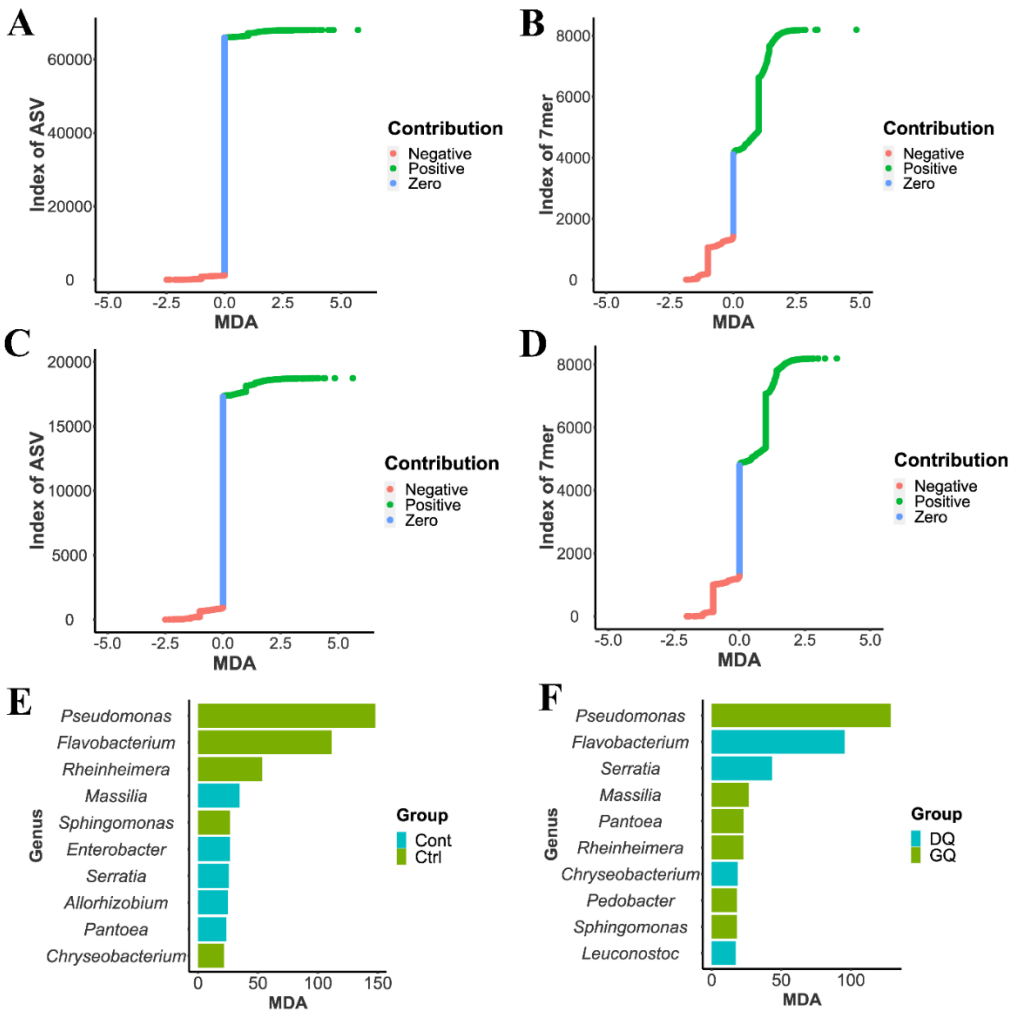

Figure 5. Importance of features evaluated by mean decrease accuracy (MDA) provided by random forests-based classifiers. (A) Contribution of ASV features to PS classification identified from RF-based IPS classifiers. (B) Contribution of 7-mer hash features to PS classification identified from RF-based IPS classifiers. (C) Same as (A), but for PQ. (D) Same as (B), but for PQ. (E) Identified top 10 genera with a positive contribution to IPS classification. (F) Identified top 10 genera with a positive contribution to IPQ classification. Ctrl represents the non-contaminated group and Cont represents the contaminated group. GQ represents good quality and DQ represents decreasing quality.

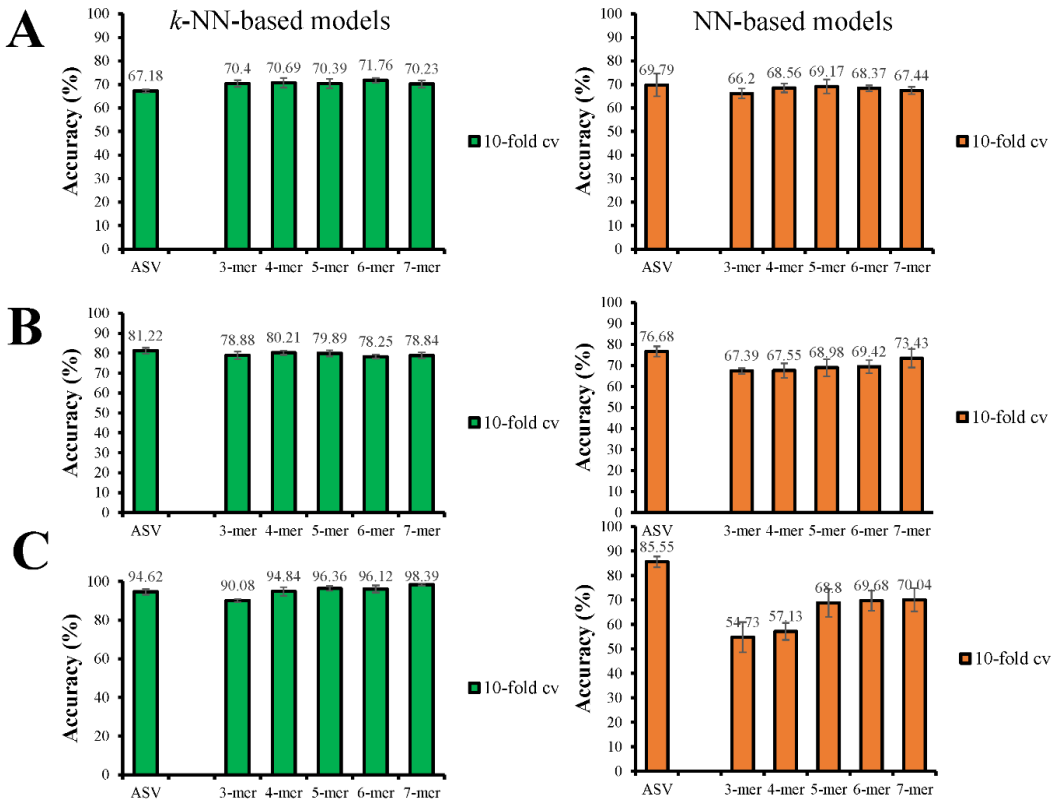

Figure S1. Comparison of the classification performance of PS classifiers based on *k*-NN and NN by using ASV datasets and hash datasets from 3-mer to 7-mer. (A) Accuracy of PS classifiers based on *k*-NN and NN using ASV and 3-mer to 7-mer hash datasets of Zhang18. The trained models were validated by using the 10-fold CV method. (B) Accuracy of PS classifiers based on *k*-NN and NN using ASV and 3-mer to 7-mer hash datasets of LiaoRl21 (Liao and Wang, 2022; this study). (C) Accuracy of PS classifiers based on *k*-NN and NN using ASV and 3-mer to 7-mer hash datasets of LiaoSm21.

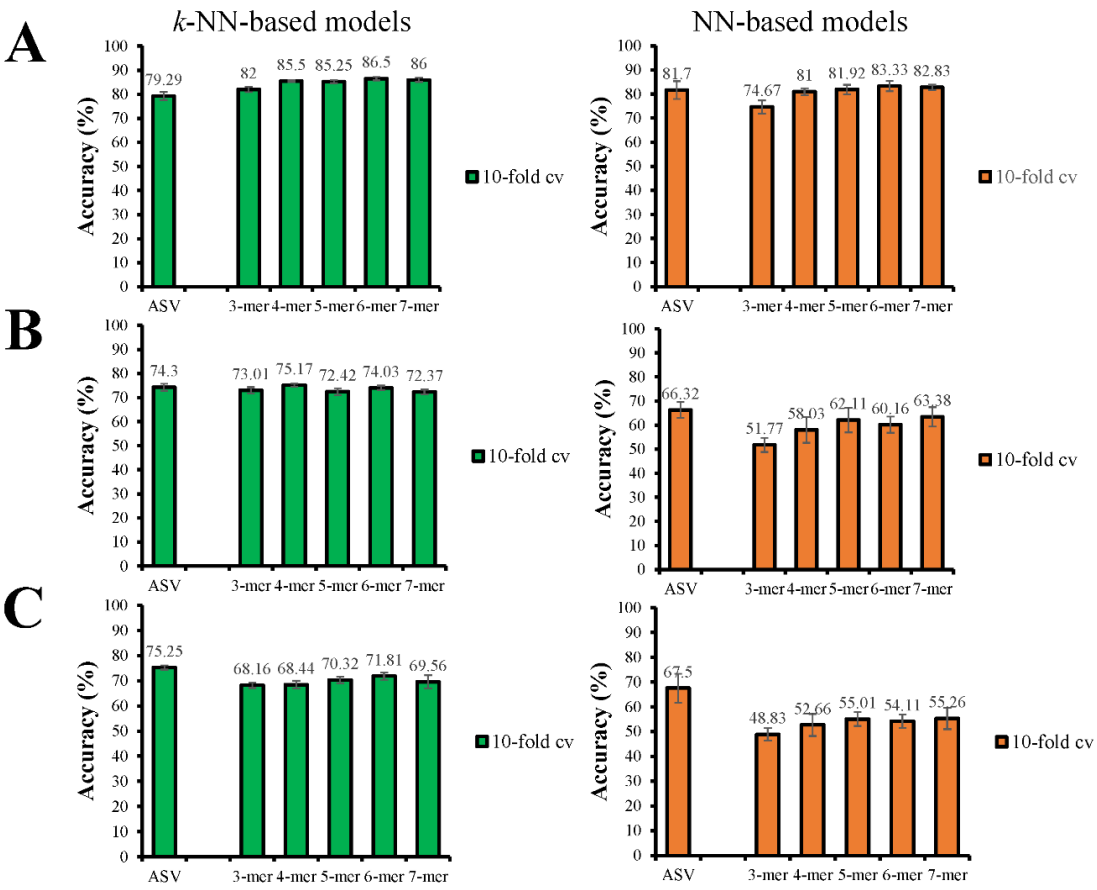

Figure S2. Comparison of the classification performance of PQ classifiers based on *k*-NN and NN by using ASV datasets and hash datasets from 3-mer to 7-mer. (A) Accuracy of PQ classifiers based on *k*-NN and NN using ASV and 3-mer to 7-mer hash datasets of Kusstatcher19 project. The trained models were validated by using 10-fold CV method. (B) Accuracy of PQ classifiers based on *k*-NN and NN using ASV and 3-mer to 7-mer hash datasets of LiaoRl21 project (this study). (C) Accuracy of PQ classifiers based on *k*-NN and NN using ASV and 3-mer to 7-mer hash datasets of LiaoSm21 project.

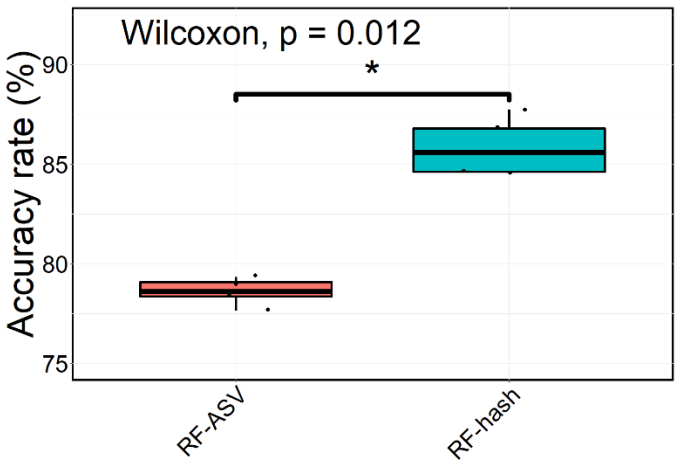

Figure S3. Accuracy of RF-based PS classifiers using integrated ASV and 7-mer hash datasets based on pathogen label. Accuracy of PS classifiers using an integrated ASV and 7-mer hash datasets associated with PS (Zhang18, LiaoRl21, and LiaoSm21). The Wilcoxon rank sum test was used for the pairwise comparison of PS and PQ models.

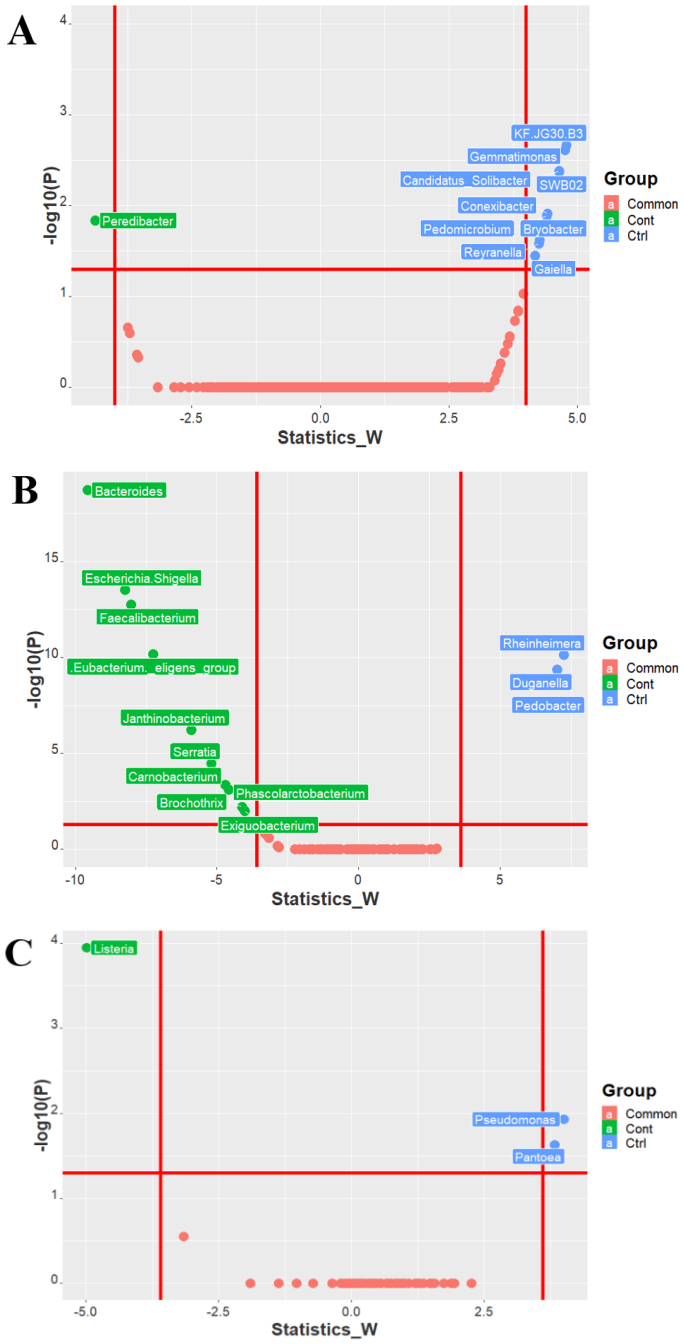

Figure S4. Taxonomic differential abundance analysis of individual fresh produce microbiota dataset related to PS. (A) Volcano plot of the significantly differential abundance of identified bacteria at the genus level in different groups of samples from Zhang18. (B) Volcano plot of the significantly differential abundance of identified bacteria at the genus level in different groups of samples from LiaoSm21. (C) Volcano plot of significantly differential abundance of identified bacteria at the genus level in samples

from LiaoRl21. Common represents bacteria with no significantly different abundance between non-contamination group (Ctrl) and contamination group (Cont).

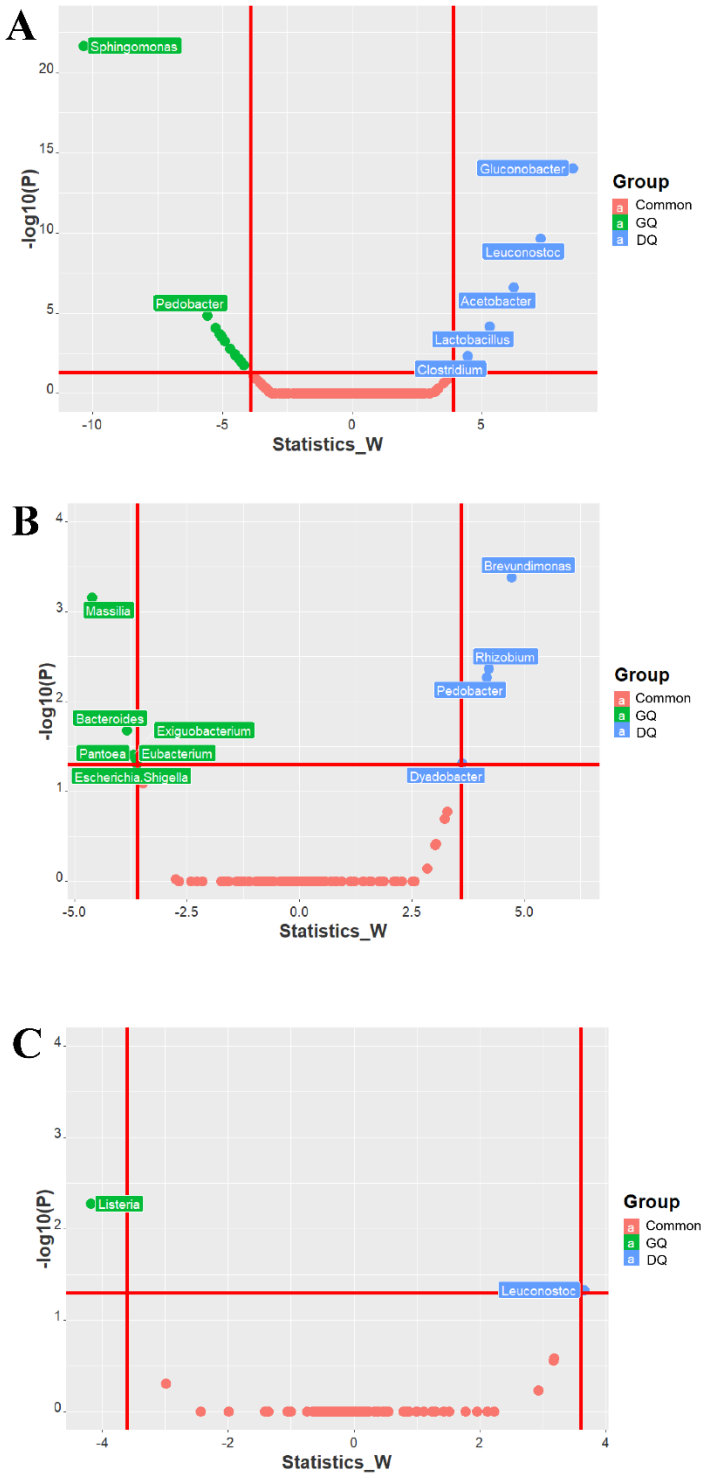

Figure S5. Taxonomic differential abundance analysis of individual fresh produce microbiota dataset related to PQ. (A) Volcano plot of significantly differential abundance of identified bacteria at the genus level in different group of samples from Kusstatscher_2019. (B) Volcano plot of significantly differential

abundance of identified bacteria at the genus level in different groups of samples from LiaoSm21. (C) Volcano plot of significantly differential abundance of identified bacteria at the genus level in samples from LiaoRl21. Common represents bacteria with no significantly different abundance between good-quality group (GQ) and decreasing-quality group (DQ).

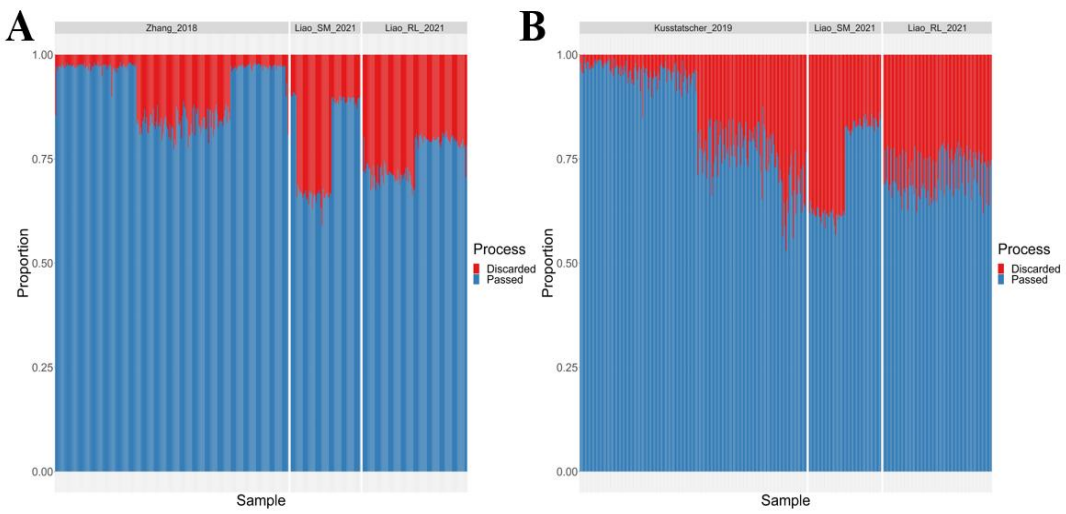

Figure S6. The proportion of sequence discarded from 16S rRNA sequence datasets of fresh produce microbiota related to PS and PQ during the denoising step by using the DADA2 plugin in QIIME 2. (A) Sequence loss from three microbiota sequence datasets associated with PS, including projects Zhang18, LiaoSm21, and LiaoRl21. (B) sequence loss from three microbiota sequence datasets associated with PQ, including projects Kusstatscher19, LiaoSm21, and LiaoRl21.

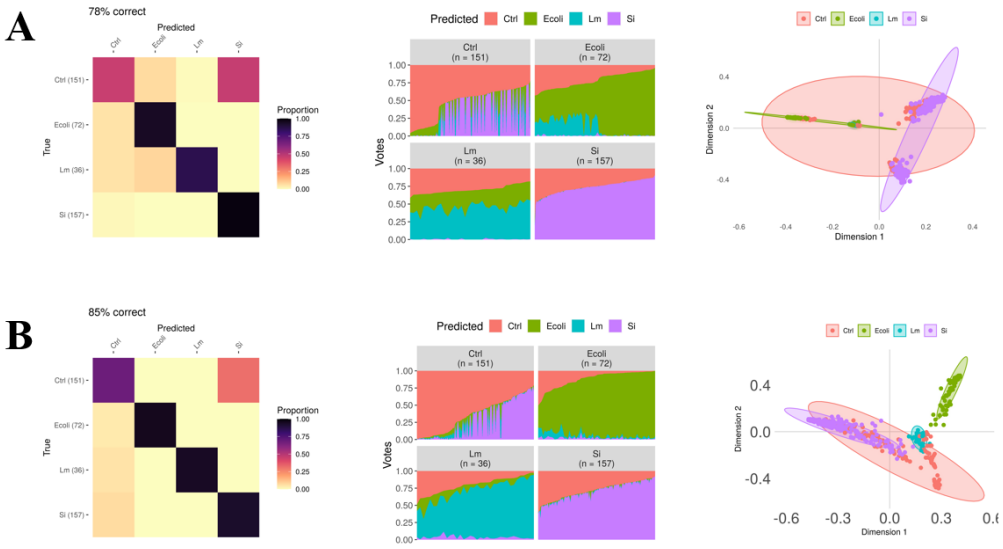

Figure S7. Plots of the confusion matrix (left), classification votes (middle), and MDS (right) generated from the RF-based models by using integrated ASV and 7-mer hash datasets associated with PS based on pathogen labels. (A) RF-based models established by using integrated ASV datasets. (B) RF-based models established by using integrated 7-mer hash datasets. The circles with color around the sample dots on MDS plots indicate the classification to Ctrl: control samples (red), Ecoli: *E. coli* O157 contaminated samples (green), Lm: *L. monocytogenes* contaminated samples, and Si: *Salmonella* Infantis contaminated samples (purple).