IDENTIFICATION AND CONTROL OF MICROBIAL HAZARDS ALONG THE PRODUCE SUPPLY CHAIN

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Alexandra Marie Belias

August 2021

IDENTIFICATION AND CONTROL OF MICROBIAL HAZARDS ALONG THE PRODUCE SUPPLY CHAIN

Alexandra Marie Belias, Ph. D.

Cornell University 2021

Microbial hazards, including *Listeria monocytogenes*, *Salmonella*, and Enterohemorrhagic *Escherichia coli* (EHEC), present complex challenges for the produce industry. The work here aimed to characterize these hazards and test relevant control strategies.

A review was conducted to identify hazards, risks and challenges associated with *Listeria* along the food supply chain. This review showed *Listeria* is prevalent across various environments (e.g., processing environments) and presents both public health and business risks. As such, identification of the root causes of contamination and implementation of control measures are warranted. To characterize *Listeria* presence and diversity in the preharvest environment, feces and water also were collected from a produce farm and tested for *Listeria*. A highly diverse *Listeria* population and evidence of *Listeria* transfer between sample types were observed, indicating tracing *Listeria* contamination of produce or the processing environment to a specific pre-harvest source is likely difficult without large sampling and subtyping efforts. To control *Listeria* contamination of the post-harvest environment, a root cause analysis (RCA) procedure was developed to guide intervention identification, implementation, and testing. Interventions using quaternary ammonium compound powder around forklift stops and a floor crack appeared to be effective. Two additional interventions were tested (i.e., use of a chlorinated cleaner in drains and dead-end pipe

removal) that were not immediately effective but highlighted that RCA should often be an iterative process.

Machine learning models were developed to predict *Salmonella* and EHEC marker (i.e., *eaeA* and *stx*) presence in canal water using spatial and temporal factors. These models indicated machine learning shows promise for predicting pathogen contamination of agricultural water. Additionally, leafy greens were inoculated with *E. coli* and *Salmonella* via simulated irrigation events in 3 distinct climates, and produce samples were collected for 4 days following inoculation. The *E. coli* and *Salmonella* die-off rates were calculated and the effect of weather on die-off was assessed. These results showed weather, and particularly relative humidity, is important in determining the effectiveness of using die-off as a control strategy.

Overall, this work further illustrates the complexities of managing microbial hazards along the produce supply chain and provides potential solutions for controlling these hazards.

BIOGRAPHICAL SKETCH

Alexandra Belias was born in Rochester, NY and graduated from Pittsford Mendon High School. She then attended Purdue University where she earned a Bachelor of Science in Food Science. While at Purdue, she conducted produce safety research under Dr. Amanda Deering, which led her to pursue a graduate degree in the field. In addition, Alexandra served as a teaching assistant in the Introduction to Microbiology Lab and the Food Microbiology lecture at Purdue. Following her bachelor's degree, she worked as a summer intern at Wegman's where she assisted in shelf-life determination of new products and assisted the Wegman's manufacturing facilities in product and environmental testing. After her internship, Alexandra started her graduate degree at Cornell University under the advisement of Dr. Martin Wiedmann and Dr. Renata Ivanek. While at Cornell, Alexandra had the opportunity to work on a variety of produce safety projects spanning the pre- and post-harvest produce environments. She also had the opportunity to collaborate with food companies in the United States and abroad to help them develop and improve their *Listeria* environmental monitoring programs, as well as give training and webinars to food industry members. After completing her Ph.D., Alexandra has accepted a position to work in the produce industry.

To my family, and especially by parents, for always providing me with the support to achieve my goals.

ACKNOWLEDGMENTS

There have been many people that have provided me with support and encouragement throughout graduate school that I am infinitely grateful for. I would like start by thanking Dr. Martin Wiedmann for serving as my advisor. Martin always challenged me to reach my fullest potential and become a better scientist. In addition, Martin always went out of his way to teach me, to help me when things went wrong in my projects, and to provide me with opportunities to reach my goals.

I would like to thank Dr. Renata Ivanek for taking me into her lab and spending an extensive amount of time mentoring and teaching me. Renata has always gone out of her way to provide me with support and life and career advice. It has been a pleasure getting to know her over the years. I would also like to thank Dr. Randy Worobo for serving on my committee. Randy has always been willing to share his learnings to further my research.

I am incredibly grateful to have been a member of the Food Safety Lab; it has been a wonderful experience being able to get to know and learn from all the members of the Wiedmann lab over the past four years. Specially, I would like to thank Dr. Daniel Weller for mentoring me, always being eager to teach me everything there is to know about produce safety, modeling, and GIS. I would like to thank Dr. Genevieve Sullivan for always being willing to discuss our projects and being a great travel companion. I would like to thank Dr. Sarah Murphy and Ruixi Chen for making our trip to China memorable and one of my favorite memories of graduate school. In addition, I would like to thank Sherry Roof, Maureen Gunderson, Alan Bitar, Sean Guo, Samantha Bolten, Nathanial Henderson, and Sriya Sunil for helping me on various aspects of my projects; I would not have been able to complete graduate degree without all your help.

I would also like to thank the members of the Ivanek lab for welcoming me in as a guest in their lab and always being willing to provide me with support when I needed it.

Lastly, I would like to thank my friends and family for encouraging me to pursue a graduate degree and providing me with support to make it through. Specifically, I would like to thank my parents. They always encouraged me to go after my goals and have taught me the importance of hard work. Also a special thanks to my grandparents, brothers, Matthew King, and Hannah Walpole for being there for me throughout the years.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Control of foodborne microbial hazards (e.g., *Listeria monocytogenes*, *Salmonella* spp., and Enterohemorrhagic *Escherichia coli*; EHEC) and the associated risks remain substantial concerns for industry. In fact, there have been several outbreaks caused by *L. monocytogenes (23, 25, 26, 30)*, *Salmonella* spp. *(21, 24, 27, 31)*, and EHEC *(22, 28, 29, 32)* contamination of fresh produce. Of notable significance, in 2014 there was a *L. monocytogenes* illness outbreak linked to contamination of cantaloupes that caused 147 illnesses and 33 deaths *(23)*. In 2015 and 2016, there was a *Salmonella* Poona illness outbreak linked to cucumbers that caused 907 illnesses and 6 deaths *(27)*. In addition, in 2018 there was an *E. coli* O157:H7 illness outbreak linked back to romaine lettuce that caused 210 illnesses and 5 deaths *(28)*. Furthermore, there have been 32, 28, and 6 recalls due to *L. monocytogenes*, *Salmonella* spp., and EHEC contamination of produce in the United States from 2019 to 2020, respectively *(33)*.

Microbial contamination of produce can occur at many stages along the supply chain, such as during planting, growing, harvesting, packing, processing, or transportation. In the pre-harvest environment (i.e., during the planting, growing, and harvesting stages), introduction of foodborne pathogens can come from a variety of sources *(2, 29, 32, 26)*. For instance, surface waters (e.g., ponds, streams, and canals) are used as agricultural water for irrigation, agrochemical applications, frost protection, etc. Agricultural water has been shown to be contaminated with foodborne pathogens *(1, 7, 41)*, and application of contaminated water to produce has been shown to be associated with the contamination of produce *(2, 29, 32, 26, 34)*. Furthermore, foodborne pathogens have also been isolated from soil *(34, 35)*. Transfer of pathogens from soil to produce can occur due to splash cause by rain or irrigation, wind, and

movement in the field by workers, harvest equipment, among other routes. Wildlife *(10, 11, 19)* and livestock *(2, 29, 15)* may also be carriers of foodborne pathogens. Wildlife feces can be found in produce fields and can directly contaminate produce or lead to produce contamination via splash (e.g., during rain events) *(40, 36)*. Run-off from livestock operations (e.g., concentrated animal feeding operations) can enter water sources or deposit directly onto produce fields *(2, 29)*. In addition, if biological soil amendments of animal origin are not treated properly to reduce microbial loads, they can also introduce pathogens into the preharvest produce environment *(3, 9)*. While these represent some of the routes of pre-harvest produce contamination, additional sources may also exist (e.g., cross-contamination from harvest equipment or employees introducing pathogens into the environment). As such, controls must be put in place to prevent initial contamination and additional protections must be put in place to reduce risks in the case that contamination has occurred.

In addition, some foodborne pathogens (i.e., *L. monocytogenes* and *Salmonella* spp.) have been isolated from the post-harvest produce environment (i.e., during the packing and processing stages) *(16, 38)*. In particular, *L. monocytogenes* can establish persistence within produce operations (i.e., packinghouses or processing facilities), especially in cool and wet areas *(16)*. *L. monocytogenes* can be introduced into produce operations via produce, produce storage crates, employees acting as fomites, forklifts, transportation vehicles, among other routes *(14, 19)*. Once introduced, the *L. monocytogenes* can enter into niches in the equipment and facility structure where it is able to evade cleaning and sanitation and establish persistence *(16, 19)*. Once in the niche, the *L. monocytogenes* can grow (i.e., when nutrients and moisture are present) and eventually move throughout the operation via mobile equipment and employees; this *L. monocytogenes* can then contaminate the product *(16, 17)*. In addition, *L. monocytogenes* may continually be introduced into an operation, which is referred to as

"persistent transient *Listeria*." Both persistent and persistent transient *L. monocytogenes* can lead to recalls and outbreaks, and therefore, control of *L. monocytogenes* in the produce operation environment must be maintained. In addition, *Salmonella* spp. has been shown to survive in food processing facilities, however, it is typically isolated from in dry, warm environments *(6, 20)* as compared to the wet, cool environments where *Listeria* is typically isolated from. While dry, warm areas are less common in produce facilities, they can exist (e.g., nut facilities). As such, control measures and environmental sampling for *Salmonella* spp. may be necessary in these areas in addition to control measures for *L. monocytogenes*.

Therefore, foodborne pathogen contamination of produce can occur through a variety of different routes, making control of foodborne pathogens along the produce supply chain complex. Due to these complexities, the produce industry must deploy multiple control measures at each stage along the supply chain (e.g., antimicrobial treatment of water, composting of biological soil amendments, no harvest buffers around wildlife feces, wait times between irrigation and harvest, among others). However, additional research is required to better understand foodborne pathogen dynamics so these control strategies can be optimized. As such, this work aimed to (i) characterize *L. monocytogenes*, *Salmonella* spp., and EHEC presence, diversity, and dynamics at various stages of the produce supply chain, and (ii) test strategies for identifying and controlling *L. monocytogenes*, *Salmonella* spp., and EHEC along the produce supply chain. In particular, this work (i) reviewed the hazards, risks, and challenges associated with *Listeria* spp. and *L. monocytogenes* along the food supply chain, (ii) evaluated the presence and diversity of *Listeria* spp. and *L. monocytogenes* in the pre-harvest produce environment, (iii) evaluated a root cause analysis (RCA) procedure for eliminating and reducing persistent and persistent transient *Listeria* spp. and *L. monocytogenes* from the packinghouse environment, (iv) evaluated the use of

predictive modeling to identify southwestern U.S. canal water used as agricultural water that is contaminated with EHEC markers (i.e., *stx* and *eaeA*) or *Salmonella* spp., and (v) characterized the effect of weather on die-off of *E. coli* and *Salmonella* spp. on baby spinach and lettuce following a simulated irrigation event in 3 distinct climatic regions.

REFERENCES

1. Cooley, M. B., B. Quinones, D. Oryang, R. E. Mandrell, and L. Gorski. 2014. Prevalence of shiga toxin producing *Escherichia coli*, *Salmonella enterica*, and *Listeria monocytogenes* at public access watershed sites in a California Central Coast agricultural region. *Front. Cell. Infect. Microbiol*. 4(3):1-13. doi: 10.3389/fcimb.2014.00030

2. Delaquis, P., S. Bach, and L. D. Dinu. 2007. Behavior of *Escherichia coli* O157:H7 in leafy vegetables. *J. Food Prot*. 70(8):1966-1974. doi: 10.4315/0362-028x-70.8.1966

3. Erickson, M. C., C. Smith, X. Jiang, I. D. Flitcroft, and M. P. Doyle. 2015. Manure source and age affect survival of zoonotic pathogens during aerobic composting at sublethal temperatures. *J. Food Prot*. 78(2):302-310.

4. Falardeau, J., R. P. Johnson, F. Pagotto, and S. Wang. 2017. Occurrence, characterization, and potential predictors of verotoxigenic *Escherichia coli*, *Listeria monocytogenes*, and *Salmonella* in surface water used for produce irrigation in the Lower Mainland of British Columbia, Canada. *PLoS One*. 12(9):e0185437.

5. Ferreira, V., M. Wiedmann, P. Teixeira, and M. J. Stasiewicz. 2014. *Listeria monocytogenes* persistence in food-associated environments: epidemiology, strain characteristics, and implications for public health. *J Food Prot*. 77(1):150-170.

6. Finn, S., O. Condell, P. McClure, A. Amezquita, and S. Fanning. 2013. Mechanisms of survival, responses, and sources of *Salmonella* in low-moisture environments. *Front. Microbiol*. 4(331).

7. Hanning I. B., J. D. Nutt, and S. C. Ricke. 2009. Salmonellosis outbreaks in the United States due to fresh produce: sources and potential intervention methods. *Foodborne Pathog. Dis.* 6(6):635-648.

8. Heaton J. C. and Jones K. 2008. Microbial contamination of fruit and vegetables and the behaviour of enteropathogens in the phyllosphere: a review. *J. Appl. Microbiol.* 104:613-626.

9. Hutchison, M. L., L. D. Walters, S. M. Avery, F. Munro, and A. Moore. 2005. Analyses of livestock production, waste storage, and pathogen levels and prevalences in farm manures. *Appl. Environ. Microbiol*. 71(3):1231-1236.

10. Jay, M. T., M. Cooley, D. Carychao, G. W. Wiscomb, R. A. Sweitzer, L. Crawford-Miksza, J. A. Farrar, D. K. Lau, J. O'Connell, A. Millington, R. V. Asmundson, E. R. Atwill, and R. E. Mandrell. 2007. *Escherichia coli* O157:H7 in feral swine near spinach fields and cattle, central California coast. *Emerg. Inf. Dis*. 13(12):1908-1911. doi: 10.3201/eid1312.070763

11. Jay-Russell, M. T., A. F. Hake, Y. Bengson, A. Thiptara, and T. Nguyen. 2014. Prevalence and characterization of *Escherichia coli* and *Salmonella* strains isolated from stray dog and coyote feces in a major leafy greens production region at the United States- Mexico border. *PLoS ONE*. 9(11):e113433. doi: 10.1371/journal.pone.0113433

12. Jeamsripong S., J. A. Chase, M. T. Jay-Russell, R. L. Buchanan, and E. R. Atwill. 2019. Experimental in-field transfer and survival of *Escherichia coli* from animal feces to romaine lettuce in Salinas Valley, California. *Microorganisms* 7(408).

13. Lyautey, E., A. Hartmann, F. Pagotto, K. Tyler, D. R. Lapen, G. Wilkes, P. Piveteau, A. Rieu, W. J. Robertson, D. T. Medeiros, T. A. Edge, V. Gannon, and E. Topp. 2007. Characteristics and frequency of detection of fecal *Listeria monocytogenes* shed by livestock, wildlife, and humans. *Can. J. Microbiol*. 53: 1158-1167. doi:10.1139/W07-084.

14. Malley, T. J. V., J. Butts, and M. Wiedmann. 2015. Seek and Destroy Process: *Listeria monocytogenes* process controls in the ready-to-eat meat and poultry industry. *J. Food Prot. 78*(2):436–445. https://doi.org/10.4315/0362-028X.JFP-13-507

15. Nightingale, K. K., Y. H. Schukken, C. R. Nightingale, E. D. Fortes, A. J. Ho, Z. Her, Y. H. Grohn, P. L. McDonough, and M. Wiedmann. 2004. Ecology and transmission of *Listeria monocytogenes* infecting ruminants and in the farm environment. *Appl. Environ. Microbiol*. 70(8):4458-4467. DOI: 10.1128/AEM.70.8.4458–4467.2004

16. Park, S., B. Szonyi, R. Gautam, K. Nightingale, J. Anciso, and R. Ivanek. 2012. Risk factors for microbial contamination in fruits and vegetables at the

preharvest level: A systematic review. *J. Food Prot*. 75:2055-2081. doi:10.4315/0362-028X.JFP-12-160.

17. Simmons, C. K. and M. Wiedmann. 2018. Identification and classification of sampling sites for pathogen environmental monitoring programs for *Listeria monocytogenes*: Results from expert elicitation. *Food Microbiol.* 75:2-17.

18. Strawn, L. K., E. D. Fortes, E. A. Bihn, K. K. Nightingale, Y. T. Grohn, R. W. Worobo, M. Wiedmann, and P. W. Bergholz. 2013. Landscape and meteorological factors affecting prevalence of three food-borne pathogens in fruit and vegetable farms. *Appl. Environ. Microbiol*. 79:588-600.

19. Tompkin, R. B. 2002. Control of *Listeria monocytogenes* in the food-processing environment. *J. Food Prot. 65*(4), 709–725. https://doi.org/10.4315/0362-028X-65.4.709

20. United Fresh Produce Association. 2021. Considerations for Fresh Produce Dry-pack Environmental Monitoring Programs (EMPs). Available at: https://info.unitedfresh.org/hubfs/Food%20Safety%20Resources/Dry%20environment%20EMPs_FINAL.pdf. Accessed 6 May 2021.

21. U.S. Centers for Disease Control and Prevention. 2006. Multistate Outbreak of *Salmonella* Typhimurium Infections Linked to Tomatoes (FINAL UPDATE). Available at: https://www.cdc.gov/salmonella/2006/tomatoes-11-2006.html. Accessed 6 May 2021.

22. U.S. Centers for Disease Control and Prevention. 2006. Multistate outbreak of *E. coli* O157:H7 infections linked to fresh spinach (FINAL UPDATE). Available at: https://www.cdc.gov/ecoli/2006/spinach-10-2006.html. Accessed 6 May 2021.

23. U.S. Centers for Disease Control and Prevention. 2011. Multistate outbreak of listeriosis linked to whole cantaloupes from Jensen Farms, Colorado (FINAL UPDATE). Available at: https://www.cdc.gov/listeria/outbreaks/cantaloupes-jensen-farms/index.html. Accessed 6 May 2021.

24. U.S. Centers for Disease Control and Prevention. 2012. Multistate outbreak of *Salmonella* Braenderup infections associated with mangoes (final update). Available at: https://www.cdc.gov/salmonella/braenderup-08-12/index.html. Accessed 6 May 2021.

25. U.S. Centers for Disease Control and Prevention. 2015. Multistate outbreak of listeriosis linked to commercially produced, prepackaged caramel apples made from Bidart Bros. Apples (final update). Available at: https://www.cdc.gov/listeria/outbreaks/caramel-apples-12-14/index.html. Accessed 6 May 2021.

26. U.S. Centers for Disease Control and Prevention. 2016. Multistate outbreak of listeriosis linked to packaged salads produced at Springfield, Ohio Dole processing facility (final update). https://www.cdc.gov/listeria/outbreaks/bagged-salads-01-16/index.html. Accessed 6 May 2021.

27. U.S. Centers for Disease Control and Prevention. 2016. Multistate outbreak of *Salmonella* Poona infections linked to imported cucumbers (final update). Available at: https://www.cdc.gov/salmonella/poona-09-15/index.html. Accessed 6 May 2021.

28. U.S. Centers for Disease Control and Prevention. 2018. Multistate outbreak of *E. coli* O157:H7 infections linked to romaine lettuce (final update). Available at: https://www.cdc.gov/ecoli/2018/o157h7-04-18/index.html. Accessed 6 May 2021.

29. U.S. Centers for Disease Control and Prevention. 2019. Outbreak of *E. coli* infections linked to romaine lettuce. Available at: https://www.cdc.gov/ecoli/2018/o157h7-11-18/index.html. Accessed 6 May 2021.

30. U.S. Centers for Disease Control and Prevention. 2020. Outbreak of *Listeria* infections linked to enoki mushrooms. Available at: https://www.cdc.gov/listeria/outbreaks/enoki-mushrooms-03-20/index.html. Accessed 6 May 2021.

31. U.S. Centers for Disease Control and Prevention. 2020. Outbreak of *Salmonella* Enteritidis infections linked to peaches. Available at: https://www.cdc.gov/salmonella/enteritidis-08-20/index.html. Accessed 6 May 2021.

32. U.S. Centers for Disease Control and Prevention. 2020. Outbreak of *E. coli* infections linked to romaine lettuce. Available at:

https://www.cdc.gov/ecoli/2019/o157h7-11-19/index.html. Accessed 6 May 2021.

33. U.S. Food and Drug Administration. 2021. Recalls, market withdrawals, & safety alerts. Available at: https://www.fda.gov/safety/recalls-market-withdrawals-safety-alerts. Accessed 6 May 2021.

34. Weller, D., M. Wiedmann, and L. K. Strawn. 2015. Irrigation is significantly associated with an increased prevalence of *Listeria monocytogenes* in produce production environments in New York State. *J. Food Prot*. 78:1132-1141. doi:10.4315/0362-028X.JFP-14-584.

35. Weller, D., M. Wiedmann, and L. K. Strawn. 2015. Spatial and temporal factors associated with an increased prevalence of *Listeria monocytogenes* in spinach fields in New York State. *Appl. Environ. Microbiol*. 81:6059-6069. doi:10.1128/AEM.01286-15. doi:10.1128/AEM.02491-12.

36. Weller D. L., J. Kovac, D. J. Kent, S. Roof, J. I. Tokman, E. Mudrak, and M. Wiedmann. 2019. A conceptual framework for developing recommendations for no-harvest buffers around in-field feces. *J. Food Prot*. 82(6):1052-1060. doi: 10.4315/0362-028X.JFP-18-414.

37. Wilkes, G., T. A. Edge, V. P. J. Gannon, C. Jokinen, E. Lyautey, N. F. Neumann, N., Ruecker, A. Scott, M. Sunohara, E. Topp, and D. R. Lapen. 2011. Associations among pathogenic bacteria, parasites, and environmental and land use factors in multiple mixed-use watersheds. *Water Res*. 45:5807-5825. doi:10.1016/j.watres.2011.06.021.

38. Zoellner, C., K. Ceres, K. Ghezzi-Kopel, M. Wiedmann, and R. Ivanek. 2018. Design elements of *Listeria* environmental monitoring programs in food processing facilities: a scoping review of research and guidance materials. *Compr. Rev. Food Sci. F*. 17:1156-1171.

CHAPTER 2

HAZARDS, RISKS AND CHALLENGES OF *LISTERIA* IN THE FOOD SUPPLY

Published in: Food Safety Management in Practice (2021).

ABSTRACT

Listeria monocytogenes contamination of Ready-To-Eat (RTE) food products and food-associated built environments (e.g., processing plants) represents a food safety issue with major public health and business risk implications. A number of factors make *L. monocytogenes* control a particular challenge, including (i) its frequent presence in different environments, (ii) its propensity for establishing persistence in food-associated environments, (iii) its ability to grow under a variety of stressful conditions, and (iv) its ability to cause severe illness, particularly in immunocompromised individuals and pregnant women. Key sources of *L. monocytogenes* contamination of RTE foods are food-associated built environments. However, raw material can also be an important source, particularly for products without a "kill step" (e.g., fresh produce, raw dairy products, cold-smoked seafood). While certain RTE foods (e.g., deli meats, soft cheeses, produce) have commonly been linked to listeriosis outbreaks, cases, and recalls, a number of factors will influence the specific public risk a given RTE food represents, including likelihood of contamination, ability to support *L. monocytogenes* growth, and consumer related factors (including consumption by pregnant women or immunocompromised individuals). Consequently, a risk-based approach presents the most appropriate strategy to minimize the public health and business impact of *L. monocytogenes*. Key challenges to control *L. monocytogenes* include (i) development and implementation of food safety systems that prevents *L. monocytogenes* persistence in food-associated built environments, (ii) minimizing *L. monocytogenes* contamination

of raw material sources, (iii) implementation of effective root cause analysis procedures, (iv) minimizing *L. monocytogenes* growth in finished product, and (v) consumer education.

INTRODUCTION

Listeria monocytogenes is a bacterial foodborne pathogen that causes an estimated 1,591 *(38)* illnesses per year in the United States and an estimated 23,510 *(25)* illnesses per year in the world. While the total number of illnesses caused by *L. monocytogenes* is small compared to other foodborne pathogens (e.g., nontyphoidal *Salmonella* spp., which cause an estimated 1,027,561 foodborne illnesses per year in the United States), *L. monocytogenes* has a high case-fatality rate of approx. 16% *(38)*. This makes it a serious public health concern. The risk of illness and death caused by *L. monocytogenes* is especially high for pregnant women, the old, and other individuals with compromised immune systems. *L. monocytogenes* can cause an invasive infection with symptoms including sepsis and meningitis in immunocompromised individuals, as well as rarely in healthy individuals *(9)*, and spontaneous abortions in pregnant women *(9)*. *L. monocytogenes* can also cause gastrointestinal illness in healthy individuals. Contamination of ready-to-eat (RTE) food products with *L. monocytogenes* is also a common cause of food recalls, which can damage company reputation and cause substantial financial losses. In fact, there have been 125 recalls of foods and beverages in the United States due to *L. monocytogenes* contamination between 2018 and 2019 *(46, 44)*. As such, public health officials and the food industry have substantial stakes in reducing contamination of RTE food products with *L. monocytogenes*, as well as in the implementation of other strategies that reduce human listeriosis cases (e.g., reducing growth of *L. monocytogenes* in foods, education campaigns targeting susceptible consumers).

The need to effectively control *L. monocytogenes* has been further heightened with the broad use of molecular subtyping tools and specifically whole genome sequencing (WGS), which has led to improved detection of listeriosis outbreaks, including small outbreaks and/or outbreaks that occur over prolonged time periods (e.g.,

years) *(28)*. More specifically, the US Centers for Disease Control and Prevention has been performing WGS of all human *L. monocytogenes* isolates since 2013 *(11)*. Similarly, routine WGS of human clinical *L. monocytogenes* isolates has also been performed by the European Centre for Disease Prevention and Control and the Public Health Agency of Canada since 2014 *(13)* and 2017 *(35)*, respectively, with other public health agencies around the world also increasingly switching to routine use of WGS for characterization of *L. monocytogenes*. Even with WGS, detailed epidemiological data are still needed to reliably and definitively identify the specific food source responsible for a given outbreak. Importantly, subtyping and specifically WGS data are also increasingly used by regulatory agencies to characterize *Listeria* spp. and *L. monocytogenes* isolates obtained from foods and food-associated built environments. Subtype data, such as WGS, may provide evidence for *L. monocytogenes* persistence in processing plants, which in some countries and jurisdictions may be used by regulatory agencies to help identify unhygienic conditions in a given facility, which can lead to facility shut-downs and recalls even in the absence of finished product contamination. Hence, environmental *L. monocytogenes* contamination can also represent a considerable business risk, even when there is a limited public health risk.

Key factors that affect the risk of human foodborne listeriosis cases linked to a specific food include (i) initial contamination of the food; (ii) ability of the food to support *L. monocytogenes* growth, and (iii) the susceptibility of the consumers of a specific food product. As (ii) and (iii) have been detailed in a number of publications and reviews (e.g., *5, 15, 21, 22, 34*), the paper presented here specifically focuses on the risks associated with initial contamination of foods as well as the challenges the food industry faces in its efforts to reduce initial contamination. Efforts to reduce *L. monocytogenes* contamination of food products are complicated by the fact that this organism (i) is frequently found in a variety of different environments, making

introduction into raw materials and processing facilities a high risk, (ii) is able to survive and grow under adverse environmental conditions *(26)*, and (iii) has a propensity to establish persistent populations in food associated built environments (e.g., processing plants) and equipment. Consequently, contamination of food products can occur through a variety of different routes, including natural environments (e.g., for raw materials harvested from nature, such as wild-caught seafood), primary production environments (i.e., livestock or produce farms), raw materials, and the food-associated built environments (e.g., processing plants, retail establishments) and equipment, among others *(16)*.

The reminder of this article will (i) detail sources of *Listeria*, (ii) discuss public health and business risks associated with *L. monocytogenes* and how to develop and implement risk-based systems to address these food safety issues, and (iii) outline key control strategies and associated challenges.

SOURCES OF *LISTERIA*

While *L. monocytogenes* is the only human pathogen in the genus *Listeria*, testing for the presence of *Listeria* spp. is often used by industry to monitor processing plant environments for the presence of conditions that would facilitate presence, survival, and/or growth of *L. monocytogenes (12)*. *Listeria* spp., as well as *L. monocytogenes*, have been isolated from a wide variety of environments, including from soil, water, feces, and vegetation in the primary production environment *(18, 29, 41, 42, 48, 49, 50)*; in pristine environments such as national parks *(37)*; in urban environments from sidewalks, ATM machines, among others *(37)*; in processing environments *(16)*; and in retail and food service environments *(20)*. Consequently, many different sources can be responsible for introduction of *Listeria* spp. and *L. monocytogenes* into finished product and food associated environments, including primary production environments

and raw materials, natural environments, food-associated built environments and processing equipment (including processing plants and retail). Importantly, while employees may act as fomites, there is essentially no evidence that human fecal carriers play a role as sources of *L. monocytogenes* in foods and food associated environments.

Primary Production Environments and Raw Materials. Primary production environments (e.g., farms, fields) and raw materials can play two distinct roles as sources of *Listeria,* including (i) introduction into raw materials that do not undergo a kill step and where *L. monocytogenes* can be carried over into the finished RTE product (e.g., fresh-cut produce, raw milk dairy products, cold-smoked seafood), and (ii) introduction into food-associated environments (e.g., processing plants) with the potential of subsequent environmental transmission into finished RTE products. Control of *Listeria* in raw materials is hence particularly important for production of RTE foods that do not involve an effective kill step. Contamination of raw materials can occur from a variety of sources and at a variety of points in the supply chain prior to materials reaching a processing facility (or a retail establishment), including (i) primary production (e.g., produce or livestock farms), (ii) at an upstream facility (e.g., a storage facility, a packing house, or a slaughterhouse), and (iii) during transportation.

One key supply chain where raw material contamination is of particular concern is raw and fresh-cut produce. In produce primary production environments, *Listeria* has been isolated from agricultural water sources, soil, vegetation, and wildlife feces *(12, 41, 42, 49, 50)*. With a variety of possible sources, there are many transmission pathways that can lead to pre-harvest produce contamination with *Listeria*. For instance, contaminated irrigation water could directly deposit *Listeria* onto product or could deposit it into soil with the possibility of subsequent transmission onto produce *(32, 49)*. Soil could also harbor *Listeria* populations that can be transferred to produce. Finally, wildlife could directly deposit *Listeria* contaminated feces onto produce or into the soil.

Livestock could also present a source of *Listeria* in produce at the preharvest level. The possible role of livestock as a source of *L. monocytogenes* is supported by the fact that a listeriosis outbreak in the Maritime Provinces, Canada was linked to coleslaw that appeared to have been contaminated from sheep feces *(39)*, as well as the frequent high prevalence of *Listeria* spp. and *L. monocytogenes* in livestock. In fact, Golden et al. *(18)* found a *Listeria* spp. (including *L. monocytogenes*) prevalence of 15.9% (245/1,537) and a *L. monocytogenes* prevalence of 1.8% (28/1,537) from soil and fecal samples collected from 11 poultry farms in the southeastern United States, further supporting the importance of livestock-associated sources.

A number of factors can influence how effectively *Listeria* are transferred from soil (or other environmental sources) to produce. During a rain or irrigation event, splashing of soil or wildlife feces, can facilitate transfer; in addition, run-off and flooding can also facilitate contamination of the produce *(31)*. Increased wind speed has also been shown to be associated with an increased prevalence of *Listeria*, possibly due to increased transfer of *Listeria* from surrounding environments (e.g., farms) *(31)*. Unlike for *Salmonella* and pathogenic *E. coli*, to date, very few listeriosis outbreaks have been definitively linked to contamination of produce that occurred at the pre-harvest level. With the frequent presence of *L. monocytogenes* in natural and farm environments it is however extremely likely that a number of finished product contamination events are linked to the pre-harvest environment, particularly for products that do not undergo extensive anti-microbial wash treatments or heat treatments. The limited number of outbreaks directly linked to pre-harvest sources may be a reflection of the fact that pre-harvest contamination events more likely lead to individual listeriosis cases rather than outbreaks, however additional information is required to confirm low levels of *Listeria* are typically present in the preharvest environment.

Another supply chain where pre-harvest contamination is relevant, is cold-smoked seafood. While a number of studies have provided convincing evidence that contamination of finished RTE cold-smoked products (in particular, cold-smoked salmon) can be traced back to raw materials *(24)*, it is rare to identify the specific contamination sources of the raw material (e.g., fish farms, fish slaughter facilities, transport equipment, etc.). Importantly, however, cold-smoked salmon represents a model for a supply chain that has started to implement innovative approaches to reduce contamination of incoming raw materials, ranging from stringent supplier qualification supported by intensive raw material testing to non-thermal treatments to reduce *Listeria* loads on incoming raw material *(24)*.

Finally, *L. monocytogenes* contamination of raw materials is also important for raw milk (which is legal for sale in some locations) and raw milk dairy products, particularly as *L. monocytogenes* have been shown to often be highly prevalent on livestock farms. For instance, Nightingale et al. *(29)* found a *L. monocytogenes* prevalence of 20.1% (414/2,056) in a survey of 52 ruminant farms where samples were collected of feces, soil, feedstuff, and water. There was also a *L. monocytogenes* illness outbreak linked to raw chocolate milk, which caused 2 illnesses and 1 death *(10)*.

In addition to raw materials being a direct source of *Listeria* in finished product and an indirect source (via introduction from farm environments into processing plants), *Listeria* can also establish itself within vehicles and crates used to transport product from primary production environments to packing houses, processing facilities, or directly to retail. Cross-contamination during transport is possible whenever product is exposed to the open environment. To prevent cross-contamination at this stage, regular cleaning and sanitation of trailers that transport crates and raw materials is thus necessary.

Employees. Employees are sometimes brought up as potential sources or

vectors that contribute to introduction of *Listeria* into the processing environment. However, unlike other pathogens, such as *Salmonella*, it is unlikely for humans to be fecal carriers of *Listeria (36)*. Employees however can act as fomites, particularly since *Listeria* can often be found at high frequencies in urban, rural, and natural environments. For example, in a survey for *Listeria* in four urban environments in New York State, Sauders et al. *(37)* found the overall *Listeria* spp. (including *L. monocytogenes*) prevalence to be 23.4% (N=907). Once in the facility, *Listeria* can survive over time, particularly if effective food safety and sanitation practices are lacking. In order to prevent the introduction of *Listeria* into the processing environment or onto the product via employees, effective good manufacturing practices (GMPs) must be followed. These GMPs include hand washing and sanitation, wearing gloves, wearing clean coats designated for use only inside of the processing areas, boot washes (or footbaths) at the entrance to the processing areas, and captive boot policies (i.e., each employee has a set of work boots kept at the facility which may only be used inside the processing areas).

Processing Plant and Packing House Environment. The majority of *L. monocytogenes* outbreaks have been linked to contamination of RTE foods originating from processing plant or packing house environments. *Listeria* presence in food processing associated environments can conceptually be broken down into two components: (i) introduction into the facility and (ii) the subsequent fate of *Listeria* in a facility, which includes rapid elimination (e.g., through sanitation) or survival subsequent to introduction (also often referred to as "persistence"). As indicated above, *Listeria* introduction can originate from a number of sources outside a processing facility; preventing introduction through fomites (e.g., people, equipment, etc.) is thus key part of a *Listeria* control programs. Without the proper zoning of equipment and employees, once introduced, *Listeria* can then be moved throughout the facility and can be introduced into a "niche" where it can evade cleaning and sanitation. A lack of

sanitary design and proper cleaning and sanitation programs can prevent the elimination of *L. monocytogenes* from the facility (thus facilitating "persistence"); hence persistence can typically be traced back to failures associated with pre-requisite programs (or non-process preventive controls under US FDA FSMA), such as sanitation. The importance of *L. monocytogenes* persistence in processing facilities has been defined though numerous studies *(16)*. For instance, Beno et al. *(4)*, isolated persistent *L. monocytogenes* PFGE types (PFGE types isolated on more than one sampling date) from the processing environments of 4 out of 9 small cheese processing facilities during a survey where environmental samples were collected monthly; 31 of the 57 *L. monocytogenes* strains subjected to PFGE in this study were persistent in a given facility. Furthermore, in a 2011 *L. monocytogenes* outbreak linked to cantaloupes in the United States *(6)*, samples of the brushes used to wash the cantaloupes in the packing house were discovered to be positive for *L. monocytogenes* of the same PFGE types as the strains causing illnesses, implicating the brushes as a likely persistent source of contamination in this outbreak *(27)*. These are two of the many examples which highlight the importance of strong food safety programs at the processing environment level that are designed to control the presence and persistence of *Listeria*.

Importantly, if the proper conditions exist, *L. monocytogenes* can survive in facilities for weeks to years. In fact, Orsi et al. *(30)* found *L. monocytogenes* persisted in a food processing facility for at least 12 years. During persistence, *L. monocytogenes* can be transferred to food contact surfaces and contaminate the product, or a harborage point may develop (a niche where *Listeria* is present and can continually contaminate product or other areas of the processing environment) within a food contact surface. Some common harborage points include product coolers, forklifts, forklift stops, hollow equipment legs, dead-end pipes, drains, floor-wall junctures, junctures between equipment legs and the floor, floor cracks, among other sites that are difficult to clean

and sanitize *(40)*.

Retail and Food Services. *L. monocytogenes* recalls and outbreaks have also been linked back to the retail and food service environment. As with processing facilities, *Listeria* presence in retail and food service environments is dependent upon (i) its introduction into the environment, and (ii) its ability to persist within the environment after it has been introduced. Compared with a processing environment, retail and food service environments are more open to outside environments, as there is limited control over what consumers bring into the retail space. Therefore, in addition to *Listeria* being introduced on raw materials or with employees, it can also be transported into the retail environment via customers. Hoelzer et al. *(20)* conducted a survey of 120 retail delis that were classified as small (<10 employees, N=60) or having failed an inspection (N=60) where they collected samples of food and non-food contact surfaces, including slicers, utensils, the deli case, floors, drains, and sinks, among others. The *L. monocytogenes* prevalence in these delis ranged from <6% (0/18) to 92% (11/12) *(20)*. Common sites positive for *L. monocytogenes* included milk crates, floors of walk-in coolers, and drains. As such, frequent and stringent cleaning and sanitation is required to eliminate *Listeria* as it is brought into the retail setting, including breaking down equipment (e.g., slicers) as far as possible prior to cleaning and sanitation to effectively eliminate *Listeria* within niches in the equipment. In addition, easy-to-clean equipment should be used when possible (e.g., no wooden utensils). Once introduced into a retail setting, *Listeria* persistence can develop, and cross-contamination of RTE products can occur (similar to what is observed in processing facilities) if *Listeria* is not eliminated through cleaning and sanitation.

RISKS

L. monocytogenes contamination of food products and food processing

environments can represent both a public health risk and a business (or enterprise) risk, which both need to be managed and minimized. Importantly, currently used approaches to control *L. monocytogenes* can use either risk or hazard-based approaches, however, risk-based approaches tend to be more impactful on reducing *L. monocytogenes* illnesses *(3)*.

Public Health Risks. *L. monocytogenes* contamination of RTE foods has been identified as the cause of a number of outbreaks across the world. The risk of a *L. monocytogenes* illness or an outbreak linked to a specific food product is affected by a combination of (i) the likelihood the food will become contaminated with *L. monocytogenes*, (ii) the food's ability to support the growth of *L. monocytogenes*, (iii) possible inactivation steps before consumption (e.g. cooking), (iv) susceptibility of the products' consumers, (v) the dose of L. monocytogenes consumed with the contaminated food, and (vi) the virulence of the *L. monocytogenes* strain(s) present in the food. The mean r (i.e., probability of a person becoming ill from one cell of *L. monocytogenes*) is estimated to range from 7.9×10^{-12} to 9.6×10^{-9} depending on the underlying conditions of the person (e.g., age, pregnancy, and other co-morbidities) *(34)*. Due to the low probability of illness from a single cell, it is not likely for a product to be contaminated with *L. monocytogenes* at a level high enough to cause illness without *L. monocytogenes* growth in the product. Therefore, the ability of *L. monocytogenes* to grow in a given product plays an important role in its ability to cause an infection. However, it is important to note that while a food can be inherently low risk due to the inability of *L. monocytogenes* to replicate in it, it is still possible for it to cause illnesses. In particular, the risk of illness increases when the state of the food product is changed in a way that allows it to support growth of *L. monocytogenes*. For instance, in 2014 there was a *L. monocytogenes* outbreak linked back to caramel apples, which caused 35 illnesses and 7 deaths *(7)*. Through an investigation into the cause of

the outbreak, it was discovered that the apples were contaminated with *L. monocytogenes* by the brush beds at the packing house. Then, when the apples were pierced with a stick and covered with caramel, the *L. monocytogenes* was provided with nutrients to allow it to grow. Therefore, while apples are not an inherently high-risk product (due to the containment of nutrients within their waxy skin), downstream processing of the apples increased their risk of causing illness. In addition, the importance of *Listeria* control in frozen vegetables and other frozen products is becoming more apparent. Frozen vegetables are not traditionally considered a RTE product, as consumers are generally instructed to cook these products prior to consumption. However, fruit and vegetable smoothies have become increasingly popular and are often prepared using frozen fruits and vegetables (e.g., berries, spinach, kale), which are generally not cooked prior to blending. While *L. monocytogenes* is unable to grow at freezing temperatures (<0° C), it can survive *(2)*. If these smoothies are not consumed immediately following preparation, but are left at temperatures that permit growth (e.g., room temperature, refrigeration temp) for sufficient time, *L. monocytogenes* can replicate to a level that increases the likelihood of causing an illness. Similarly, preparation of shakes or smoothies from ice cream with subsequent storage at temperatures that allow growth of *L. monocytogenes* converts a product that would be considered low risk (due to its inability to support *L. monocytogenes* growth) to a high risk product, as suspected in a recent outbreak in the US linked to ice cream *(8)*. As such, it is important to consider all potential uses of a product when designing a food safety program, which could include effective cooking labels and instructions as one component of a food safety plan, that could decrease both the public health and the business risk associated with that product.

In addition, when assessing the public health risk associated with different RTE products, one must consider the target consumers, as some individuals are at a higher

risk of infection, with old, pregnant, and immunocompromised individuals being particularly susceptible to systemic listeriosis. For instance, in the listeriosis outbreak linked to ice cream *(8)*, 5 of the 10 cases were reported to have consumed the ice cream as a milkshake while hospitalized for non-related ailments (i.e., they were immunocompromised), hence putting them at a higher of listeriosis *(8)*. This example illustrates how combinations of different factors can increase the public health risk of a product that may otherwise be considered low risk. Similarly, certain products may be consumed at an increased frequency by pregnant women or by immunocompromised individuals or cancer patients, who are also at an increased risk of listeriosis (some products may even be specifically targeted towards one of these groups). Hence, an assessment of the listeriosis associated public health risk of a RTE product should also consider potential product uses (with a focus on those that increase the potential for *L. monocytogenes* growth) as well as target consumers. The outcomes of these assessments may indicate the need for additional control strategies, such as specific labelling (e.g., detailed cooking instructions etc.).

Business Risks. While the public health risk (i.e., the risk of human listeriosis) associated with a product would typically be the driver of *L. monocytogenes* focused food safety efforts, individual firms may also want to assess the business and enterprise risks associated with *Listeria*. The predominant business risks associated with *Listeria* typically relate to (i) human disease cases and outbreaks linked to a product, and (ii) recalls due to detection of *L. monocytogenes* contamination in product or repeat *L. monocytogenes* (or possibly even repeat *Listeria*) detection in the processing environment. In this context, it is important to note that in a number of countries (e.g., US), any RTE product that tests positive for *L. monocytogenes* would be considered adulterated and hence would have to be recalled, even if the product represents an extremely low public health risk (e.g., sunflower seeds) and even if there are no

associated human disease cases. Therefore, *L. monocytogenes* may pose a reasonably high enterprise risk for some products that represent a limited public health risk, which could lead to situations where it may be prudent for firms to make considerable investments into *L. monocytogenes* control even for products where this organism represents a limited public health risk. Enterprise risks associated with *L. monocytogenes* detection differ considerably based on the regulatory environments. For example, in the United States, there is a so called "zero-tolerance" policy for *L. monocytogenes* in RTE foods. This means no detectable *L. monocytogenes* may be present in two 25g samples of US FDA regulated products and one 25g sample of USDA regulated products *(1)*. If *L. monocytogenes* is found in any RTE product in the United States, the product must be recalled (if product is in commerce), regardless of if there are any known illnesses traced back to the product. If the product is still under control of the company (i.e., not in commerce), the food must be reprocessed with a validated listericidal treatment, repurposed such that it will not be consumed by humans or animals, or destroyed. In addition, it must be determined if other product lots are also potentially contaminated regardless of if the products have entered commerce *(45)*. Alternatively, in the European Union, the regulation includes zero tolerance for *L. monocytogenes* in infant foods or medicines, while there are different criteria for RTE foods depending on the potential for *L. monocytogenes* growth. RTE food that are not able to support the growth of *L. monocytogenes* may have up to 100 cfu/g of product for the entirely of the product's shelf-life. However, for RTE foods that support growth of the bacterium, or RTE foods without data to prove the product's ability to limit *L. monocytogenes* growth to 100 cfu/g at the end of the product's shelf-life, *L. monocytogenes* must be absent in five 25g samples of the product at the time the product leaves the production facility *(14)*. In Canada, food products are grouped into two categories. Category 1 products are those that support the growth of *Listeria* and are

commonly implicated in outbreaks; for category 1 products, *L. monocytogenes* must be absent in a 125g sample of the product. Category 2 products are those that support limited to no growth of *L. monocytogenes*; for category 2 products, *L. monocytogenes* must be less than 100 CFU/g in five 10g samples *(19)*.

Quantification of the business risk associated with *L. monocytogenes* should take into account a number of different costs, including (i) costs of illnesses (which a company may be liable for); (ii) costs of product destruction or re-processing (if permitted), (iii) legal fees; (iv) loss of sales of destroyed product; and (v) and loss of future sales due to poor company reputation. The magnitude of the business risk associated with *L. monocytogenes* can be illustrated by the number of recalls due to *L. monocytogenes* contamination, or suspected contamination. For example, in the US alone there were 56 *L. monocytogenes* related recalls in 2019, including 17 recalls involving produce, 10 recalls involving sandwiches/ deli meats, 3 recalls due to smoked seafood, and 4 recalls due to dairy products *(46, 44)*.

CHALLENGES

The major challenges associated with *L. monocytogenes* control include (i) development and consistent implementation of programs that prevent *L. monocytogenes* introduction and persistence in food associated environments (e.g., processing facilities); (ii) management of raw material contamination in products that do not have an effective kill step (fresh produce, cold-smoked seafood, raw milk and dairy products); (iii) implementation of appropriate root cause analysis procedures that allow for identification of sources of environmental and product contamination; and (iv) appropriate use of subtyping methods, including WGS, and appropriate interpretation of the resulting data.

Development and Consistent Implementation of Programs that Prevent *L.*

***monocytogenes* Introduction and Persistence in Food Associated Environments**. A continued main challenge industry faces is the development and consistent implementation of programs that (i) prevent *L. monocytogenes* introduction and (ii) prevent *L. monocytogenes* persistence in food associated environments (with persistence being defined as survival of a specific *Listeria* subtype in a processing facility over time, see below for details). A related challenge is for regulatory agencies to develop, implement, and enforce regulations that encourage industry to develop and implement stringent programs that minimize the likelihood of *L. monocytogenes* introduction and persistence (for example, regulatory agencies may want to eliminate negative consequences associated with *Listeria* detection in a processing environment to encourage processors to find *Listeria* if it is present). Some regulations or practices may, unintentionally, provide incentives for companies to not implement stringent environmental and finished product testing strategies. For instance, this may be the case if consequences of positive test results (which are expected due to the high prevalence of *L. monocytogenes*) are not commensurate to public health risk (e.g., recalls or other severe regulatory consequences due to finding low levels of *L. monocytogenes* in finished products that do not support its growth or finding *Listeria* spp. in drains in processing facilities).

Minimizing *L. monocytogenes* introduction from outside environments is particularly challenging due to the high prevalence of *L. monocytogenes* in many different environments, as detailed above. Achieving "zero" introduction of *L. monocytogenes* into processing plants is essentially impossible, particularly if raw materials, which in most cases would have to be expected to at least occasionally be contaminated, are introduced in a facility. Key strategies to minimize introduction of *L. monocytogenes* include GMPs (e.g., employees should wear clean coats designated for use only within the processing area, wearing gloves), captive footwear policies, foot

baths, door foamers, regular cleaning and sanitation of trailers used to transport raw materials, regular cleaning and sanitation of forklifts, and regular cleaning and sanitation of creates and bins (including trash bins) that carry materials inside and outside of the facility. However, it is important to note that interventions that introduce additional moisture into the facility (e.g., door foamers and foot baths) can facilitate *L. monocytogenes* growth and survival if not properly maintained (e.g., if appropriate sanitizer concentrations are not consistently maintained, leading to wet areas and moisture). Verifying sanitizer concentrations as well as testing the areas around foot baths and foamers for *Listeria* presence can be useful in identifying a lack of *Listeria* control.

Prevention or management of *L. monocytogenes* persistence in food facilities is a well-documented issue for the industry. Persistent *Listeria* refers to the *Listeria* that remains in the processing environment for an extended time and is able to survive cleaning and sanitation. Once introduced into the processing environment, *Listeria* can enter niches within the equipment or building infrastructure where cleaners and sanitizers are not able to reach and eliminate the *Listeria*, allowing it to become persistent. In comparison to persistent *Listeria*, there also "transient" *Listeria*. Transient *Listeria* refers to *Listeria* introduced into the processing environment but removed during regular cleaning and sanitation activities. Since *Listeria* is prevalent in a variety of environments, it is expected for *Listeria* to enter the processing environment on occasion. As long as *L. monocytogenes* is quickly (e.g., by the end of a one-day shift) removed by cleaning and sanitation and not allowed to survive in a niche within the processing environment, it is unlikely to pose a substantial public health or business risk, making it less of a problem compared to persistent *L. monocytogenes*.

An effective *Listeria* sampling program is key to identifying *Listeria* presence and persistence. For any RTE foods that are exposed to the processing plant

environment, appropriate testing programs need to include a robust environmental monitoring program and may also include finished product testing, although often at substantially lower frequencies than environmental monitoring. Finished product testing, particularly if conducted in the absence of a strong environmental monitoring program, is of limited values as *L. monocytogenes* is often present sporadically and at low levels, which can make it difficult to identify contaminated product via final product testing. On the other hand, environmental monitoring programs often allow for early detection of potential sources and routes of contamination. A key challenge with environmental monitoring programs however is that many lack clear and defined goals, such as verification and validation of *Listeria* control strategies. For instance, routine environmental monitoring programs can be used for verification of *Listeria* control strategies (e.g., cleaning and sanitation). Another important consideration for *Listeria* sampling programs is whether to test for *Listeria* spp. or *L. monocytogenes*. When performing finished product testing, samples should always be tested for *L. monocytogenes*, as a *Listeria* spp. positive result in a finished product would require further speciation. When performing environmental monitoring, testing for *Listeria* spp. generally allows for a more comprehensive testing strategy, as several non-pathogenic species of *Listeria* often inhabit similar environments as *L. monocytogenes* and act as index organisms. Additional information on environmental monitoring programs can be found in the further reading section (e.g., 3M, 2018). While strong food safety programs, including environmental monitoring programs, are typically expensive, they provide a significant return on investment if they facilitate identification and elimination of *Listeria* within the processing environment before detection by regulatory agencies or before a public health issue.

In order to identify persistent *Listeria* (and differentiate them from transient *Listeria*), subtyping (e.g., PFGE or WGS) should be used to determine which subtype

or strain of *Listeria* is present. If the same or related subtypes are found over time, it is an indication of persistent *Listeria* or continuous introduction of the same subtype. Identifying if a subtype is persistent or continuously being introduced into the environment can be difficult. However, sampling a suspected piece of equipment before and after cleaning and sanitation, then determining if the same subtype of *Listeria* is present at both times can provide evidence of persistence. Proper food safety programs should be put in place to protect against persistent *Listeria*. These programs must emphasize proper sanitary design of equipment (i.e., elimination of areas within the equipment or facility infrastructure that are difficult to clean and sanitize); a one-directional flow of employees, equipment, and food products through the processing area; and proper cleaning and sanitation programs, including pre-cleaning disassembly of equipment to a level that allows for effective elimination of *Listeria* from niches through cleaning and sanitation.

In addition to persistent *Listeria* and transient *Listeria*, there is a scenario that can be labeled "persistent transient *Listeria*," we use this term to refer to a continual introduction of one or more *Listeria* subtypes into the processing environment. While this scenario may not be as much of a concern as persistent *Listeria*, the continuous introduction of *Listeria* also indicates a lack of proper *Listeria* control. Persistent transient *Listeria* are likely to be introduced into the processing environment with raw materials, crates, employees, among other routes. In order to reduce the prevalence of persistent transient *Listeria*, more frequent cleaning and sanitation, improved supplier verification, and additional controls to prevent employees from tracking *Listeria* into the processing environment (e.g., captive footwear programs and boot washes at the entrance of processing areas) should be considered.

Management of Raw Material Contamination in Products that Do Not Have an Effective Kill Step. While *Listeria* can be easily destroyed by heat, there are

a number of products and raw materials which do not receive an effective kill step during processing. Produce, cold-smoked seafood products, and raw milk dairy products represent examples of commonly consumed RTE products that do not include a kill a step as part of the process. For these products in particular, robust supplier verification programs for raw materials are essential to reduce the likelihood that raw materials lead to contamination of the final product. If suppliers are unable to prove effective control of *Listeria*, including in their processing environment if applicable (e.g., through environment monitoring programs), alternative suppliers should be identified. In addition, non-thermal and thermal heat treatments that reduce *L. monocytogenes* (although not at the level of a "kill step", which is typically defined as a 5-log reduction) can be used. For example, cold-smoked seafood producers may use antimicrobial washes of incoming raw materials, or the addition of antimicrobial treatments (e.g., nisin) to final products to reduce *Listeria* levels and growth *(24)*. Produce packinghouses and fresh-cut facilities may also use antimicrobials in wash water to reduce cross-contamination between produce items *(17)*.

Implementation of Appropriate Root Cause Analysis Procedures that Allow for Identification of Sources of Environmental and Product Contamination. Since *L. monocytogenes* is prevalent across a variety of different environments, identifying the source of *L. monocytogenes* found in finished product or the environment can be difficult (e.g., is it coming in on a raw material, or where within the processing environment is *L. monocytogenes* present or persisting). While a formal well-defined root cause analysis (RCA) approach provides one of the most effective ways to define the root cause of a *Listeria* "issue," implementing good RCA procedures remains a challenge for many companies. RCA is a strategy which aims to identify the true or initial cause of a final event, such that without this initial cause the final event could not occur. Therefore, using RCA pivots corrective actions from being responsive in nature

to being preventive (i.e., with proper RCA similar problems will be prevented from happening in the future). In particular, RCA can be used to provide a more systematic method for identifying and controlling the presence and persistence of *Listeria* in the food supply chain. In order to perform an RCA, a multidisciplinary team (e.g., someone from quality and food safety, maintenance, and operations) should be formed. Due to the complexity of many food safety problems, getting input from a diverse set of thinkers can help to identify novel causes, as well as innovative corrective actions. Once assembled, the team should clearly define the problem and discuss what information is needed to help solve the problem; this information or data should then be gathered. There are a variety of techniques that can then be used to identify root causes, including fishbone diagrams, the "five whys" technique, change analysis, fault tree analysis, among others. Each technique can be used on its own or they can be used in combination and each technique may be optimal in different situations *(43)*. Fishbone diagrams, in particular, can be helpful when laying out all components of a problem. Then using the "5 whys" technique (i.e., continually asking why some event or practice occurred or is the way it is until reaching the root cause) can be useful when identifying the root cause associated with each bone of the diagram that has been deemed important. For instance, if a persistent *Listeria* subtype is present in a given processing environment, RCA should be performed to (i) identify the source of the persistent *L. monocytogenes*, (ii) eliminate this persistent *L. monocytogenes*, and (iii) identify a control strategy that will prevent a similar persistent *Listeria* in the future. Figure 2.1 lays out an example of a fishbone diagram, which can be used as a starting point for identifying potentially important preliminary causes. In this fishbone diagram, there are six major bones: company practices/ food safety culture, personnel, facilities, cleaning and sanitation, raw material introduction, and processing equipment. Each of these major bones represent an over-arching category where many food safety, or *Listeria* persistence,

problems originate. Using this fishbone diagram as a template, RCA can be performed; Figure 2.2 outlines a potential procedure for performing RCA.

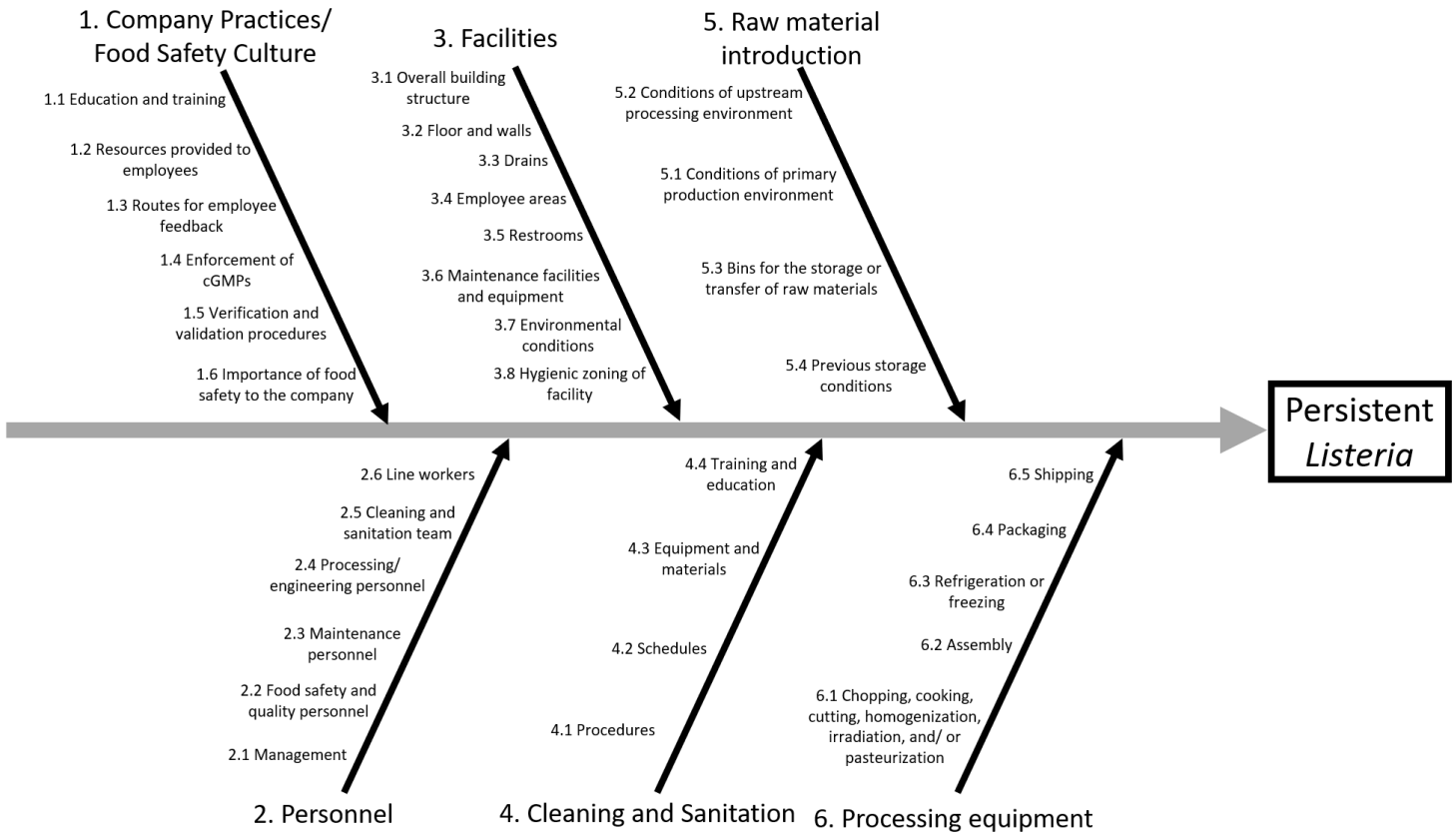


Figure 2.1. Example of a fishbone diagram to identify the cause of *Listeria* persistence within a food processing environment.

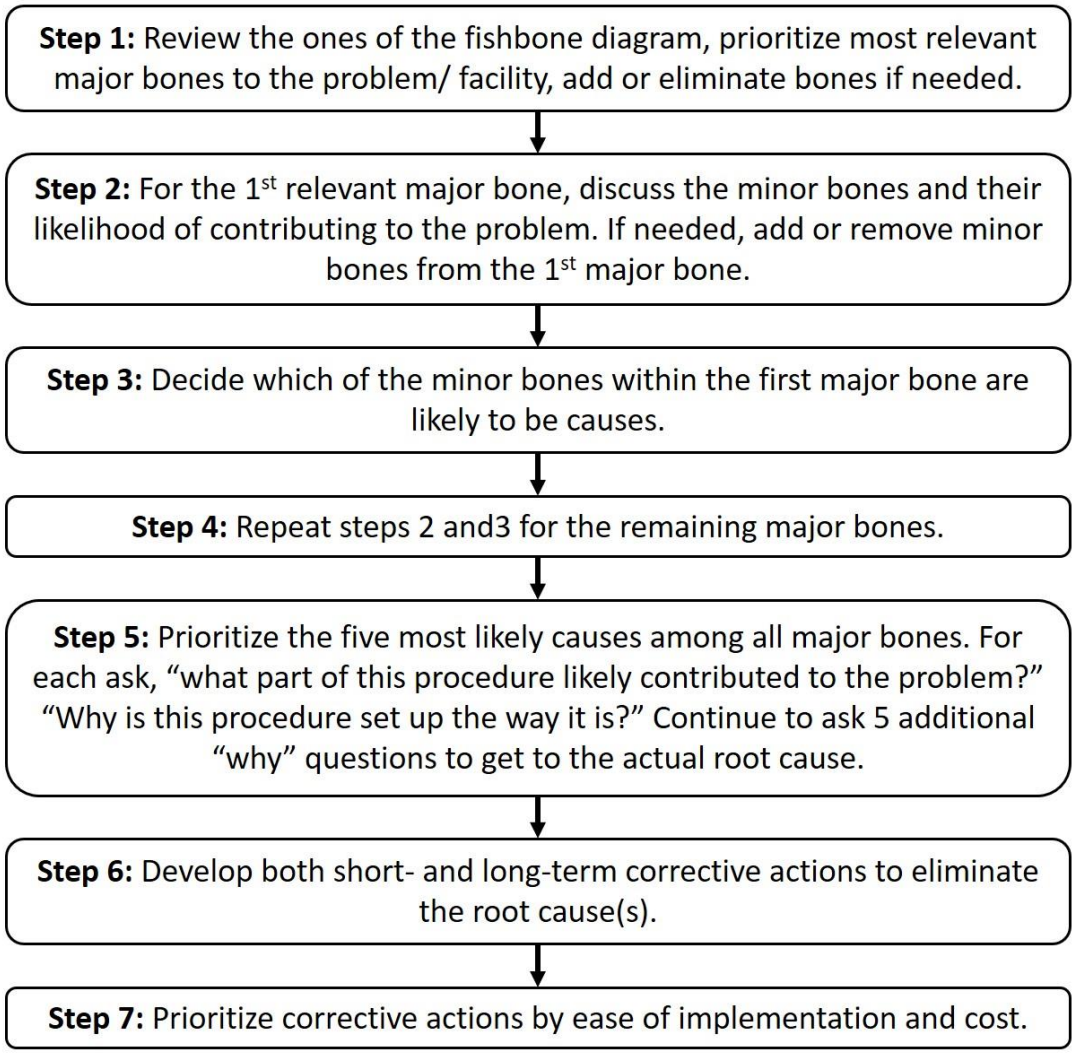


Figure 2.2. Steps for conducting a root cause analysis. Major bones refer to the six broad categories in the fishbone diagram in Figure 2.1. Minor bones refer to the sub-categories within each of the six major bones.

An RCA for identifying the source of contamination or persistent *Listeria* is typically part of a "for-cause" investigation that also includes efforts to gather sufficient data to facilitate the RCA. This type of "for-cause" investigation typically requires an intensified sampling of the implicated parts of the processing environment to identify contaminated sites and contamination sources; this type of sampling often involves collection of 100s to 1,000s of samples. While collection of environmental samples is typically key for *Listeria* RCA, raw material and finished product testing can also be useful and needed. Overall sampling as part of RCAs often represents an iterative process where the RCA identifies possible root causes that require sampling for confirmation (or exclusion), often followed by additional sampling to guide further discussions on the root cause and potential corrective actions to eliminate and prevent similar contamination problems in the future.

Appropriate Use of Subtyping Methods, Including Whole Genome Sequencing, and Appropriate Interpretation of the Resulting Data. In many parts of the world, subtyping ("DNA fingerprinting") methods are increasingly used as part of efforts to manage *L. monocytogenes*. These methods may be used by either (i) an individual company or (ii) regulatory and public health agencies. For example, individual companies may perform subtyping of all *L. monocytogenes* or all *Listeria* spp. isolates that are obtained as part of their routine environmental monitoring programs. While still uncommon, this practice is increasingly used, particularly in locations where regulatory agencies also perform subtyping on *L. monocytogenes* or *Listeria* spp. that may be obtained by an agency as part of either a routine inspection or for-cause investigations. Routine subtyping of all isolates helps companies to identify persistent contamination, particularly if positive test results are only sporadically obtained and subtyping is needed to determine whether two positive samples represent contamination with the same *Listeria* or independent events. In addition, some

companies do not perform routine subtyping but may perform subtyping only as part of investigations and RCA efforts. Many companies experience challenges with the use of molecular subtyping methods, including (i) the decisions of whether and when to perform subtyping; (ii) the decision of which subtyping method to use, and (iii) performing and interpreting the data analysis (particularly for WGS). The decisions of whether and when to perform subtyping are complex and involve a number of considerations (e.g., regulatory climate, food safety budget, history of *Listeria* issues, etc.), but companies with a strong food safety culture increasingly use these tools. Companies that require advanced information to successfully identify the root cause of a *Listeria* "issue" also typically use subtyping. As for selection of subtyping methods, commonly used methods include ribotyping, PFGE, and WGS; while industry often still uses methods such as PFGE and ribotyping, due to typically lower costs and a reduced need for highly technical data analysis, WGS is being increasingly used by industry *(23)*.

With regard to subtyping use by government agencies across the world, public health and regulatory agencies increasingly use WGS to characterize human and food-associated *L. monocytogenes* isolates. For example, in the United States, WGS is performed on all human *L. monocytogenes* isolates as well as on any isolates from foods and food processing environments obtained by either the Food and Drug Administration (US FDA) or the USDA-FSIS. These WGS data are uploaded into the NCBI pathogen detection database (https://www.ncbi.nlm.nih.gov/pathogens/) and hence are publicly available, even though the metadata provided do not typically allow for identification of the facility an isolate was obtained from. As part of this process, isolates from foods and food processing environments are also clustered with closely related human isolates and other isolates from foods and food processing facilities. This clustering, and subsequent follow up genome comparisons, can be used to (i) identify possible human

cases that may be linked to a product or facility (providing hypothesis for subsequent epidemiological investigations) or (ii) possible instances where a specific strain may persist in an environment. However, these analyses do not identify definitive linkages, but rather WGS data need to be interpreted in conjunction with epidemiological data. Industry often struggles with these data analyses, particularly since they often need to be performed and interpreted rapidly to make correct decisions on recalls, recall scopes, and other matters with considerable public health and business impact. While some guidance documents and reviews on interpretation of WGS have been published *(23, 33)*, appropriate interpretation of WGS data and associated decision making are not trivial and should typically conducted in consultation with experts to avoid costly errors and misinterpretations.

CONCLUSIONS

Listeria spp. and *L. monocytogenes* are prevalent in a variety of environments, including natural and urban environments as well as primary production, processing, and retail environments, among others. As such, there are a variety of points along the supply chain where RTE food products can become contaminated with *L. monocytogenes*. While *L. monocytogenes* poses substantial public health risks, it also poses business risks, including when it is found in RTE products that represent a low risk of human disease (e.g., products that do not support *L. monocytogenes* growth [*15*]). As such, risk-based stringent *Listeria* control programs should be implemented, which include GMPs; regular cleaning and sanitation programs; the use of equipment with sanitary designs (i.e., equipment without niches); and one-directional flow of employees, equipment, and food products. In addition, food products and their processing environments should be monitored for *Listeria* presence and persistence; these environmental monitoring programs need to be linked to specific goals (e.g.,

verification of certain food safety programs, such as sanitation programs). Furthermore, robust supplier verification programs and non-thermal antimicrobial treatments are especially important for RTE products produced without a kill step. The large number of potential sources of *Listeria* throughout the food supply chain make it difficult to identify the true source of contamination when detected in the environment or products. The use of formal and well executed root cause analysis and "for-cause" investigations and subtyping tools are hence essential in investigations of *Listeria* positives in order to not only address the specific issue at hand, but also to create and implement control measures to prevent similar events from occurring in the future.

FURTHER READING

3M (2018). *Environmental Monitoring Handbook for the Food and Beverage Industries*, 1st Edition. Minnesota: *3M*.

US Department of Agriculture (2014). FSIS compliance guideline: Controlling *Listeria monocytogenes* is post-lethality exposed ready-to-eat meat and poultry products. Accessed 10/1/2020. https://www.fsis.usda.gov/wps/wcm/connect/d3373299-50e6-47d6-a577-e74a1e549fde/Controlling-Lm-RTE-Guideline.pdf?MOD=AJPERES

REFERENCES

1. Archer, D. L. (2018). The evolution of FDA's policy on *Listeria monocytogenes* in ready-to-eat foods in the United States. *Current Opinions in Food Science* 20: 64-68.

2. Azizoglu, R. O., Osborne, J., Wilson, S., et al. (2009). Role of growth temperature in freeze-thaw tolerance of *Listeria* spp. *Applied and Environmental Microbiology* 75(16): 5315-5320.

3. Barlow, S. M., Boobis, A. R., Bridges, J., et al., (2015). The role of hazard- and risk-based approaches in ensuring food safety. *Trends in Food Science and Technology* 45:176-188.

4. Beno, S. M., Stasiewicz, M. J., Andrus, A. D., et al. (2016). Development and validation of pathogen environmental monitoring programs for small cheese processing facilities. *Journal of Food Protection* 79(12):2095-2106.

5. Buchanan, R. L., Gorris, L. G. M., Hayman, M. M., et al. (2017). A review of *Listeria monocytogenes*: An update on outbreaks, virulence, dose-response, ecology, and risk assessments. *Food Control* 75:1-13.

6. Centers for Disease Control and Prevention (2012). Multistate outbreak of listeriosis linked to whole cantaloupes from Jensen Farms, Colorado (final update). Accessed 9/10/2020. https://www.cdc.gov/listeria/outbreaks/cantaloupes-jensen-farms/index.html

7. Centers for Disease Control and Prevention (2015a). Multistate outbreak of listeriosis linked to commercially produced, packaged caramel apples made from Bidart Bros apples (final update). Accessed 9/10/2020. https://www.cdc.gov/listeria/outbreaks/caramel-apples-12-14/index.html

8. Centers for Disease Control and Prevention (2015b). Multistate outbreak of listeriosis linked to blue bell creameries products (final update). Accessed 9/10/2020. https://www.cdc.gov/listeria/outbreaks/ice-cream-03-15/index.html

9. Centers for Disease Control and Prevention (2016a). *Listeria* (listeriosis): symptoms. Accessed 9/10/2020. https://www.cdc.gov/listeria/symptoms.html

10. Centers for Disease Control and Prevention (2016b). Multistate outbreak of listeriosis linked to raw milk produced by Miller's Organic Farm in Pennsylvania (final update). Accessed 9/10/2020. https://www.cdc.gov/listeria/outbreaks/raw-milk-03-16/index.html

11. Centers for Disease Control and Prevention (2019). Whole genome sequencing. Accessed 9/28/2020. https://www.cdc.gov/ncezid/dfwed/keyprograms/tracking-foodborne-illness-wgs.html

12. Chapin, T. K., Nightingale, K. K., Worobo, R. W., et al. (2014). Geographical and meteorological factors associated with isolation of *Listeria* species in New York State produce production and natural environments. *Journal of Food Protection* 77: 1919-1928.

13. European Centre for Disease Prevention and Control (2018). Monitoring the use of whole genome sequencing in infectious disease surveillance in Europe, 2015-2017. Accessed 10/2/2020. https://www.ecdc.europa.eu/sites/default/files/documents/monitoring-WGS-infectious-disease-surveillance-in-Europe-2015-2017-updated-Dec-2018.pdf

14. European Commission (2005). COMMISSION REGULATION (EC) No 2073/2005 of 15 November 2005 on microbiological criteria for foodstuffs. Accessed 11/3/2020. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32005R2073&from=EN

15. Farber, J. M., Zwietering, M., Wiedmann, M., et al. (2020). Alternative approaches to the risk management of *Listeria monocytogenes* in low risk foods. *Food Control* 107601.

16. Ferreira, V., Wiedmann, M., Teixeira, P., et al. (2014). *Listeria monocytogenes* persistence in food-associated environments: Epidemiology, strain characteristics, and implications for public health. *Journal of Food Protection* 77(1): 150-170.

17. Gil, M. I, Selma, M. V., Lopez-Galvez, F., and Allende, A. (2009). Fresh-cut product sanitation and wash water disinfection: Problems and solutions. *International Journal of Food Microbiology* 134(1-2):37-45.

18. Golden, C. E., Rothrock, M. J. Jr, and Mishra, A. (2019). Comparison between random forest and gradient boosting machine methods for predicting *Listeria* spp.

prevalence in the environment of pastured poultry farms. *Food Research International* 122: 47-55.

19. Health Canada (2011). Policy on *Listeria monocytogenes* in ready-to-eat foods. Accessed 11/3/2020. https://www.canada.ca/content/dam/hc-sc/migration/hc-sc/fn-an/alt_formats/pdf/legislation/pol/policy_listeria_monocytogenes_2011-eng.pdf

20. Hoelzer, K., Sauders, B. D., Sanchez, M. D., et al. (2011). Prevalence, distribution, and diversity of *Listeria monocytogenes* in retail environments, focusing on small establishments and establishments with a history of failed inspections. *Journal of Food Protection* 74: 1083-1095.

21. Hoelzer, K., Pouillot, R., and Dennis, S. (2012). *Listeria monocytogenes* growth dynamics on produce: A review of the available data for predictive modeling. *Foodborne Pathogens and Disease* 9(7):661-673.

22. ILSI Research Foundation/ Risk Science Institute, Expert Panel on *Listeria monocytogenes* in Foods (2005). Achieving continuous improvement in reductions in foodborne listeriosis – A risk-based approach. *Journal of Food Protection* 68(9):1932-1994.

23. Jagadeesan, B., Gerner-Smidt, P., Allard, M. W., et al. (2019). The use of next generation sequencing for improving food safety: Translation into practice. *Food Microbiology* 79:96-115.

24. Jahncke M. L., Collette, R., Hicks, D., et al. (2004). Treatment options to eliminate or control growth of *Listeria monocytogenes* on raw materials and on finished product for the smoked seafood industry. *Food Protection Trends* 24(8):612-619.

25. Maertens de Noordhout, C., Devleesschauwer, B., Angulo, F. J., et al. (2014). The global burden of listeriosis: A systematic review and meta-analysis. *Lancet Infectious Diseases* 14(11): 1073-1082.

26. McClure, P. J., Roberts, T. A., and Otto Oguru, P. (1989). Comparison of the effects of sodium chloride, pH and temperature on the growth of *Listeria monocytogenes* on gradient plates and in liquid medium. *Letters in Applied Microbiology* 9: 95-99.

27. McCollum, J. T., Cronquist, A. B., Silk, B. J. (2013). Multistate outbreak of listeriosis associated with cantaloupe. *The New England Journal of Medicine* 369(10): 944-953.

28. Moura, A., Tourdjman, M., Leclercq, A., et al. (2017). Real-time whole-genome sequencing for surveillance of *Listeria monocytogenes*, France. *Emerging Infectious Disease* 23(9):1462-1470.

29. Nightingale, K. K., Schukken, Y. H., Nightingale, C. R., et al. (2004). Ecology and transmission of *Listeria monocytogenes* infecting ruminants and in the farm environment. Applied and Environmental Microbiology 70(8): 4458-4467.

30. Orsi, R. H., Borowsky, M. L., Lauer, P., et al. (2008). Short-term genome evolution of *Listeria monocytogenes* in a non-controlled environment. *BMC Genomics* 9(539): 1-17.

31. Pang, H., McEgan, R., Mishra, A., et al. (2017). Identifying and modeling meteorological risk factors associated with preharvest contamination of *Listeria* species in a mixed produce and dairy farm. *Food Research International* 102: 355-363.

32. Park, S., Szonyi, B., Gautam, R., et al. (2012). Risk factors for microbial contamination in fruits and vegetables at the preharvest level: A systematic review. *Journal of Food Protection* 75: 2055-2081.

33. Pightling, A. W., Pettengill, J. B., Luo, Y., et al. (2018). Interpreting whole-genome sequencing analysis of foodborne bacteria for regulatory applications and outbreak investigations. *Frontiers in Microbiology* 9:1482.

34. Pouillot, R., Hoelzer, K., Chen, Y., et al. (2015). Listeria monocytogenes dose response revisited – incorporating adjustments for variability in strain virulence and host susceptibility. Risk Analysis 35(1): 90-108.

35. Public Health Agency of Canada (2019). The public health agency of Canada's experience with whole genome sequencing transition: Success and pitfalls. PulseNet/ OutbreakNet West Coast Regional Meeting, San Diego, CA.

36. Sauders, B. D., Pettit, D., Currie, B., et al. (2005). Low prevalence of *Listeria monocytogenes* in human stool. *Journal of Food Protection* 68(1):178-181.

37. Sauders, B. D., Overdevest, J., Fortes, E., et al. (2012). Diversity of *Listeria* species in urban and natural environments. *Applied and Environmental Microbiology* 78(12): 4420-4433.

38. Scallan, E., Hoekstra, R. M., Angulo, F. J., et al. (2011). Foodborne illness acquired in the United States- major pathogens. *Emerging and Infectious Disease* 17: 7-15.

39. Schlech, W. F. III, Lavigne, P. M., Bortolussi, R. A., et al. (1983). Epidemic listeriosis – evidence for transmission by food. *New England Journal of Medicine* 308:203-206.

40. Simmons, C. K. and Wiedmann, M. (2018). Identification and classification of sampling sites for pathogen environmental monitoring programs for *Listeria monocytogenes*: Results from expert elicitation. *Food Microbiology* 75:2-17.

41. Strawn, L. K., Fortes, E. D., Bihn, E. A., et al. (2013a). Landscape and meteorological factors affecting prevalence of three food-borne pathogens in fruit and vegetable farms. *Applied and Environmental Microbiology* 79: 588-600.

42. Strawn, L. K., Grohn, Y. T., Warchocki, S., et al. (2013b). Risk factors associated with *Salmonella* and *Listeria monocytogenes* contamination of produce fields. *Applied and Environmental Microbiology* 79: 7618-7627.

43. The PEW Charitable Trust (2020). A guide for conducting a food safety root cause analysis: Approaches for investigating contamination incidents and preventing recurrence. Accessed 10/2/2020. https://www.pewtrusts.org/en/research-and-analysis/reports/2020/03/a-guide-for-conducting-a-food-safety-root-cause-analysis

44. US Department of Agriculture (2020). Recall case archive. Accessed 11/3/2020. https://www.fsis.usda.gov/wps/portal/fsis/topics/recalls-and-public-health-alerts/recall-case-archive

45. US Food and Drug Administration (2017). Control of *Listeria monocytogenes* in ready-to-eat foods: Guidance for industry. Accessed 9/28/2020. https://www.fda.gov/media/102633/download

46. US Food and Drug Administration (2020). Recalls, market withdrawals, and safety alerts. Accessed 11/3/2020. https://www.fda.gov/safety/recalls/

47. Walker, S. J., Archer, P., and Banks, J. G. (1990). Growth of *Listeria monocytogenes* at refrigeration temperatures. *Journal of Applied Bacteriology* 68: 157-162.

48. Vilar, M. J., Yus, E., Sanjuan, M. L., et al. (2007). Prevalence of and risk factors for *Listeria* species on dairy farms. *Journal of Dairy Science* 90: 5083-5088.

49. Weller, D., Wiedmann, M., and Strawn, L. K. (2015a). Irrigation is significantly associated with an increased prevalence of *Listeria monocytogenes* in produce production environments in New York State. *Journal of Food Protection* 78: 1132-1141.

50. Weller, D., Wiedmann, M., and Strawn, L. K. (2015b). Spatial and temporal factors associated with an increased prevalence of *Listeria monocytogenes* in spinach fields in New York State. *Applied and Environmental Microbiology* 81: 6059-6069.

CHAPTER 3

SMALL PRODUCE FARM ENVIRONMENTS CAN HARBOR DIVERSE *LISTERIA MONOCYTOGENES* AND *LISTERIA* SPP. POPULATIONS

Published in: Journal of Food Protection (2021), 84 (1): 113–121

ABSTRACT

A comprehensive understanding of foodborne pathogen diversity in pre-harvest environments is necessary to effectively track pathogens on farms and identify sources of produce contamination. As such, this study aimed to characterize *Listeria* diversity in wildlife feces and agricultural water collected from a New York State produce farm over a growing season. Water samples were collected from a pond (N=80) and stream (N=52). Fecal samples (N=77) were opportunistically collected from areas <5m from the water sources; all samples were collected from a <0.5km^2 area. Overall, 41% (86/209) and 24% (50/209) of samples were positive for *Listeria monocytogenes* and *Listeria* spp. (excluding *L. monocytogenes*), respectively. For each positive sample, one *L. monocytogenes* or *Listeria* spp. isolate was speciated by sequencing the *sigB* gene, which allowed for additional characterization based on the *sigB* allelic type (AT). The 86 *L. monocytogenes* and 50 *Listeria* spp. isolates represented 8 and 23 different ATs, respectively. A subset of *L. monocytogenes* isolates (N=44) from pond water and pond-adjacent feces (representing a ~5,000m^2 area) were further characterized by PFGE; these 44 isolates represented 22 PFGE types, which is indicative of considerable diversity at a small spatial scale. Ten PFGE types were isolated more than once, suggesting persistence or re-introduction of PFGE types in this area. Given the small spatial scale, the prevalence of *L. monocytogenes* and *Listeria* spp., as well as the considerable diversity amongst isolates, suggests traceback investigations may be

challenging. For example, traceback of finished product or processing facility contamination with specific subtypes to pre-harvest sources may require collection of large sample sets, and characterization of a considerable number of isolates. Our data also support the adage, “absence of evidence does not equal evidence of absence” applies to *L. monocytogenes* traceback efforts at the pre-harvest level.

HIGHLIGHTS

- There is considerable *Listeria* diversity in the farm environment investigated.
- *Listeria* subtypes were re-introduced or persisted over the growing season.
- Four *L. monocytogenes* PFGE types were shared between feces and pond samples.

INTRODUCTION

It is estimated that *Listeria monocytogenes* is responsible for approximately 1,600 illnesses, 1,455 hospitalizations, and 255 deaths annually in the United States (*1*), and several high-profile outbreaks and recalls have been traced to *L. monocytogenes* contamination of fresh produce (*8, 24, 44*). For example, a 2011 outbreak was traced back to *L. monocytogenes* contamination of cantaloupes that resulted in 147 illnesses and 33 deaths (*2*). Additionally, there have been >30 recalls since 2016 due to *L. monocytogenes* contamination of produce and ready-to-eat produce products (*36*). While the majority of produce outbreaks and recalls due to *L. monocytogenes* contamination are traced back to the processing environment, *L. monocytogenes* can be introduced, or re-introduced, in the processing environment from the pre-harvest environment [e.g., on produce, in soil, on harvesting equipment, or on workers shoes; (*32*)] or could be introduced onto product in the field and remain on product until consumption.

Following detection of a *Listeria*-positive sample during routine monitoring or an outbreak investigation, processors or investigators often conduct extensive sampling to identify potential contamination sources (i.e., traceback analysis). Identifying how a *Listeria* isolate was introduced onto product or into the processing environment is a critical step in performing a root cause analysis and establishing effective corrective actions (to control potential contamination pathways). To determine if an isolate originated in the pre-harvest environment, investigators and operators must collect environmental samples from farms that supply the product. However, the ubiquity of *Listeria* in the environment (*1, 7, 12, 11*) may make it challenging to prove a causal relationship between *Listeria* isolated from the pre-harvest environment and *Listeria* isolated from post-harvest environments and/or product. Previous studies of pre-harvest environments isolated *Listeria* from soil (*3, 25, 35, 34, 48, 49*), wildlife feces (*5, 19, 37,*

49), agricultural water (*1, 6, 17, 20, 33*), and biological soil amendments (*25*). As such, a more comprehensive understanding of the prevalence, diversity, and distribution of *Listeria* on produce farms is needed to better understand on-farm sources of *Listeria*, understand transfer of *Listeria* between sample types on produce farms, and facilitate traceback analysis. While some studies have investigated the diversity of *Listeria* in pre-harvest produce environments (*3, 19, 20, 33, 34, 35*), these studies have examined diversity over large geographic areas. For instance, Chapin et al. (*3*) and Strawn et al. (34) investigated *Listeria* diversity on 5 New York State (US) farms located up to 205 km from each other, while Stea et al. (*33*) examined *Listeria* diversity within an agricultural watershed in Nova Scotia, Canada with a drainage area of 135 km^2. As there is limited data on *Listeria* diversity at small spatial scales (e.g., within individual farms), we aimed, in this study, to address this knowledge gap by examining the diversity of *Listeria* on a single produce farm.

Agricultural water and wildlife intrusion in produce fields have been identified as key sources and pathways for dispersal of *Listeria* contamination in pre-harvest environments (*14*, *26*, *37*, *48*). Multiple studies have isolated *L. monocytogenes* from agricultural water and wildlife feces collected in produce growing environments (6, *7*, *19*, *25*, *34*, *48*). For instance, Weller et al. (*48*) isolated *L. monocytogenes* from 45% (N=11) of wildlife feces collected on 10 produce farms in New York State. Falardeau et al. (*7*) isolated *L. monocytogenes* from 10% (N=223) of agricultural water samples collected from 2 watersheds in British Colombia, Canada. Furthermore, Cooley et al. (*6*) isolated *L. monocytogenes* from 43% (N=1405) of water samples collected from the California Central Coast agricultural region. As such, the primary objective of the study reported here was to characterize and compare the *L. monocytogenes* and *Listeria* spp., excluding *L. monocytogenes*, diversity in agricultural water and wildlife feces collected from a single produce farm in New York State.

MATERIALS AND METHODS

Study design and sample collection. The study reported here was conducted between May and July 2014 on a produce farm in the Finger Lakes region of New York State. Two separate irrigation water sources are located on the farm: (i) a pond and (ii) a stream. The water sources are within a rectangular 0.5km^2 (124 acres) area; the sampling sites in the pond and stream are approx. 775 m (2,546 ft) apart.

Agricultural water and fecal samples were collected on 26 different days. Fifty-two and 80 water samples were collected from the stream and pond, respectively; the *Listeria* prevalence and number of *sigB* allelic types for the 52 stream samples and the 13 fecal samples collected near the stream were previously reported by Weller et al. (*49*). However, in order to better understand the *Listeria* diversity on the studied farm, this data was also included here. All water samples were collected as described in Weller et al. (*49*). Briefly, 250 mL of water were collected into a sampling cup (Nalgene, Rochester, NY) using a sampling pole and stored on ice until processing. Fecal samples (N=77) located within 5m of the pond (N=64) or stream (N=13) were collected opportunistically, as previously described (*49*); data for the fecal samples collected near the pond were not previously reported. A 5m buffer around the pond and stream was used for practical reasons. All samples were stored at 4°C and processed within 3h of collection.

Listeria enrichment and isolation. All samples were processed as previously described (*49*). Briefly, 10 g of the fecal samples were aliquoted and transferred to separate Whirl-Pak bags (Nasco, Fort Atkinson, WI). The water samples were each passed through separate 0.45 µm filters, and each filter was transferred to a separate Whirl-Pak bag. Each sample was enriched in 90 mL of Buffered *Listeria* Enrichment Broth (BLEB; Becton Dickinson, Franklin Lakes, NJ) at 30°C, and after 4h, *Listeria*

Selective Enrichment Supplement (Oxoid, Cambridge, UK) was added to each sample. Each sample was incubated for a total of 48h at 30°C. After 24h and 48h of incubation, the enrichments were sub-streaked onto Modified Oxford Agar (MOX; Benton Dickinson) and *Listeria monocytogenes* Plating Medium (LMPM; Biosynth International, Itasca, IL). MOX plates were incubated at 30°C and LMPM plates were incubated at 35°C for 48h. After incubation, characteristic *Listeria* colonies on 24 and 48h MOX plates (i.e., grey, donut-shaped colonies) were streaked onto LMPM for isolation and incubated at 35°C for 48h. One colony per presumptive *Listeria*-positive LMPM plate (up to 4 colonies total; 1 from the 24h LMPM plate, 1 from the 48h LMPM plate, 1 from the LMPM plate sub-streaked from the 24h MOX plate, and 1 from the LMPM plate sub-streaked from the 48h MOX plate) was selected, sub-streaked onto brain heart infusion agar (BHI, Becton Dickinson), and incubated at 37°C for 24h. If blue colonies were present (characteristic of *L. monocytogenes*, *L. ivanovii*, hemolytic *L. innocua* due to phospholipase C action, *29*) on a given plate, they were preferentially selected to be sub-streaked on BHI. If a blue colony was not present, a round white colony was selected to be sub-streaked on BHI; this colony phenotype is characteristic of *Listeria* species other than *L. monocytogenes*, *L. ivanovii*, and hemolytic *L. innocua*. If no round white or blue colonies were present on any of the LMPM plates, the sample was considered negative for *Listeria*. If a round white or blue colony was present, *sigB* PCR was performed as a confirmation step and speciation of *Listeria*, as described below.

sigB sequencing and allelic typing. All isolates per presumptive-positive *Listeria* sample (i.e., up to 4 isolates per sample) were confirmed as *Listeria* via PCR amplification of *sigB* as described by Nightingale et al. (*22*); all samples with presumptive-positive *Listeria* isolates in this study were confirmed as positive via *sigB* PCR. Sequencing of the *sigB* gene was then performed on 1 isolate per positive sample

for further characterization; if the colony characterized was identified as *L. monocytogenes* the sample was considered *L. monocytogenes* positive, if the colony characterized was identified as *Listeria* spp. other than *L. monocytogenes*, then sample was considered *Listeria* spp. positive. While only subtyping 1 isolate per sample may lead to an underestimation of the true diversity present in the sample types investigated here, this approach allowed us to get a preliminary estimate of the diversity present. There were 22 samples with both blue and white colonies present on LMPM plates (out of 209 total samples); for 21 of these samples the blue colony was selected for characterization (2/3 fecal samples, 8/8 stream samples, and 11/11 pond samples). Since only *L. monocytogenes* was selected for PFGE, preferentially selecting these blue colonies provided us with the larger sample size for further subtyping. A BLAST search was performed against an internal database (http://www.foodmicrobetracker.com/login/intro.aspx) to determine its *sigB* allelic type (AT) of each sequence. All isolates were archived as 15% glycerol stocks and stored at -80°C. A *sigB* AT phylogeny was constructed based on the 660 bp nucleotide sequence of the *sigB* gene using the maximum likelihood method in MEGA7 version 7.0.26; the nucleotide substitution Tamura 3-parameter model was used, and 1,000 bootstraps were performed.

Pulsed field gel electrophoresis. All confirmed *L. monocytogenes* isolates from the pond and the feces collected near the pond (N=44) were characterized by pulsed field gel electrophoresis (PFGE). Only isolates from pond water and feces were characterized by PFGE as a subset analysis to investigate diversity over a small spatial scale (i.e., the approx. 5,000m^2 around the pond). PFGE was performed using *AscI* and *ApaI* restriction enzymes, according to the CDC PulseNet protocol (*10*).

Statistical analysis. All statistical analyses were performed using R version 3.5.1 (*27*). Chi-square tests were used to determine if *L. monocytogenes* and *Listeria*

spp. (excluding *L. monocytogenes*) prevalence differed between sample types. The Bonferroni correction was used to account for the 3 comparisons made for *Listeria* spp. (excluding *L. monocytogenes*) and the 3 comparisons made for *L. monocytogenes* (pond vs. fecal, pond vs. stream, fecal vs. stream; $\alpha=0.05/3=0.0167$). While separate data for the feces collected near the pond and feces collected near the stream are provided in this study, these sample types were combined into a single fecal sample category for analysis due to the small number of fecal samples collected near the stream (N=13 of 77 total fecal samples). To determine the number of unique ATs and PFGE types, AT richness and PFGE type richness were estimated using the breakaway package version 4.6.7 (*43*), which calculates the transformed species richness estimation with 95% confidence intervals as outlined in Rocchetti et al. (*28*) and Willis and Bunge (*42*). To estimate AT and PFGE type diversity for each sample type, Simpsons Index of Diversity was calculated using the vegan package version 2.5-3 (*23*); 95% confidence intervals (95% CI) were calculated for the Simpson's Index of Diversity, according to the methods described by Hunter and Gaston (*13*). The effective number of subtypes was also calculated to communicate Simpson's Index of Diversity in a more intuitive form (i.e., the estimated number of subtypes present in a sample type given a certain value of Simpson's Index of Diversity; *15*).

RESULTS AND DISCUSSION

Seek and you shall find. If a reasonably large number of samples are collected from the pre-harvest environment in New York State and tested for *Listeria* ("seek"), it is likely *Listeria* will be "found." The overall prevalence of *Listeria* in the study reported here was 65% (136/209); the prevalence of *Listeria* spp. (excluding *L. monocytogenes*) was 24% (50/209) and the prevalence of *L. monocytogenes* was 41% (86/209; Table 3.1). While the *Listeria* spp. (excluding *L. monocytogenes*) prevalence in the pond

(31%; 25/80) and stream samples (27%; 14/52) were not significantly different (P=0.735), the *L. monocytogenes* prevalence in the pond (39%; 31/80) and stream samples (63%; 33/52) were significantly different even after the Bonferroni correction (P=0.009; Table 3.1). The frequency of *Listeria* isolation from agricultural water has been reported in several previous studies conducted in New York State (*3, 34, 35, 48, 40*) and other produce-growing regions (*1, 6, 7, 18, 33*). In previous New York State studies, the *L. monocytogenes* prevalence ranged between 9% (*48*) and 59% (*35*) in pond samples, and between 10% (*40*) and 50% (*48*) in stream samples. As such, the frequency of *L. monocytogenes* isolation from pond samples in the current study falls within the range reported in the literature, while the frequency of *L. monocytogenes* isolation from the stream samples in the current study is higher than in previous studies. The higher *L. monocytogenes* prevalence in the stream sampled here, compared to those sampled in other studies, may be due to presence of two dairy operations immediately upstream of the sampling sites along the stream. Additionally, run-off from the dairy operations could explain the significantly higher *L. monocytogenes* prevalence in the stream compared to the pond. Previous studies have indicated upstream livestock operations are associated with an increased prevalence of *L. monocytogenes* in stream water (*7, 19, 41*). Furthermore, another study suggested cattle may increase the prevalence and may distribute *L. monocytogenes* in the farm environment (*22*). Of note, the current study only tested 250 mL of water per sample, whereas other studies collected upwards of 1 L per sample. As such, the prevalence reported for the pond and stream investigated in the current study may be an underestimation of the true prevalence. Regardless, while the current study supports agricultural water as a source of *L. monocytogenes* in the New York State preharvest produce environment, our findings also suggest future studies on agricultural water need to consider the impact of spatiotemporal factors (e.g., upstream land use) on downstream microbial water quality.

Table 3.1. Frequency and prevalence of *Listeria spp.* and *L. monocytogenes* in fecal and agricultural water samples.

Sample Type[b]	Location	No. of Samples	No. of samples positive for (Prevalence in %)				
			Listeria spp.[a]	*L. innocua*	*L. seeligeri*	*L. welshimeri*	*L. monocytogenes*
Fecal	-	77	11 (14)	6 (8)	4 (5)	1 (1)	22 (29)
	Near Pond	64	9 (14)	4 (6)	4 (6)	1 (2)	13 (20)
	Near Stream	13	2 (15)	2 (15)	0 (0)	0 (0)	9 (69)
Water	-	132	39 (30)	14 (11)	20 (15)	5 (4)	64 (48)
	Pond	80	25 (31)	8 (10)	15 (19)	2 (3)	31 (39)
	Stream	52	14 (27)	6 (12)	5 (10)	3 (6)	33 (63)

[a]*Listeria* spp. includes *L. innocua*, *L. seeligeri*, and *L. welshimeri*.

[b]In addition to the water samples collected from the pond water, one sediment sample was also collected and tested positive for *L. innocua*.

The prevalence of *Listeria* spp. (excluding *L. monocytogenes*) and *L. monocytogenes* in the fecal samples collected here were 14% (11/77) and 29% (22/77), respectively (Table 3.1). Based on fecal shape and consistency, the majority of the fecal samples appeared to be produced by Canadian geese (*Branta canadenesis*; N=71). This is logical given the large flock of Canadian geese that nested in the wetlands adjacent to the pond throughout the study. The remainder of the feces were produced by canids (N=5) and an unknown animal (N=1). The *Listeria* spp. (excluding *L. monocytogenes*) prevalence was 13% (9/71) in the goose and 40% (2/5) in the canid fecal samples. The *L. monocytogenes* prevalence was 28% (20/71) in the goose and 40% (2/5) in the canid fecal samples. Neither *Listeria* spp. (excluding *L. monocytogenes*), nor *L. monocytogenes,* were isolated from the one unidentified fecal sample. While multiple studies have isolated *Listeria* from wildlife feces (*5, 19, 25, 34, 37, 48, 49*), relatively few studies have examined the prevalence of *L. monocytogenes* in Canadian goose feces, specifically (*5, 19*). The *L. monocytogenes* prevalence in the goose fecal samples collected as part of the current study was greater compared to previous studies (*5, 19*). For instance, a study conducted in Ontario, Canada, did not detect *L. monocytogenes* in any of the 18 goose fecal samples collected (*19*), while Converse et al. (*5*) isolated *Listeria* spp. from 10% (47/495) of goose fecal samples. Although, the prevalence of *L. monocytogenes* in wildlife fecal samples varied between studies, our findings support previous studies' findings (e.g., *33, 36, 37*) that wildlife can act as source of *L. monocytogenes* in pre-harvest environments.

Absence of evidence does not equal evidence of absence. The composition of the *Listeria* population in the agricultural water and fecal samples was characterized using *sigB* allelic typing. Allelic typing was used to speciate the isolates and provide a preliminary estimate of diversity. A total of 136 isolates were selected for *sigB* allelic typing (i.e., one isolate per *Listeria*-positive sample); *L. monocytogenes* isolates were

preferentially selected for subtyping based on colony morphology (blue colonies on LMPM). Of the 136 isolates, 20, 24, 6, and 86 isolates were identified as *L. innocua*, *L. seeligeri*, *L. welshimeri*, and *L. monocytogenes*, respectively. We identified 15 *sigB* ATs among the 50 *Listeria* spp. (excluding *L. monocytogenes*) isolates (approx. 1 unique AT per 3 isolates) and 8 ATs among the 86 *L. monocytogenes* isolates (approx. 1 unique AT per 11 isolates; Tables 3.2 & 3.3; Table S1). In addition, we estimated two diversity parameters, richness and the Simpson's Index for *Listeria* spp. (excluding *L. monocytogenes*; Table 3.2) and *L. monocytogenes* (Table 3.3) in the pond, stream, and fecal samples. For example, the estimated *L. monocytogenes* AT richness for the pond water, stream water and fecal samples were 8 (95% CI= 7, 15), 9 (95% CI= 6, 73) and 5 (95% CI= 5, 8) ATs, respectively, indicating the presence of a considerable diversity on the farm investigated. However, it should be noted that this may be an underestimation of the true diversity present in these sample types on this produce farm, as only one isolate per sample type was characterized by allelic typing. Even with this limitation our data indicate considerable *Listeria* diversity in the small area of the produce farm investigated here. The proportion of ATs identified relative to the total number of isolates subtyped in the current study was similar to the proportion reported by previous studies conducted on New York State produce farms (*3*, *34*). For example, applying the same analysis as in the current study, Chapin et al. identified 50 *sigB* ATs among 186 *Listeria* spp. (excluding *L. monocytogenes*) isolates [(approx. 1 AT per 4 isolates), (*3*)] and Strawn et al. identified 12 *sigB* ATs among 107 *L. monocytogenes* isolates [(approx. 1 AT per 8 isolates), (*34*)]; both studies isolated *Listeria* from soil, drag swabs, water, and fecal samples collected from the five produce farms across New York State. Since all samples collected in the study reported here were from a <0.5 km^2 area, compared to past studies which sampled larger areas (between 33 and 205 km apart), the similarity in the ratio of unique ATs relative to the total number of isolates

subtyped in this and these previous studies (*3*, *34*) provides *Listeria* richness may be similar across spatial scales. However, additional research is required to confirm this finding, especially in other geographical areas.

Table 3.2. Diversity of *Listeria* spp. *sigB* allelic types (AT) in agricultural water and fecal samples.

Sample Type		No. of *Listeria* spp.[a]	No. of ATs	Estimated Subtype Richness (95% CI)[b,c]	Simpson's Index (95% CI)[c]	Effective No. of AT[d]
Fecal	-	11	8	35 (8, 2871)	0.84 (0.75, 0.94)	6
Water	-	39	14	23 (14, 259)	0.84 (0.76, 0.92)	6
	Pond	25	11	32 (11, 1445)	0.79 (0.67, 0.92)	5
	Stream	14	8	11 (8, 39)	0.84 (0.76, 0.91)	6

[a]Excluding *L. monocytogenes*

[b]Estimated using weighted linear regression

[c]95% CI: 95% confidence interval

[d]Effective number of AT is calculated based on Simpson's Index of Diversity

Table 3.3. Diversity of *Listeria monocytogenes sigB* allelic types and pulsotypes in agricultural water and fecal samples.

Sample Type	Location	No. of *L. monocytogenes*	Subtyping Method[a]	No. of Subtypes	Estimated Subtype Richness (95% CI)[b,c]	Simpson's Index (95% CI)[c]	Effective No. of Subtype[d]
Fecal	-	22	AT	5	5 (5, 8)	0.63 (0.46, 0.80)	3
	Near Pond	13	PT	7	9 (7, 33)	0.78 (0.63, 0.93)	5
Water	-	64	AT	7	7 (7, 8)	0.54 (0.41, 0.68)	2
	Stream	33	AT	6	9 (6, 73)	0.41 (0.20, 0.62)	2
	Pond	31	AT	7	8 (7, 15)	0.66 (0.49, 0.82)	3
	Pond	31	PT	19	29 (19, 207)	0.93 (0.91, 0.95)	15

[a] AT = allelic type; PT = PFGE type. PFGE was only performed on *Listeria monocytogenes* isolates and not on non-pathogenic *Listeria* species isolates.

[b] Estimated using weighted linear regression

[c] 95% CI: 95% confidence interval

[d] The effective number of subtype is calculated based on Simpson's Index of Diversity.

Furthermore, multiple *Listeria* (including *L. monocytogenes*) *sigB* ATs were isolated from a single sample type on a given day on multiple occasions. In the pond, stream, and fecal samples there were an average of 1.9, 1.5, and 1.7 *Listeria* (including *L. monocytogenes*) *sigB* ATs isolated per sampling day, respectively. There was also a maximum of 3, 2, and 4 *Listeria* (including *L. monocytogenes*) *sigB* ATs isolated per sampling day from the pond, stream, and fecal samples, respectively, indicating considerable *Listeria* diversity in the area sampled even on a given day. Six of the 8 *L. monocytogenes sigB* ATs were isolated from at least one sample type on multiple dates (see Figure 3.1, which shows the isolation of *L. monocytogenes sigB* ATs over time; equivalent data for *Listeria* spp. [excluding L. *monocytogenes*] are shown in Figure S1). Most notably, AT 57 was isolated from all three sample types (i.e., pond water, stream water, and feces) on multiple occasions, and was isolated from at least one sample type on 23 of the 26 sampling dates. AT 57 represents 13 PFGE types (Figure 3.2), consistent with previous reports that PFGE is considerably more discriminatory then *sigB* allelic typing (*4*). The high prevalence of a single AT (i.e., AT 57) makes it likely that less prevalent ATs are not isolated, as for example, AT 57 may outcompete other ATs. In addition, the considerable AT diversity observed also make it likely that samplings that do not include large sample numbers will not detect and isolate some *Listeria sigB* ATs. Hence, traceback investigations could mistakenly conclude that a given *Listeria* subtype is not present in a given pre-harvest environment, particularly in environments similar to the one investigated here.

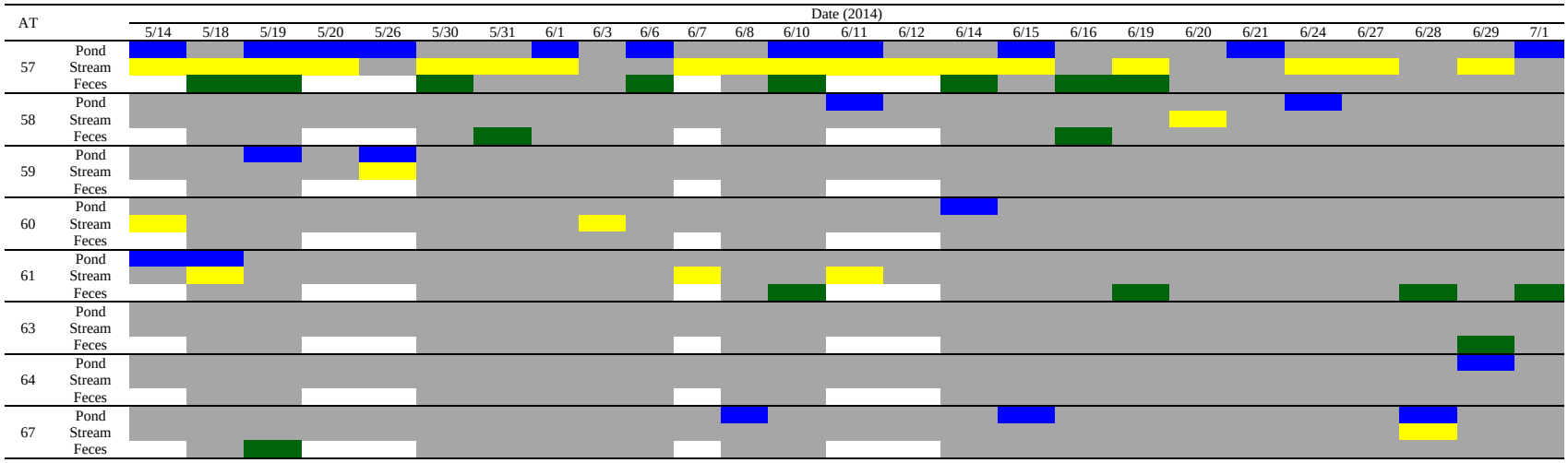
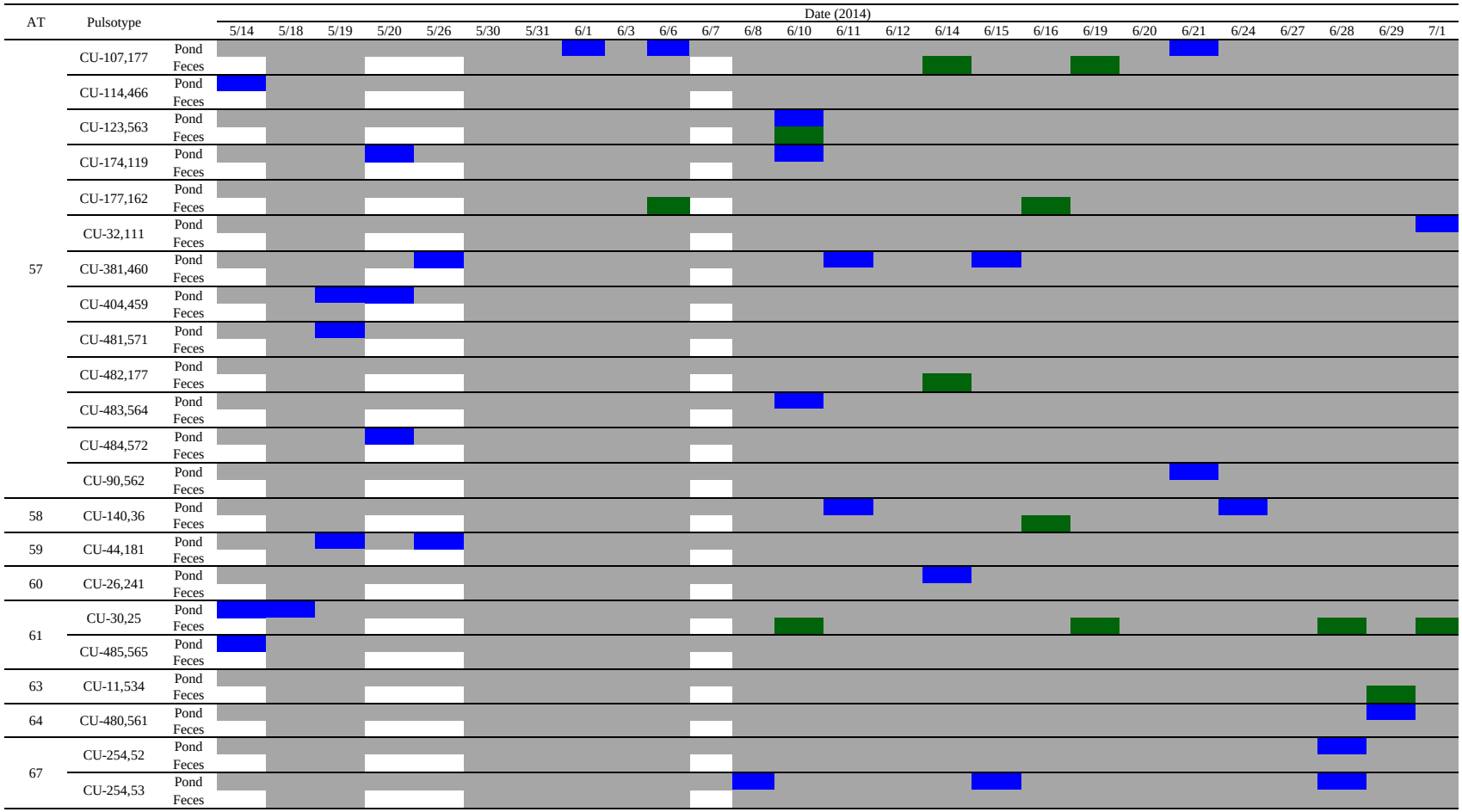


Figure 3.1. Date of *L. monocytogenes sigB* AT isolation from water and fecal samples. In total, 31, 33, and 22 isolates from pond, stream, and fecal samples were allelic typed, respectively. The colored squares indicate at least one isolate of the given AT was isolated on the given date from pond ( ), stream ( ), or fecal ( ) samples; dates when a sample was collected, but the specified AT was not isolated are represented by , and dates when a sample was collected but was negative for *Listeria* are represented by . At each sampling event, 3 pond and 3 stream samples were collected. Fecal samples were collected opportunistically (i.e., when feces were present). If no fecal samples were collected, the square is white. The dates of *Listeria* spp. (excluding *L. monocytogenes*) *sigB* AT isolation is shown in Figure S1.

Figure 3.2. Date of *L. monocytogenes* PFGE type isolation from pond and fecal samples collected near the pond. In total, PFGE was performed on 31 and 13 isolates from pond and fecal samples, respectively. The colored squares indicate at least one isolate of the given PFGE type was isolated on the given date from pond ( ) or fecal ( ) samples; dates when a sample was collected, but the specified PFGE type was not isolated are represented by , and dates when a sample was collected but was negative for *L. monocytogenes* are represented by . At each sampling event, 3 pond samples were collected. Fecal samples were collected opportunistically (i.e., when feces were present). If no fecal samples were collected, the square is white.


AT
Pulsotype
Date (2014)
5/14
5/18
5/19
5/20
5/26
5/30
5/31
6/1
6/3
6/6
6/7
6/8
6/10
6/11
6/12
6/14
6/15
6/16
6/19
6/20
6/21
6/24
6/27
6/28
6/29
7/1
57
CU-107,177
CU-114,466
CU-123,563
CU-174,119
CU-177,162
CU-32,111
CU-381,460
CU-404,459
CU-481,571
CU-482,177
CU-483,564
CU-484,572
CU-90,562
58
CU-140,36
59
CU-44,181
60
CU-26,241
61
CU-30,25
CU-485,565
63
CU-11,534
64
CU-480,561
67
CU-254,52
CU-254,53
Pond
Feces

Additionally, a *sigB* AT phylogeny was constructed to investigate if there was any relationship between *sigB* AT isolation and sample type (Figure 3.3). The phylogeny suggests that neither specific *sigB* ATs nor specific *sigB* clades are associated with a given sample type. A Fisher's exact test also found no significant relationship between *sigB* AT and sample type (P=0.30). While our data could also be interpreted as suggesting that none of the ATs isolated in this study have an enhanced fitness in the different habitats sampled here, future studies with larger data sets are needed to further probe niche adaptation of different *Listeria* species and clades.

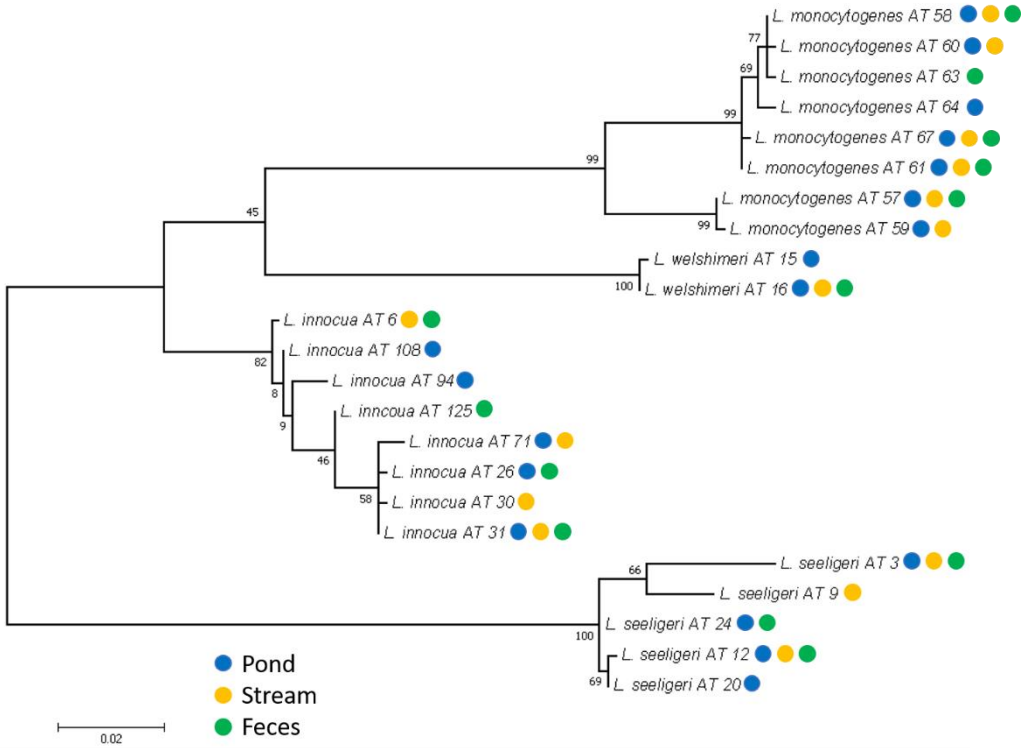


Figure 3.3. Maximum likelihood phylogeny of *Listeria* isolated from pond, stream, and fecal samples from a single produce farm in New York State based on a 660 bp fragment of the *sigB* gene. Values on the branches are bootstrap values from 1,000 replicate trees. The scale displayed below the tree represents the genetic distance shown in the branch length.

In addition to characterization by *sigB* AT, we performed PFGE on the *L. monocytogenes* isolates from the pond (N=31) and feces collected near the pond (N=13) to better understand the diversity of the isolates at a very small spatial scale (approx. 5,000 m^2). We identified 19 PFGE types among the 31 *L. monocytogenes* isolates obtained from pond water samples (approx. 1 PFGE type per 2 isolates) and 7 PFGE types among the 13 *L. monocytogenes* isolates obtained from fecal samples (approx. 1 PFGE type per 2 isolates). The estimated PFGE type richness for *L. monocytogenes* isolated from the pond water and fecal samples were 29 (95% CI= 19, 207) and 9 (95% CI= 7, 33) PFGE types, respectively (Table 3.3). Two concurrent studies conducted in a 200 km^2 agricultural watershed in Ontario, Canada also performed PFGE on *L. monocytogenes* isolates from agricultural water and wildlife feces (*19, 20*); these studies used the same PFGE protocol as the current study (*10*). In the two Canadian studies, Lyautey et al., (*19, 20*) identified 21 PFGE types from 75 *L. monocytogenes* isolates from water samples (approx. 1 PFGE type per 3 isolates) and 18 PFGE types from 84 *L. monocytogenes* isolates from fecal samples (approx. 1 PFGE type per 5 isolates). As such, the ratio of PFGE types identified relative to the number of *L. monocytogenes* isolates subtyped in the current study, which was conducted in a 40,000-fold smaller area, was larger than the ratio found in the Canadian studies (*19, 20*) for both the fecal samples and water samples. While this may suggest the *L. monocytogenes* population in the Finger Lakes region of New York is more diverse than Ontario, Canada, the greater diversity on the farm sampled in the study reported here may also be due to factors unique to the sampled farm. For example, since two dairy farms are located uphill of the sampled farm, continuous introduction of *L. monocytogenes* from the dairies may have resulted in a more diverse population of *L. monocytogenes*, compared to the other areas. To identify potential drivers of the greater PFGE type diversity observed in the study reported here, compared to Lyautey et al. (*19, 20*), a systematic

survey of *Listeria* populations across and within regions is needed. Regardless, the diversity of *Listeria* identified in this study, and other studies, suggests it often may be difficult to determine if an isolate found in a processing environment originated in a particular pre-harvest environment, as multiple subtypes may be present in a single environment. In fact, in order to either rule in or out any particular source in an environment with a considerable diversity, such as the one investigated in the current study, a large number of samples would have to be collected and a large number of isolates would have to be subtyped, including multiple isolates per sample, as previously suggested (*9, 30*). Consequently, the "absence of evidence" for presence of a given *Listeria* subtype (i.e., the *Listeria* subtype was not detected in any of the samples collected) in the pre-harvest environment in New York State and other pre-harvest environments with prevalent *Listeria* does not necessarily provide evidence that the given subtype is truly absent in this environment, particularly if sample size is small.

Should I stay or should I go. The *L. monocytogenes* PFGE data collected here indicate the survival ("stay") or continuous re-introduction ("go") of *L. monocytogenes* into the farm environment. Specifically, 8 of the 19 PFGE types isolated from the pond and 3 of the 7 PFGE types isolated from the feces collected near the pond were isolated on more than one date (Figure 3.2). The repeated isolation of multiple PFGE types suggests these PFGE types were surviving or were continuously re-introduced in these sample types. For example, PFGE type CU-30,25 was isolated a total of 8 times in the study reported here, making it the most commonly isolated PFGE type (Figure 3.2). This could indicate PFGE type CU-30,25 has an increased fitness in this environment or there is a source of this PFGE type continuously contaminating the environment. Additionally, 4 out of the 22 PFGE types identified here were isolated from both the pond and feces collected near the pond (Figure 3.2). Lyautey et al. (*19, 20*) found similar results in two of their studies, as 5 of the 18 PFGE types from wildlife feces in the South

Nation watershed in Ontario, Canada, were also isolated in water samples collected from the same watershed. Since the same PFGE types can be found in agricultural water and feces collected in proximity of the sampled water in both studies, this supports transfer of *L. monocytogenes* between feces and agricultural water, and/or a common source leading to the contamination of both sample types in the investigated environment.

Work smarter not harder. Allelic typing of the *sigB* gene has been used to speciate *Listeria* isolates and can provide initial characterization of *Listeria* strains present (*16*), but is not of high enough discriminatory power to elucidate the actual diversity of a *Listeria* population in an environment. For example, in the current study AT 57 represents 13 different PFGE types, AT 61 represents 2 different PFGE types and AT 67 represents 2 different PFGE types (Figure 4.2). Strawn et al. (*34*) also performed PFGE on *L. monocytogenes* isolates with matching *sigB* ATs from a single sample type in a survey of soil, feces, drag swabs, and water samples from 5 farms in New York State (i.e., repeatedly isolated *sigB* ATs). For 3 out of the 4 repeatedly isolated *sigB* ATs, they found multiple PFGE types corresponding to each individual AT. PFGE hence clearly provides improved discriminatory power (over *sigB* allelic typing), facilitating investigations of persistence, transfer between sample types, and continuous re-introduction (*4*). Interestingly, in the study reported here, all isolates of the same PFGE type also had the same AT (Figure 3.2). For instance, all isolates of PFGE type CU-371,460 were AT 57, all isolates of PFGE type CU-107,177 were AT 57, all isolates of PFGE type CU-254,53 were AT 67, etc. This finding has not been directly reported in other studies, however, Liao et al. (*16*) have determined the *sigB* gene is genetically stable, as it is not a frequent site of homologous recombination or positive selection. As such, we expect this relationship between *sigB* ATs and PFGE types to hold true across different environments. While *sigB* allelic typing is not as discriminatory as PFGE, it can still be useful in performing traceback analysis, as it can

be used as an initial screen to determine which isolates should be selected for PFGE, or other fine resolution typing methods (e.g., whole genome sequencing, WGS) during a traceback investigation. As a result, to determine which isolates from an environment match an isolate from another environment (e.g., processing environment), allelic typing should first be performed ("*work smarter*"). Then, isolates from the pre-harvest environment that have the same allelic type as the isolate from the processing environment should be characterized by PFGE or WGS ("*work harder*") to determine which are truly the same. This will save time and money, as PFGE and WGS are considerably more expensive and time intensive to perform.

Overall, the data presented here and in previous studies (*49*) indicate *Listeria* is prevalent in the New York State pre-harvest produce environment, and as a result, it is likely for *Listeria* from the pre-harvest environment to contaminate produce, and/or be introduced to the processing environment (from the farm). The considerable *Listeria* diversity observed in the small area within the produce farm studied, and the evidence of transfer between sample types or continuous re-introduction of *Listeria*, indicate that it often may be difficult to successfully trace isolates from finished products or the processing environment (e.g., environmental swab) back to farm sources (e.g., feces, water), unless a large number of samples are collected and a large number of isolates are subtyped. However, it should be noted that these conclusions may be limited to similar preharvest produce environments that are continuously contaminated with livestock run-off. As such, additional research is needed to determine if the findings reported as part of this pilot study hold true for other produce farms and produce growing regions.

ACKNOWLEDGMENTS

This work is supported by United States Department of Agriculture (USDA) Specialty Crop Research Initiative grant number 2019-51181-30016 from the National Institute of Food and Agriculture. Manuscript preparation and revision were also partially supported by the National Institute of Environmental Health Sciences, National Institutes of Health (NIH) under award number T32ES007271 and the Virginia Agricultural Experiment Station and the Hatch Program of the National Institute of Food and Agriculture, USDA. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the USDA or NIH.

SUPPLEMENTAL MATERIAL

Supplemental material can be found at: https://doi.org/10.4315/JFP-20-179.s1

REFERENCES

1. Bradshaw, J. K., B. J. Snyder, A. Oladeinde, D. Spidle, M. E. Berrang, R. J., Meinersmann, B. Oakley, R. C. Sidle, K. Sullivan, and M. Molina. 2016. Characterizing relationships among fecal indicator bacteria, microbial source tracking markers, and associated waterborne pathogen occurrence in water and sediments in a mixed land use watershed. *Water Res.* 101:498-509.

2. Centers for Disease Control and Prevention. 27 August 2012. Multistate outbreak of listeriosis linked to whole cantaloupes from Jensen Farms, Colorado (final update). Available at: https://www.cdc.gov/listeria/outbreaks/cantaloupes-jensen-farms/index.html

3. Chapin, T. K., K. K. Nightingale, R. W. Worobo, M. Wiedmann, and L. K. Strawn. 2014. Geographical and meteorological factors associated with isolation of *Listeria* species in New York State produce production and natural environments. *J. Food Prot.* 77:1919-1928. doi:10.4315/0362-028X.JFP-14-132

4. Chen, J. Q., P. Regan, P. Laksanalamai, S. Healey, and Z. Hu. 2017. Prevalence and methodologies for detection characterization and subtyping for *Listeria monocytogenes* and *L. ivanovii* in foods and environmental sources. *Food Sci Hum Wellness*. 6(3):97-120.

5. Converse, K., M. Wolcott, D. Docherty, and R. Cole. 1999. Screening for potential human pathogens in fecal material deposited by resident Canada Geese on areas of public utility. *USGS Natl. Wildl. Health Cent. Rep.*

6. Cooley, M. B., B. Quinones, D. Oryang, R. E. Mandrell, and L. Gorski. 2014. Prevalence of shiga toxin producing *Escherichia coli, Salmonella enterica*, and *Listeria monocytogenes* at public access watershed sites in a California Central Coast agricultural region. *Front. Cell. Infect. Microbiol*. 4(article 30). doi:10.3389/fcimb.2014.00030

7. Falardeau, J., R. P. Johnson, F. Pagotto, and S. Wang. 2017. Occurrence, characterization, and potential predictors of verotoxigenic *Escherichia coli*, *Listeria monocytogenes*, and *Salmonella* in surface water used for produce irrigation in the Lower Mainland of British Columbia, Canada. *PLoS One*. 12(9):e0185437.

8. Garner, D. and S. Kathariou. 2016. Fresh produce–associated listeriosis outbreaks, sources of concern, teachable moments, and insights. *J. Food Prot*. 79:337-344. doi:10.4315/0362-028X.JFP-15-387.

9. Gorski L., S. Walker, A. S. Liang, K. M. Nguyen, J. Govoni, D. Carychao, M. B. Cooley, and R. E. Mandrell. 2014. Comparison of Subtypes of *Listeria monocytogenes* isolates from naturally contaminated watershed samples with and without a selective secondary enrichment. *PLOS ONE*. 9(3): e92467. https://doi.org/10.1371/journal.pone.0092467

10. Graves, L. M. and B. Swaminathan. 2001. PulseNet standardized protocol for subtyping *Listeria monocytogenes* by macrorestriction and pulsed-field gel electrophoresis. *Int. J. Food Microbiol*. 65:55-62.

11. Haase, J. K., X. Didelot, M. Lecuit, H. Korkeala, *L. monocytogenes* MLST Study Group, and M. Achtman. 2014. The ubiquitous nature of *Listeria monocytogenes* clones: A large-scale multilocus sequence typing study. *Environ. Microbiol*. 16(2):405-416. doi:10.1111/1462-2920.12342

12. Heaton, J. C. and K. Jones. 2008. Microbial contamination of fruit and vegetables and the behaviour of enteropathogens in the phyllosphere: A review. *J. Appl. Microbiol*. 104:613-626. doi:10.1111/j.1365-2672.2007.03587.x

13. Hunter, P. R. and M. A. Gaston. 1988. Numerical index of the discriminatory ability of typing systems: An application of Simpson's Index of Diversity. *J. Clin. Microbiol*. 26:2465-2466.

14. Jay-Russell, M. T. 2013. What is the risk from wild animals in food-borne pathogen contamination of plants? *CAB Rev*. 8:1-16. doi:10.1079/PAVSNNR20138040.

15. Jost, L. 2006. Entropy and diversity. *OIKOS*. 113:363-375.

16. Liao, J., M. Wiedmann, and J. Kovac. 2017. Genetic stability and evolution of the sigB allele, used for *Listeria sensu stricto* subtyping and phylogenetic inference. *Appl. Environ. Microbiol*. 83(12). https://doi.org/10 .1128/AEM.00306-17.

17. Linke, K., I. Ruckerl, K. Brugger, R. Karpiskova, J. Walland, S. Muri-Klinger, A. Tichy, M. Wagner, and B. Stessl. 2014. Reservoirs of *Listeria* species in three environmental ecosystems. *Appl. Environ. Microbiol*. 80:5583-5592. doi:10.1128/AEM.01018-14

18. Lopez-Galvez, F., A. Allende, F. Pedrero-Salcedo, J. J. Alarcon, and M. I. Gil. 2014. Safety assessment of greenhouse hydroponic tomatoes irrigated with reclaimed and surface water. *Int. J. Food Microbiol*. 191:97-102. http://dx.doi.org/10.1016/j.ijfoodmicro.2014.09.004

19. Lyautey, E., A. Hartmann, F. Pagotto, K. Tyler, D. R. Lapen, G. Wilkes, P. Piveteau, A. Rieu, W. J. Robertson, D. T. Medeiros, T. A. Edge, V. Gannon, and E. Topp. 2007. Characteristics and frequency of detection of fecal *Listeria monocytogenes* shed by livestock, wildlife, and humans. *Can. J. Microbiol*. 53: 1158-1167. doi:10.1139/W07-084.

20. Lyautey, E., D. R. Lapen, G. Wilkes, K. McCleary, F. Pagotto, K. Tyler, A. Hartmann, P. Piveteau, A. Rieu, W. J. Robertsonm D. T. Medeiros, T. A. Edge, V. Gannon, and E. Topp. 2007. Distribution and characteristics of *Listeria monocytogenes* isolates from surface waters of the South Nation River Watershed, Ontario, Canada. *Appl. Environ. Microbiol*. 73:5401-5410. doi:10.1128/AEM.00354-07.

21. Nightingale, K. K., Y. H. Schukken, C. R. Nightingale, E. D. Fortes, A. J. Ho, Z. Her, Y. H. Grohn, P. L. McDonough, and M. Wiedmann. 2004. Ecology and transmission of *Listeria monocytogenes* infecting ruminants and in the farm environment. *Appl. Environ. Microbiol*. 70(8):4458-4467. DOI: 10.1128/AEM.70.8.4458–4467.2004

22. Nightingale, K. K., K. Windham, and M. Wiedmann. 2005. Evolution and molecular phylogeny of *Listeria monocytogenes* isolated from human and animal listeriosis cases and foods. *J. Bacteriol*. 187:5537-5551. doi:10.1128/JB.187.16.5537–5551.2005

23. Okasanen, J., F. Guillaume Blanchet, M. Friendle, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'hara, G. L. Simpson, P. Solymos, M. Henry, H. Stevens, E. Szoecs, and H. Wagner. 2018. Vegan: Community ecology package. R package version 2.5-3. Available at: https://CRAN.R-project.org/package=vegan. Accessed 4 May 2018.

24. Painter, J. A., R. M. Hoekstra, T. Ayers, R. V. Tauxe, C. R. Braden, F. J. Angulo, and P. M. Griffin. 2013. Attribution of foodborne illnesses, hospitalizations, and deaths to food commodities by using outbreak data, United States, 1998–2008. *Emerg. Infect. Dis.* 19:407-415. doi:10.3201/eid1903.111866.

25. Pang, H., R. McEgan, A. Mishra, S. A. Micallef, and A. K. Pradhan. 2017. Identifying and modeling meteorological risk factors associated with preharvest contamination of *Listeria* species in a mixed produce and dairy farm. *Food Res. Int*. 102:355-363. doi:10.1016/j.foodres.2017.09.029.

26. Park, S., B. Szonyi, R. Gautam, K. Nightingale, J. Anciso, and R. Ivanek. 2012. Risk factors for microbial contamination in fruits and vegetables at the preharvest level: A systematic review. *J. Food Prot*. 75:2055-2081. doi:10.4315/0362-028X.JFP-12-160.

27. R Core Team. 2020. R: A language and environment for statistical computing. Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.

28. Rocchetti, I., J. Bunge, and D. Bohning. 2011. Population size estimation based upon ratios of recapture probabilities. *Ann. Appl. Stat*. 5(2B):1512-1533. doi:10.1214/10-AOAS436

29. Restaino, L., E. W. Frampton, R. M. Irbe, G. Schabert, and H. Spitz. 1999. Isolation and detection of Listeria monocytogenes using fluorogenic and chromogenic substrates for phosphatidylinositol-specific phospholipase C. J Food Prot. 62(3):244-251.

30. Ryser, E. T., S. M. Arimi, M. M. C. Bunduki, and C. W. Connelly. 1996. Recovery of different *Listeria* ribotypes from naturally contaminated, raw refrigerated meat and poultry products with two primary enrichment media. *Appl. Environ. Microbiol*. 62(5):1781-1787.

31. Scallan, E., R. M. Hoekstra, F. J. Angulo, R. V. Tauxe, M. A. Widdowson, S. L. Roy, J. L. Jones, and P. M. Griffin. 2011. Foodborne illness acquired in the United States- major pathogens. *Emerg. Infect. Dis.* 17:7-15.

32. Smith, A., E. Moorhouse, J. Monaghan, C. Taylor, and I. Singleton. 2018. Sources and survival of *Listeria monocytogenes* on fresh, leafy produce. *J. Appl. Microbiol.* 125:930-942. doi:10.1111/jam.14025

33. Stea, E. C., L. M. Purdue, R. C. Jamieson, C. K. Yost, and L. T. Hansen. 2015. Comparison of the prevalences and diversities of *Listeria* species and *Listeria monocytogenes* in an urban and a rural agricultural watershed. *Appl. Environ. Microbiol*. 81:3812-3822. doi:10.1128/AEM.00416-15

34. Strawn, L. K., E. D. Fortes, E. A. Bihn, K. K. Nightingale, Y. T. Grohn, R. W. Worobo, M. Wiedmann, and P. W. Bergholz. 2013. Landscape and meteorological factors affecting prevalence of three food-borne pathogens in fruit and vegetable farms. *Appl. Environ. Microbiol*. 79:588-600.

35. Strawn, L. K., Y. T. Grohn, S. Warchocki, R. W. Worobo, E. A. Bihn, and M. Wiedmann. 2013. Risk factors associated with *Salmonella* and *Listeria monocytogenes* contamination of produce fields. *Appl. Environ. Microbiol*. 79:7618-7627. doi:10.1128/AEM.02831-13

36. U. S. Food and Drug Administration. 14 November 2018. Recalls, market withdrawals, and safety alerts. Available at: https://www.fda.gov/safety/recalls/

37. Weis, J. and H. P. R. Seeliger. 1975. Incidence of *Listeria monocytogenes* in nature. *Appl. Microbiol*. 30(1): 19-32.

38. Weller, D., M. Wiedmann, and L. K. Strawn. 2015. Irrigation is significantly associated with an increased prevalence of *Listeria monocytogenes* in produce production environments in New York State. *J. Food Prot*. 78:1132-1141. doi:10.4315/0362-028X.JFP-14-584.

39. Weller, D., M. Wiedmann, and L. K. Strawn. 2015. Spatial and temporal factors associated with an increased prevalence of *Listeria monocytogenes* in spinach fields in New York State. *Appl. Environ. Microbiol*. 81:6059-6069. doi:10.1128/AEM.01286-15. doi:10.1128/AEM.02491-12.

40. Weller, D., A. Belias, H. Green, S. Roof, and M. Wiedmann. 2019. Landscape, water quality, and weather factors associated with an increased likelihood of foodborne pathogen contamination of New York streams used to source water for produce production. *Conflicts and Compromises between Food Safety Policies and Environmental Sustainability, a special issue of Frontiers in Sustainable Food Systems* 3(124).

41. Wilkes, G., T. A. Edge, V. P. J. Gannon, C. Jokinen, E. Lyautey, N. F. Neumann, N., Ruecker, A. Scott, M. Sunohara, E. Topp, and D. R. Lapen. 2011.

Associations among pathogenic bacteria, parasites, and environmental and land use factors in multiple mixed-use watersheds. *Water Res*. 45:5807-5825. doi:10.1016/j.watres.2011.06.021.

42. Willis, A. and J. Bunge. 2015. Estimating diversity via frequency ratios. *Biometrics*. 71(4):1042-1049.

43. Willis, A., B. D. Martin, D. Trinh, K. Barger, and J. Bunge. 2018. Breakaway: Species richness estimation and modeling. R package version 4.6.7. https://adw96.github.io/breakaway/

44. Zhu, Q., R. Gooneratne, and M. A. Hussain. 2017. *Listeria monocytogenes* in fresh produce: Outbreaks, prevalence and contamination levels. *Foods*. 6(21):1-11. doi:10.3390/foods603002

CHAPTER 4

ROOT CAUSE ANALYSIS CAN BE USED TO IDENTIFY AND REDUCE A HIGHLY DIVERSE *LISTERIA* POPULATION IN AN APPLE PACKINGHOUSE: A CASE STUDY

Published in: Food Protection Trends (2021), In Press

ABSTRACT

Root cause analysis (RCA) was utilized to identify *Listeria* elimination strategies in an apple packinghouse. While most of the *Listeria* was not persistent according to *sigB* allelic typing (i.e., 16 allelic types were isolated from 22 positive samples), the same packinghouse sites were continually positive. The root cause was identified as a limited understanding of how to eliminate *Listeria*. Based on these findings, we provided instructions on proper *Listeria* elimination strategies and supported implementation, including (i) increasing cleaning and sanitation from once to twice a week, (ii) use of quaternary ammonium compound (quat) powder around forklift stops and floor cracks, and (iii) removal of a dead-end pipe. Five samplings were conducted to test intervention effectiveness. While increased cleaning and sanitation frequency did not significantly reduce the log-odds of a site testing *Listeria* positive [11% (4/35) before and 13% (18/137) after, P=0.787 by logistic regression], the site-specific interventions appeared to have controlled the *Listeria* at the respective sites. Specifically, after utilizing quat powder, *Listeria* was not isolated from the forklift stops or floor crack (0/5 samples positive after versus 6/10 samples positive before intervention). Therefore, this study provides a roadmap for performing RCA and implementing interventions for controlling *Listeria* in packinghouses.

INTRODUCTION

Listeria contamination of fruits and vegetables is an issue that can be traced back to contamination from pre-harvest or post-harvest sources (e.g., packinghouses, processing plants). There have been 32 recalls caused by *L. monocytogenes* contamination of fruit and vegetable products in 2019 and 2020 *(38)*. For example, there was a recall of apples in 2019 caused by *L. monocytogenes* contamination, which involved 2 bulk bins and 2,297 cases of apples *(37)*. In addition, there have been several *L. monocytogenes* illness outbreaks linked to produce *(8, 9, 10, 11, 12, 13)*. A *L. monocytogenes* illness outbreak linked to cantaloupes in 2011 *(8)* and a *L. monocytogenes* illness outbreak linked to caramel apples in 2014 *(10)* were of particular relevance; both outbreaks were linked to contamination in the packing environment *(10, 24)*.

Packinghouses (i.e., a facility that washes, sorts, culls, grades, and/ or packages produce but does not perform any processing steps) *(15, 17, 27, 33, 35)* and food processing facilities *(5, 16, 19, 22, 25, 32)*, in general have been shown to often harbor *Listeria*. *Listeria* can be brought into the packing environment on produce or produce bins (via the preharvest environment), with employees serving as fomites, via forklifts or other transportation vehicles, among other routes *(16, 17, 28, 34, 35, 48, 49)*. Once *Listeria* has been transferred into a packinghouse, it can contaminate different pieces of equipment and the packinghouse structure itself (e.g., floors, walls, or drains). A lack of *Listeria* control in a packinghouse can be identified through environmental sampling when the same sites within a packinghouse are repeatedly positive. Through further subtyping (e.g., using pulsed field gel electrophoresis or whole genome sequencing) of the *Listeria* present in "repeat positive" sites, two scenarios can be identified: (i) "persistent *Listeria*" and (ii) "persistent transient *Listeria*." "Persistent *Listeria*" refers to the *Listeria* not removed from the packing environment (e.g., through cleaning and

sanitation, facilitated by proper sanitary design of equipment), allowing it to survive and replicate over time and facilitating its spread in the packing environment as well as possible contamination of finished product *(16)*. On the other hand, "persistent transient *Listeria*" describes the scenario when the same sites within the packing environment are continuously positive but not necessarily with the same subtype of *Listeria* (e.g., due to continuous re-introduction of *Listeria* from the same sources) *(4)*. Direct traceback of product contamination to persistent transient *Listeria* populations is more difficult compared to persistent *Listeria* populations because of the larger number of subtypes present. However, similar to persistent *Listeria*, persistent transient *Listeria* can still lead to product contamination and subsequent recalls and illnesses. There is a third scenario referred to as "transient *Listeria*," which refers to the *Listeria* that enters the processing environment but is eliminated through routine cleaning and sanitation activities.

To avoid recalls and outbreaks it is necessary to identify sites of persistent *Listeria* and persistent transient *Listeria* within the packing environment and to subsequently implement corrective actions that eliminate contamination. However, only performing superficial "corrective" actions (e.g., re-cleaning and sanitation) may not truly eliminate contamination, leading to recurrence of contamination linked to the same root cause (e.g., an introduction vector that was not addressed). Therefore, identifying the root cause of the contamination or persistent *Listeria* in the packing environment is essential *(36)*. As such, this study aimed to develop, implement, and test a protocol for performing root cause analysis (RCA) to identify interventions to eliminate *Listeria*. While the study reported here used an apple packinghouse as a model system, the RCA protocol and procedures will be broadly applicable to different produce and food processing operations.

MATERIALS AND METHODS

Packinghouse characteristics and study design. An apple packinghouse in the northeastern US was investigated in this case study to evaluate a root cause analysis (RCA) procedure for reducing or eliminating *Listeria* in produce operations. The packinghouse is approx. 25,000 ft^2 (approx. 2,300 m^2), has a single packing line, and runs from September to May each year. The packinghouse is regulated under the US FDA rule for preventive controls for human food *(1)*, as >50% of apples they pack come from growers not under their management; the packinghouse is not located on a primary production farm *(2)*. The packinghouse performed daily dry cleaning (i.e., removal of leaves and debris from the equipment and floor) and performed full "wet" cleaning and sanitation weekly. The same employees who worked on the line during production also performed cleaning and sanitation activities.

Initial data on *Listeria* detection in the same packinghouse investigated in the case study reported here was reported by Sullivan and Wiedmann *(33)*, who designated this facility as "packinghouse A." This previously reported data was utilized in the current study (i) as proof of *Listeria* presence in the packinghouse, (ii) for the identification of sampling sites with a higher likelihood of being positive for *Listeria*, and (iii) as evidence in the RCA. In addition to this previously collected data, sample collection was also performed as part of the current study to test the effectiveness of the interventions.

In the current study, root cause analysis was performed to help identify interventions for reducing or eliminating persistent *Listeria* in the packing environment. To test the effectiveness of selected interventions, sampling of the packinghouse environment was performed before and after intervention implementation (see Figure 4.1 for a timeline of sampling and intervention implementation).

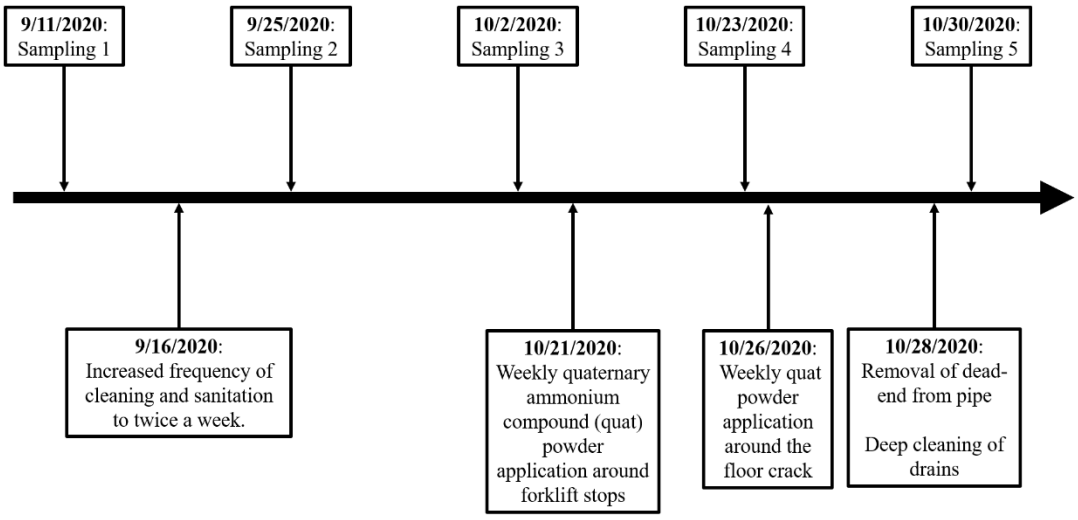


Figure 4.1. Sampling events and intervention implementation timeline for eliminating *Listeria* in the apple packinghouse.

Root cause analysis and intervention implementation. RCA was performed using a fishbone diagram and the "5 why's procedure," utilizing the procedure recently described by Belias and Wiedmann *(4)*. RCA was performed in a 1.5 h meeting with the packinghouse manager, the quality assurance manager, the maintenance manager, and two members from the Cornell team (i.e., the RCA team) on October 25th, 2019. Briefly, a pre-constructed fishbone diagram was utilized as a brainstorming tool in the RCA (Figure 4.2); this fishbone diagram was adapted from Belias and Wiedmann *(4)* and modified to be specific to produce operations (i.e., packinghouse and processing facilities). There are 7 major bones in the fishbone diagram that represent high-level categories that can lead to *Listeria* problems: (i) company practices/ food safety culture, (ii) personnel, (iii) facilities, (iv) cleaning and sanitation, (v) produce introduction, (vi) packinghouse equipment, and (vii) produce processing equipment. The apples were not further processed, so the major bone "vii: produce processing equipment" was deemed irrelevant and was therefore removed. Within each of these major bones, there are

minor bones specifying more specific qualities of a produce operation. As with the major bones, these were modified, removed, or added based on relevance; for instance, forklifts and forklift stops were added as minor bones under the “facilities” major bone.

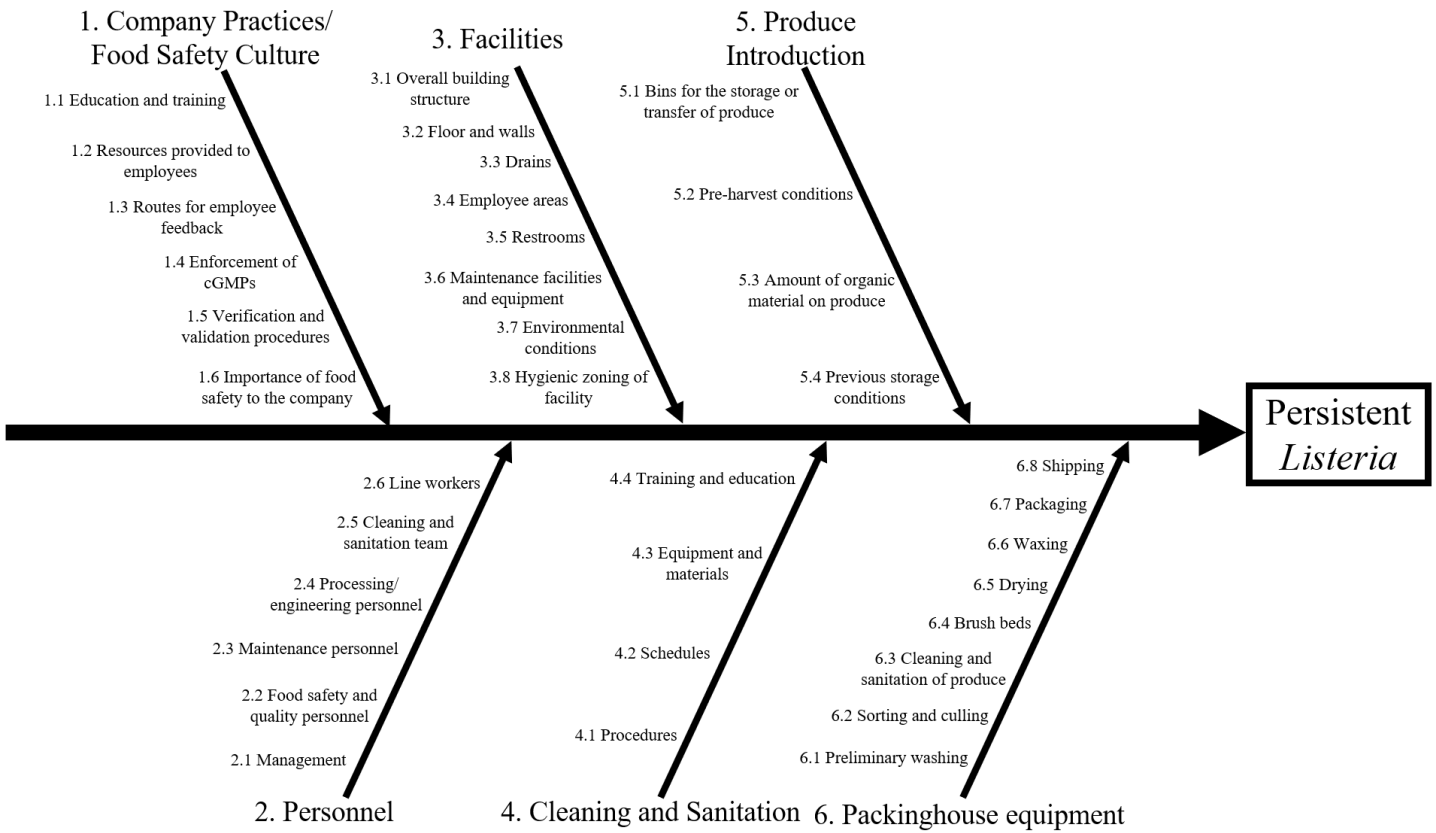


Figure 4.2. Fishbone diagram utilized in root cause analysis to identify interventions to eliminate *Listeria* in the packinghouse environment. This fishbone diagram was adapted from: Belias, A. and M. Wiedmann. 2021. Hazards, risks, and challenges of *Listeria* in the food supply. Food Safety Management in Practice. In press.

Based on the RCA, the following interventions were tested: (i) increasing cleaning and sanitation from once a week to twice a week, (ii) weekly (applied on Monday of each week) use of quat powder around forklift stops and floor cracks, (iii) a site-specific niche (a dead-end pipe) removal, and (iv) implementing a deep cleaning protocol in the drains.

Sample collection. To test the effectiveness of intervention implementation, we performed sampling of the packinghouse environment (Figure 4.1) in addition to previous pre-intervention sampling reported by Sullivan and Wiedmann *(33)*. Five samplings were performed from September 1st, 2020 to October 30th, 2020, including one sampling before and 4 samplings after intervention implementation. All sampling events were performed on Fridays to test if the second cleaning and sanitation event in a given week (performed on Wednesdays) was associated with a lower log odds of a sample testing positive for *Listeria*; in pre-intervention sampling a higher percent of *Listeria* positive samples were seen in end-of-week sampling as compared to early-in-week samples. On each visit, 35 samples were collected from the "wet-area" of the packing house (i.e., the area with the bin dump, flume, brush beds and waxing equipment); only 32 samples were collected on the final sampling event (October 30th) because 3 sampling sites were no longer present due to removal of the dead-end pipe. All samples were collected from zones 2 and 3 (i.e., no food contact surface samples were collected). The sampling sites included: (i) 11 PVC pipe samples, (ii) 9 drain samples, (iii) 2 forklift stop samples, (iv) 1 floor crack sample, and (v) 12 equipment frame samples (Table S1). Samples were collected using sponges hydrated with Dey-Engley broth (3M, Saint Paul, MN). Sampling was performed at least 2 hours into production. All samples were transported back to the lab on ice, then stored at 4°C until processing.

TABLE 4.1. List of root causes and corrective actions identified, and interventions tested to control repeatedly isolated *Listeria* in the packinghouse.

Identified Problem	Minor bone	Root Cause	Corrective Actions[a]		Interventions[b]
			Short-term	Long-term	
Increase in *Listeria* positives over the course of a week (between cleaning and sanitation events)	Cleaning and sanitation protocols and schedules	Not knowing how frequently cleaning and sanitation must be performed to control *Listeria*	(1) Increasing the frequency of cleaning and sanitation of the wet end to more than once a week (2) Improved cleaning and sanitation protocols (e.g., use of foamers to apply chemicals)	Hiring additional employees for a cleaning and sanitation (as opposed to having the line workers stay after their shifts)	Increasing cleaning and sanitation frequency of the wet end to twice a week (once on the weekend and once on Wednesday)
A persistent transient *Listeria* population around the forklift stops	Forklift stops	The forklift stops trap moisture and nutrients underneath to support *Listeria* growth and are difficult to clean and sanitize	(1) Use of quat[c] powder around forklift stops (2) Seasonal removal and deep cleaning of the forklift stops	Raise forklift stops to facilitate easier cleaning and sanitation	Weekly quat[c] powder application around the forklift stops
A persistent transient *Listeria* population in a floor crack	Floor crack	Floor cracks created when drains were constructed were never filled in	Use of quat[c] powder around floor crack	Filling in crack to eliminate the harborage point	Weekly quat[c] powder application in the floor crack
A persistent *Listeria* population present in the catch pan and a dead-	Catch pan area under brush beds	The catch pain area has several plastic-metal and metal-metal	Removal of the dead-end from the pipe	Redesign the catch pan area to eliminate juncture points that	Removal of the dead-end from the pipe

end pipe that drains the catch-pan		junctures, as well as a dead-end pipe that are difficult to clean and sanitize		can harbor *Listeria*	
Persistent and persistent transient *Listeria* populations in a square drain and a connected trench drain under the bin dump/ flume	Drains	The drains were not designed to be easily cleaned and sanitized	(1) Use of a cleaner sanitizer in a deep cleaning event (2) Use of quat[c] powder around the drain or quat[c] ring in the drain	Installation of a diaphragm pump in the square drains to facilitate draining and prevent the buildup of organic matter	Use of a cleaner sanitizer to perform deep cleaning of the drains

[a]Corrective actions refer to actions that were identified as potential solutions to address each root cause. These corrective actions were not all implemented in the current study due to time and budget constraints but provide examples of additional steps that can be taken to control *Listeria* associated with the identified root causes.

[b]Interventions refer to the actions taken in the current study to control *Listeria* associated with the identified root causes.

[c]quat = quaternary ammonium compound

***Listeria* enrichment and isolation.** All samples were processed within 24h of collection using a modified version of the FDA BAM method *(18)*. Briefly, 90 mL of buffered *Listeria* enrichment broth (BLEB, BD, Franklin Lakes, NJ) was added to each sponge sample, followed by stomaching at 230 RPM for 1 minute and incubation at 30°C for a total of 48h. After the initial 4h of incubation, 360 µL of *Listeria* selective enrichment supplement (LSES, Oxoid, Basingstoke, UK) was added. At 24h and 48h into incubation, 50 µL of the enriched samples are streaked for isolation onto modified Oxford agar (MOX, BD) and *Listeria monocytogenes* plating medium (LMPM, Biosynth International, Itasca, IL). The MOX plates were incubated at 30°C for 48h and the LMPM plates were incubated at 35°C for 48h. After incubation, characteristic *Listeria* colonies were sub-streaked from the MOX and LMPM plates onto brain heart infusion agar plates (BHI, BD); characteristic *Listeria* colonies on MOX are dimpled and pewter and characteristic *L. monocytogenes* colonies on LMPM are round and blue. Up to 16 characteristic colonies per sample were sub-streaked onto BHI, such that up to 4 colonies were selected from 24h MOX plates, 24h LMPM plates, 48h MOX plates, and 48h LMPM plates. The BHI plates were incubated at 37°C for 24h.

***sigB* sequencing and allelic typing.** PCR and subsequent sequencing of a 660 bp fragment of *sigB* was performed on all characteristic *Listeria* colonies sub-streaked to BHI for species identification; allelic type (AT) assignment based on the *sigB* sequence was performed for a preliminary assessment of the *Listeria* subtypes present in the packinghouse *(16)*. PCR and sequencing were performed according to the protocol described by Sullivan and Wiedmann *(33)*. All isolates confirmed as *Listeria* were stored as 15% glycerol stocks at -80°C.

Statistical analysis. All data visualization, cleaning, and analyses were performed in R version 4.0.0 *(27)*. Logistic regression was performed to determine if there were significant differences in the percent of samples positive for *Listeria* before

and after increasing cleaning and sanitation frequency to twice a week. Date of sampling and if the sample was collected before or after increased cleaning or sanitation were tested for inclusion in the models as potential explanatory factors. The model outcome was the percentage of samples positive for *Listeria*. To identify which of the explanatory factors were associated with the outcome, univariable logistic regression was first performed. Any variable with $P<0.1$ was then included in a multivariable regression model. To identify which explanatory factors should be included in the final multivariable logistic regression model, backwards selection was performed to identify the model with the lowest Bayesian Information Criterion (BIC) value. For a model to be selected as the final model, its BIC value had to be at least 2 less than the next simplest model.

In addition, the relationship between the increased cleaning and sanitation frequency and the percent of *Listeria* positive samples was also assessed using the sampling data collected over multiple seasons before intervention implementation to determine if the interventions were effective compared to historical data. To do so, the dataset in the current study was combined with the *Listeria* presence/absence data reported by Sullivan and Wiedmann *(33)* for the same packinghouse. However, the sites from Sullivan and Wiedmann *(33)* included in the analyses in the current study were reduced to only include sites collected from the "wet-end" of the packinghouse, as this was the only area sampled in the current study. In addition, only samplings on Thursdays or Fridays were included in the analyses performed here to determine the difference in samplings conducted after the second weekly cleaning and sanitation event (performed on Wednesdays). Logistic regression was used to determine the relationship between the log odds of a *Listeria* positive sample and if the sample was collected before or after increasing the frequency of cleaning and sanitation. The same procedure for logistic regression described above was used. No statistical analyses were performed to test the

quat powder, pipe removal, and drain cleaning interventions because there were too few samples.

RESULTS

Root cause analysis. At the RCA meeting an RCA team was assembled, which included the packinghouse manager, the quality assurance manager, the maintenance manager, and 2 members of the Cornell team. The RCA team reviewed the historical results to identify instances of repeat *Listeria* isolation (i.e., both persistent and persistent transient) in the packinghouse and opportunities for corrective actions. The historical results showed, that while there were persistent *Listeria* strains present in the packinghouse in the first years of sampling (2017 and 2018; based on whole genome sequencing data), the persistent strains were no longer present in the following year (2019) *(34)*. However, there was a pattern of persistent transient *Listeria*, as indicated by the repeat isolation of *Listeria* of non-matching subtypes (according to WGS) from the same sites (e.g., drains, dead-end pipe and catch pan area, forklift stops) within the packinghouse. In addition, the historical results showed (i) the majority of the positives in the packinghouse were from the wet-end (i.e., the area with the dump tank, flume, brush beds, and waxing equipment), (ii) there tended to be a greater percentage of positive samples in samplings conducted at the end of the week compared to the beginning of the week (which is closer to the weekly cleaning and sanitation, which was originally performed on Saturday or Sunday), (iii) deep square drains in the packinghouse were commonly positive, however after a deep cleaning event during the first year of sampling, a decrease in drain positives was observed, (iv) forklift stops at the dump tank loading area were commonly positive, and (v) sites by the catch pan below the brush bed, which is drained by a dead-end pipe, were commonly positive.

Next, brainstorming was performed by reviewing the fishbone diagram (Figure

4.2). Major bones (i.e., bones “i” to “vi”; see Figure 4.2) were reviewed and prioritized in order of importance; the following 3 major bones were prioritized in this case: (iii) facilities, (iv) cleaning and sanitation, and (vi) packinghouse equipment. Within each of these major bones, the relevant minor bones were discussed to determine their likelihood of contributing to the repeat isolation of *Listeria*. After review and discussion of historical results with the RCA team, the minor bones that were potential causes requiring further explorations were determined. Once this was completed for all relevant minor bones, the 4 minor bones most likely to be the root cause of the repeat *Listeria* positive sites were identified. The 4 minor bones selected as being the most likely contributors to repeat isolation of *Listeria* in this case were (i) cleaning and sanitation protocols and schedules, (ii) the catch pan area (under the “packinghouse equipment” major bone), (iii) forklift stops (under the “facilities” major bone), and (iv) drains (under the “facilities” major bone). For each of these 4 minor bones, the “5 why’s procedure” was performed by asking: “what part of this procedure likely contributed to the persistent *Listeria*?” and “why is this procedure set up the way it is?” Then, we continually asked 5 additional “why” questions to get to the actual root cause. For instance, if the identified problem was a persistent *Listeria* population at the end of the catch pan and dead-end pipe, the “5 why” questions may be: (i) why is the persistent *Listeria* population found in this area? Because the *Listeria* is living in the dead-end; (ii) why is the *Listeria* living in the dead-end? Because moisture, apple juices/ organic matter, and *Listeria* cells get trapped in the dead-end; (iii) Why do these things get trapped in the dead-end? Because there is no easy way for these things to be removed from the dead-end (iv) Why is there no easy way for these things to be removed from the dead-end? Because it is difficult to get cleaning and sanitation chemicals, as well as brushes for mechanical cleaning, to reach the dead end; and (v) why is the pipe and dead-end designed the way it is and why is a dead-end pipe in use in the facility? From

there, long- and short-term corrective actions to eliminate the root cause(s) were identified. These corrective actions were then prioritized based on the cost and ease of implementation. A description of the root causes, long and short-term corrective actions that were identified, and the interventions tested in the packinghouse can be found in Table 4.1. See Figure 4.1 for the intervention implementation schedule.

***Listeria* population.** Overall, 13% (22/172) of samples were positive for *Listeria* spp. (including *L. mono-cytogenes*) across the 5 sampling events performed as part of the study reported here. The 22 positive samples came from 9 sampling sites; these sites include (i) 3 sites from a single dead-end PVC pipe that drained the catch pan under the brush beds, (ii) 2 forklift stops, (iii) 1 floor crack, (iv) and 3 sites within drains (2 sites within a single square drain and 1 site from a trench drain leading into the square drain) (Table 4.2).

TABLE 4.2. *Listeria sigB* allelic types (ATs) isolated on each sampling date from the positive samples in the packinghouse.

Site ID[a]	Site Description	*sigB* ATs Historically Isolated from Each Site[b]	*sigB* ATs Isolated on Each Sampling Date[c,d]				
			Sampling 1: 9/11/20	Sampling 2: 9/25/20	Sampling 3: 10/2/20	Sampling 4: 10/23/20	Sampling 5: 10/30/20
305	Pipe outlet	12, 61, 67	-	-	9	-	-
306	Dead-end portion of PVC pipe	Not sampled	-	-	9	-	-
309	Pipe inlet	12, 61, 67	-	-	9	9	9
318	Forklift stop, full bins	-	-	-	12	-	-
319	Forklift stop, empty bins	3, 57	23, 57, 61, 67	-	20, 24	-	-
320	Floor crack by a trench drain[e]	Not sampled	23, 61	163	-	31	-
328	Square drain, PVC pipe in drain[f]	2, 3, 12, 58, 60, 61, 103	22, 57, 67	-	4, 12, 58	-	58
329	Square drain, outer edge[f]	2, 3, 12, 58, 60, 61, 103	22, 57, 67	57, 61	7, 20, 57	-	20, 24, 58
335	Trench drain	1, 9, 57, 58, 61	-	7, 57, 60	3	20, 24	3, 20, 24, 57, 58

[a]Full descriptions of all sampling sites can be found in Table S1.

[b]Historical results were obtained from the following study: Sullivan, G. and M. Wiedmann. 2020. Detection and prevalence of *Listeria* in U. S. produce packinghouses and fresh-cut facilities. *J Food Prot*. 83(10):1656-1666.

[c]*Listeria sigB* ATs 57, 58, 60, 61, 67, and 103 are *L. monocytogenes*; *Listeria sigB* ATs, 22, 23, and 31 are *L. innocua*, and *Listeria sigB* ATs 1, 2, 3, 4, 7, 9, 12, 20, 24, and 163 are *L. seeligeri*.

[d] "-" indicates the sample was negative on the given sampling date.

[d]This is not the same trench drain as in site 335.

[f]Historical isolates represent both sites in the square drain: PVC pipe in the drain and the outer edge. Therefore, the same ATs are listed is the historical column for both sites.

The *Listeria* species isolated included *L. monocytogenes*, *L. innocua*, and *L. seeligeri*; *L. monocytogenes* was isolated from 11 samples, *L. innocua* was isolated from 5 samples, and *L. seeligeri* was isolated from 15 samples (with a total of 22 positive samples). Of the 22 positive samples, 9 had 2 species present (4 samples were positive for *L. monocytogenes* and *L. innocua*, while 5 were positive for *L. monocytogenes* and *L. seeligeri*).

sigB allelic typing was performed on a total of 199 isolates obtained from the 22 positive samples; these isolates yielded 16 *sigB* allelic types (ATs) (Table 4.2). Between 1 and 5 *sigB* ATs were isolated from each sample (mean=2 ATs per sample and standard deviation=1 AT per sample). In many cases, different *sigB* ATs were isolated from a given site on different sampling dates (Table 4.2). For instance, 6 *sigB* ATs were isolated from a forklift stop (Sample ID 319); each AT was only isolated on one sampling date (Table 4.2). In addition, among the 12 *sigB* ATs isolated from 2 sites within a single square drain and 1 site within the trench drain leading into the square drain, 6 ATs were only isolated on one sampling date (Table 4.2), suggesting a persistent transient population. In the catch pan area, *sigB* AT 9 (*L. seeligeri*) was isolated on 3 sampling dates from the inlet to the dead-end pipe, on 1 sampling date from the dead-end of the pipe, and on 1 sampling date from the outlet of the dead-end pipe (Table 4.2), suggesting persistence.

Intervention effectiveness. The effectiveness of each intervention was tested by performing sampling before and after intervention implementation (Figure 4.1). For the increased cleaning and sanitation frequency intervention (from once a week to twice a week), one sampling was conducted prior to and four samplings were conducted after implementation; 11% (4/35) of samples were *Listeria* positive before and 13% (18/137) of samples were *Listeria* positive after implementation. There was no significant difference in the log odds of a sample testing positive for *Listeria* before compared to

after implementation (*P*=0.787 based on logistic regression; Table S2). In addition, date of sampling (*P*=0.866) was not significant (Table S2). Since only 1 sampling was conducted prior to implementation, the percent of positive samples after implementation was also compared to the percent of positive samples from historical samplings in the same wet-area in this packing house *(33)*; only samplings conducted later in the week (after Wednesday when the 2nd weekly cleaning and sanitation was performed) were included in the analyses. Based on these historical results, 31% (25/81) of samples were *Listeria* positive prior to compared to 13% (18/137) of samples positive after intervention implementation. Based on univariable logistic regression, both (i) date of sampling (*P*<0.001) and (ii) if the samples were collected before or after intervention implementation (*P*=0.037) were significant. However, only date of sampling was retained in the final model (Table S3).

To test the effectiveness of an intervention that involved weekly application of quat powder around the forklift stops, 3 samplings were performed before implementation and 2 samplings were performed after implementation. One of the 2 forklift stops was positive on sampling 1 (9/11/20) and sampling 3 (10/2/20); no *sigB* AT was isolated from the forklift stop on more than one sampling date, indicating the presence of a persistent transient *Listeria* population (Table 4.2). Following the implementation of weekly quat powder application (samplings 4 and 5), *Listeria* was not isolated from this forklift stop (Table 4.2). The other forklift stop was positive on sampling 3 (10/2/20) and was also not positive after application of quat powder (Table 4.2). To test the effectiveness of weekly quat powder application to a floor crack repeatedly positive for *Listeria* in the current study, 4 samplings were performed before and 1 sampling was performed after implementation. *Listeria* was isolated from the crack on sampling 1 (9/11/20), sampling 2 (9/25/20), and sampling 4 (10/23/20); no *Listeria* with the same *sigB* AT was isolated from the crack on more than one sampling

date, indicating the presence of a persistent transient *Listeria* population (Table 4.2). In the sampling after implementation of the quat powder intervention (sampling 5), *Listeria* was not isolated from the floor crack (Table 4.2). Therefore, the weekly quat powder application around the forklift stops and floor crack appear to be effective at controlling *Listeria*, as *Listeria* was no longer isolated from these sites following implementation of this intervention.

As our initial study in this facility *(33)* found repeat positives in the outlet from the same dead-end pipe sampled in the current study (5/8 samples of the dead-end pipe were *Listeria* positive from October 2017 to April 2019 in the previous study), we also tested whether the removal of the dead-end portion of the pipe would lead to elimination of potentially persistent *Listeria* in this site; 4 samplings were performed before and 1 sampling was performed after. *Listeria* was isolated from the dead-end and the pipe outlet on sampling 3 (10/2/20) but was not isolated in the sampling after removal of the dead-end (i.e., sampling 5). However, *Listeria* was isolated from the pipe inlet both before removal of the dead-end (samplings 3 and 4; 10/2/20 and 10/23/20, respectively), as well as after (i.e., sampling 5; 10/30/20) (Table 4.2). Interestingly, only *sigB* AT 9 was isolated from the dead-end pipe (the pipe inlet, outlet, and the dead-end) in the current study, suggesting a persistent *Listeria* (Table 4.2). Since *sigB* AT 9 was isolated from the pipe inlet even after removal of the dead-end portion, *sigB* AT 9 was likely persistent and not remediated through intervention implementation; *sigB* AT 9 was not isolated from any other sites during the current sampling but was isolated from the trench drain leading into the square drain in 2018 *(33)*.

To test the effectiveness of a one-time deep cleaning event using a chlorinated cleaner in the square drain and a trench drain leading into the square drain with persistent transient *Listeria* (6 out of 12 *sigB* ATs isolated from the 3 drain sites were isolated only on 1 sampling date), 4 samplings were conducted prior to intervention implementation

and 1 sampling (sampling 5) was conducted after. *Listeria* was isolated from at least 1 of the 2 sites within the square drain on sampling 1 (9/11/20), sampling 2 (9/25/20), sampling 3 (10/2/20), and sampling 5 (10/30/20). *Listeria* was also isolated from the trench drain on sampling 2 (9/25/20), sampling 3 (10/2/20), and sampling 4 (10/23/20), as well as on sampling 5 (10/20/20) (Table 4.2). Therefore, *Listeria* was still isolated from all drain samples following utilization of a chlorinated cleaner in the drains. While all 6 *sigB* ATs isolated from the 3 drain sites on the sampling after using the chlorinated cleaner had also been isolated from at least one of the drain sites prior to intervention implementation, the large number of *sigB* ATs isolated from these sites (3 ATs from the 2 square drain sites and 5 ATs from the trench drain) is still indicative of a persistent transient population (Table 4.2). The exceptions would be the re-isolation of *sigB* AT 57 (isolated on 4 out of 5 samplings) and *sigB* AT 20 (isolated on 3 out of 5 samplings) which may indicate persistence (*Table 4.2*). However, further subtyping (e.g., using whole genome sequencing) would be needed to confirm these hypotheses.

DISCUSSION

Persistent or persistent transient *Listeria* contamination of packinghouse environments represent a considerable challenge for industry. Historical data *(33, 34)*, as well as additional sampling data collected as part of this study, indicated both persistent and persistent transient *Listeria* populations were present in the investigated packinghouse. Specifically, *Listeria* was commonly isolated from forklift stops, a dead-end pipe, and drains. In addition, based on historical results, the percentage of positive samples was higher at the end of the week compared to the beginning of the week (i.e., after the weekly cleaning and sanitation was performed). As such, this study used the packinghouse as a model to (i) implement a protocol for performing root cause analysis (RCA) to identify interventions to eliminate *Listeria*, and (ii) to implement interventions

and perform sampling to determine the effectiveness of the interventions. The following interventions were identified through RCA, implemented, and tested: (i) increasing cleaning and sanitation from once a week to twice a week, (ii) use of quat powder around forklift stops and floor cracks, (iii) a site-specific niche (a dead-end pipe) removal, and (iv) implementation of a deep cleaning protocol in the drains. The site-specific interventions (e.g., use of quat powder) appeared to be more successful at eliminating *Listeria* from the apple packinghouse compared to increasing the frequency of cleaning and sanitation. This apple packinghouse case study provides an example of how RCA could be performed to eliminate or reduce persistent or persistent transient *Listeria* populations from produce operations.

A persistent and persistent transient *Listeria* population was present in the apple packinghouse investigated in the current study. Overall, our findings show the facility used in the study reported here had a number of sites with evidence for persistent transient *Listeria*, as well as at least 1 site with evidence of persistence. While most of the *Listeria* in the packinghouse in the current study represents persistent transient *Listeria*, the same *sigB* ATs were isolated ≥3 times from the square drain and the dead-end pipe, indicating potentially persistent *Listeria*. The dead-end pipe represented 3 sampling sites (the pipe inlet, the pipe outlet, and the dead-end portion of the pipe). In the dead-end pipe, *sigB* AT 9 was isolated on 3 sampling events from the pipe inlet and on 1 sampling event from the dead end and the pipe outlet. While a typing method with better discriminatory power (e.g., pulsed field gel electrophoresis or whole genome sequencing) would be needed to confirm these isolates as truly the same, the repeat isolation of this AT within one single pipe (and the lack of isolation of any other *sigB* ATs from this site) is likely to indicate persistence. This pipe is a likely site for persistence, as the dead-end portion can accumulate apple debris and other organic matter that can support the growth of *Listeria*. Consequently, cleaning and sanitation

can be a challenge because it is difficult to reach the dead-end portion. In addition, the inlet to the pipe is taped into a catch pan; the adhesive from the tape and the portion of the PVC pipe that overlaps with the metal catch pan (on the inside of the PVC pipe) can also act as harborage sites. While only one study was identified which listed plastic tubing as a harborage point for persistent *Listeria (20)*, several studies have suggested equipment that is difficult to clean is a risk factor for persistent *Listeria (6, 14, 21)*. Our findings provides further support of the importance of sanitary design in controlling persistent *Listeria* populations in processing environments and suggest that complete root cause analyses should include a consideration as to why equipment with poor sanitary design is present in a given facility.

In addition, the site that included a square drain and a connected inflow trench drain, showed evidence for persistent and persistent transient *Listeria* populations. In the square drain and the connected trench drain, *sigB* AT 57 was isolated during 4 sampling events. While this may indicate persistence of this subtype, *sigB* AT 57 is a common *sigB* AT *(15, 28, 34, 33, 49)* and has been shown to be highly diverse *(3)*. As such, it is possible this does not truly represent a persistent *Listeria*. In addition, *sigB* AT 20 was isolated from at least 1 of the 3 drain sites during 3 sampling events, which may indicate persistence; while not as common as *sigB* AT 57, *sigB* AT 20 was also isolated from a forklift stop on 1 sampling in the current study and was isolated in 2 previous studies *(3, 28)*. Drains have been listed as harborage points for persistent *Listeria* in several previous studies *(7, 19, 20, 29)*. However, additional subtyping (e.g., whole genome sequencing) is still needed in the current study to confirm persistence. In addition to the persistent population, the isolation of a highly diverse population of *Listeria* from the 3 drain sites (i.e., 12 *sigB* ATs were isolated from the 3 sites across the 5 samplings) indicates there is also a persistent transient *Listeria* population present in the drain. The drain is located under the dump tank and flume, which deposits large

amounts of organic matter and debris into the drain. Organic matter and soils originating from outdoor environments (e.g., similar to those environments apples are grown in) are a known source of *Listeria (28, 34, 35, 48, 49)* and are therefore a likely contributor of the diverse *Listeria* population present in this drain. Furthermore, the drain is deep and has poor drainage, which creates a large number of harborage points that are difficult to clean and that have sufficient moisture and nutrients to support *Listeria* growth; this can lead to both the persistent and persistent transient *Listeria* populations.

A persistent transient *Listeria* population was isolated from one of the forklift stops sampled in the current study, as 6 different *sigB* ATs were isolated from the forklift stop and no AT was isolated more than once. The presence of a persistent transient *Listeria* population at the forklift stops can be expected, as the forklifts often go outside to pick up bins of apples to be run on the packing line. In this packinghouse there are no control measures (e.g., door foamers) for the forklift wheels that would prevent *Listeria* transfer into the packinghouse, and the outdoor environments in the northeast have been shown to harbor diverse *Listeria* populations *(3, 28)*. As such, this could facilitate the transfer of the observed persistent transient *Listeria* population in the current study. Sullivan and Wiedmann *(33)* also identified forklift stops as harborage sites for *Listeria*; 1 out of 2 of the forklift stops positive for *Listeria* in this prior study was the same forklift stop as discussed in the current study.

A persistent transient *Listeria* population was also isolated from the floor crack samples in the current study, as *Listeria* of 4 *sigB* ATs were isolated from the floor crack and no AT was isolated more than once. This floor crack is directly adjacent to a trench drain and was created when the drain was installed in the packinghouse. This is consistent with the findings of Murugesan et al (2015), who repeatedly isolated *Listeria* from a floor crack next to a trench drain. Therefore, while all floor cracks are likely harborage points, a floor crack's proximity to the drain may increase the likelihood of

Listeria being present in the floor crack itself (i.e., due to potential splash from the drain into the floor crack). However, more extensive sampling is required to determine if floor cracks adjacent to other high-risk areas (e.g., drains) are at a higher likelihood of becoming contaminated.

As such, PVC pipes (especially those with dead ends), drains, forklift stops, and floor cracks area a few examples of sites that should be included in environmental monitoring programs in produce operations. The contamination patterns (i.e., the diverse *Listeria* population and repeat *Listeria* isolation from the same sites) in the packinghouse investigated in the current study represent "persistent *Listeria*" and "persistent transient *Listeria*" populations. Persistent *Listeria* poses public health and business risks (e.g., recalls) because as the *Listeria* survives in the packing environment over time it can grow; as the *Listeria* grows it is more likely to be transferred to other areas in the packing environment (e.g., by employees or mobile pieces of equipment) and eventually contaminate product. Since persistent *Listeria* represent a single strain of *Listeria*, finished product contamination can be traced back to the packing environment through environmental and product testing and subsequent subtyping of isolates (e.g., via whole genome sequencing). While it is more difficult to link final product contamination to the environment when a persistent transient *Listeria* population is present (compared to persistent *Listeria*), it can represent an instance of continuous introduction of *Listeria* from the same sources. As such, a persistent transient population may indicate more stringent supplier verification programs are required, and in serious cases the identification of alternative suppliers may be needed. Control measures may also be necessary to prevent transfer of *Listeria* from employees, forklifts, distribution trucks, or storage crate, among other routes into the packing environment (e.g., captive boot programs, compartmentalization of forklifts).

It is important to note a highly diverse persistent transient *Listeria* population

may hide a persistent population present in the packing environment (i.e., if there are a large number of *Listeria* subtypes present at any given site, the chance of identifying the persistent subtype is less likely as compared to if only the persistent subtype was present). Regardless, identifying a persistent *Listeria* that is covered up by a persistent transient *Listeria* population is still possible through large sampling efforts, such as "swab-a-thons" that subtype multiple isolates from each positive site. However, as we characterized up to 16 isolates from each positive sample, it is unlikely that true persistence was "covered up" by persistent transient strains in this case.

Root cause analysis can be utilized to identify interventions to eliminate or reduce *Listeria* populations in the apple packing environment; however, multiple iterations of testing and intervention implementation may be required to reduce *Listeria* populations. RCA was utilized to identify likely root causes of frequent repeat *Listeria* detection in the apple packinghouse in the current study; a previous study as well as this study reported these issues represent a combination of (i) persistent *Listeria* (e.g., in a catch pan with an outflow pipe with a dead-end) as well as (ii) transient persistent *Listeria* (e.g., at the forklift stops). RCA provided a formal process for identification of possible root causes associated with (i) overall high frequency of *Listeria* detection and (ii) different areas where repeat isolation of *Listeria* was an issue. The identified root causes were then used to identity interventions deemed likely to reduce overall frequent *Listeria* isolation as well as frequent site specific isolation of *Listeria*. Overall, the facility implemented one plant-wide intervention and four site specific interventions, which are discussed in detail below. The root cause analysis and subsequent interventions were successful for the forklift stops and floor crack, but further iterations of the root cause analysis are required to control the *Listeria* populations in the drain and dead-end pipe, and to account for the increase in *Listeria* positives in the second half of the week between cleaning and sanitation events. In

addition, a multipronged approach that targets multiple interventions at the same time is likely needed to effectively reduce or eliminate persistent and persistent transient *Listeria* populations. In addition, further root cause analyses that explore the reasons for why certain practices were not implemented (e.g., selection of equipment with sanitary design) will be needed for long-term successful *Listeria* control.

The plant-wide intervention tested in the current study was to increase the cleaning and sanitation frequency from once to twice a week; this was implemented to address the increase in percentage of *Listeria* positive samples observed when sampling was conducted at the end of the week. No significant change in the log odds of isolating *Listeria* after intervention implementation was observed in the 2020 sampling results. Increasing the frequency of cleaning and sanitation can also increase the amount of moisture present in the packinghouse. The increase in moisture can allow for an increase in the growth of *Listeria* and may explain the lack of a significant reduction in *Listeria* following implementation of this intervention. In addition, while the adequacy of the cleaning and sanitation protocol in the packinghouse was assessed, no changes in the protocol were made besides increasing the frequency. It is possible improvements in how cleaning and sanitation are performed (e.g., use of a foamer to apply cleaners and sanitizers) may be needed to further reduce the percent of *Listeria* positive sites in the wet are of the packinghouse. However, as only one sampling was conducted prior to increased cleaning and sanitation in the 2020 sampling events, the lack of a significant relationship may also be a function of chance (due to low sample size) or a fluctuation in the percent of positive samples just prior to intervention implementation for some unrecorded reason (e.g., lower than usual prevalence of *Listeria* on the incoming apples). When taking into account historical results, while both increased cleaning and sanitation frequency and date of sampling were significant for the log odds of a sample testing positive for *Listeria* according to univariable regression, only date of sampling

was retained in the multivariable model. This is due to the fact that the date of sampling and increasing cleaning and sanitation frequency are related with one another (i.e., increased cleaning and sanitation was only performed on dates in 2020), and the date of sampling was better able to explain differences in the log odds of a sample testing positive for *Listeria*. There was a decrease in *Listeria* positive samples as time went on. This is consistent with the findings of Sullivan and Wiedmann *(33)*, where there was a lower percent of *Listeria* positive samples in the second half of the packing season compared to the first half of the packing season in all three packinghouses investigated, when comparing sites that were repeatedly positive. Therefore, as sampling of the packing environment continued, the percent of positive samples decreased. This may indicate the packers utilized information on which sites were at a higher likelihood of testing positive to implement corrective actions (e.g., more time spent during cleaning and sanitation) as the studies progressed. These findings also illustrate the difficulty of assessing the effectiveness of interventions without extremely large sampling data sets or without adequate pre-intervention sampling to determine the true percent of *Listeria* positive sites in a facility prior to implementation and if there was a meaningful change in the percent of *Listeria* positive samples after implementation.

Two of the four site-specific interventions did appear to be successful at controlling *Listeria*. For example, *Listeria* was no longer isolated from 2 forklift stops and 1 floor crack after weekly quat powder application at these sites. Murugesan et al. *(25)* also tested the effectiveness of using quat powder to reduce *L. monocytogenes* in a mushroom processing environment. While there was a reduction in the number of sites positive for *L. monocytogenes* in the facility, some floor sites were still positive; the authors hypothesized this was due to harborage of *Listeria* in the porous concrete *(25)*. This likely indicates quat powder is effective against transient persistent *Listeria*, but not against persistent *Listeria* because elimination of persistent *Listeria* requires true

elimination of the niche. Therefore, while quat powder is not expected to eliminate *Listeria*, it can be used as a short-term solution for sites that are difficult to modify (e.g., floor cracks and forklift stops) to reduce their ability to harbor and spread *Listeria*. Quat powder is expected to prevent *Listeria* spread (i.e., the quat powder serves as a barrier that inactivates *Listeria* cells freed from the original site upon vibrations or contact with mobile equipment or employees) from the site to other sites within the packing or processing environments. However, while this intervention appeared to be successful at controlling persistent transient *Listeria* in two locations (e.g., forklift stops and the floor crack) in the current study, it must be stated that sampling was performed over a relatively short time frame (approx. 2 months). As such, continued sampling over a longer period is required to confirm the long-term effectiveness of the intervention. It is also important to note that this study does not provide evidence that widespread use of quat powders on the floors of a facility will eliminate repeatedly isolated *Listeria* populations, rather long-term interventions (e.g., improved equipment and facility design) would likely be more effective in this case.

A chlorinated cleaner for "deep-cleaning" of floor drains identified as a site with both a persistent and a persistent transient *Listeria* population was also tested. Assessment of this intervention is highly relevant as floor drains are common sites of persistent or persistent transient *Listeria* in packing or processing environments *(15, 17, 33)*. Cleaner sanitizers (such as the chlorinated cleaner used in the current study) are typically used in areas with a high likelihood of being contaminated, as they prevent the spread of *Listeria* to other sites during the cleaning process. The drains (both the square drain and the trench drain leading into it) are located right below the dump tank, and as such, they are commonly covered in a large amount of organic matter from splash out of the dump tank putting them at high risk for *Listeria* contamination. The implemented intervention did not appear to be effective, however, only 1 sampling was performed

after intervention implementation. Therefore, continued use of the cleaner sanitizer may be necessary to see an effect, especially because the large amount of debris in the drain may take several cleaning rounds before it is removed. In addition, because of the high-risk nature of this site (i.e., the large amount of debris present and the poor sanitary design of the drain), it is possible complete elimination of *Listeria* may not be possible without re-designing the drain to eliminate niches and improve cleanability. Rather, measures to control spread of *Listeria* out of the drain should be utilized in the short term. Some control measures could include the use of a quat powder around the drain, a quat ring in the drain, and monitoring the drain for flooding that would disperse the *Listeria* to other areas of the packinghouse. Use of high-pressure hoses should also be avoided, especially in drains, as the high-pressure water facilitates splash and subsequent *Listeria* transfer from the drain to equipment. Installing a pump in the drain to improve its ability to drain could also improve cleanability.

We also identified a dead-end pipe appeared to have been contaminated with a persistent *Listeria*. While *Listeria* was no longer isolated from the pipe outlet after the removal of the dead-end portion of the pipe, the same subtype was still isolated from the pipe inlet. As such, it appears that while the dead-end portion does have potential to harbor *Listeria*, it was not the (only) harborage point in this case. Rather, the persistent *Listeria* is likely harbored in the pipe inlet (or upstream). More specifically, the pipe inlet is taped inside a hole in the catch pan (the tape is in the inside of the pipe inlet), which possibly allowed for the attachment of *Listeria* and made it difficult to eliminate the *Listeria* through cleaning and sanitation. In order to eliminate the *Listeria*, complete removal of the pipe prior to cleaning and sanitation or replacing the pipe for a new one at regular frequencies would be necessary. This specific situation illustrates how RCA, in many cases, must be an iterative process where the initial root cause or associated intervention identified may not be the correct one. In this case, it could be argued the

root cause was not correctly identified (if the root cause was “a dead-end pipe”) or the root cause was correctly identified (insufficient procedures to assure sanitary facility and equipment design), but the intervention was not correctly selected or implemented. As such, the root cause analysis should be reviewed with the newly acquired data (i.e., sampling data from the pipe and surrounding area following removal of the dead-end) to determine what further interventions should be applied. For instance, an appropriate intervention for the root cause of “insufficient procedures to assure sanitary facility and equipment design” would be to (i) correct all sanitary design deficiencies (including the aforementioned tape that holds the inflow pipe in place) and (ii) implement proactive procedures to assure sanitary equipment and facility design going forward. However, it should be noted that removal of all sanitary design deficiencies may take time, and as such, changes to the most critical pieces of equipment (i.e., most likely to lead to product contamination if *Listeria* was to develop harborage at the site) and the least expensive or time-consuming changes should be prioritized first.

Conclusions. The aim of this study was to utilize a RCA procedure to identify, implement, and test interventions to eliminate *Listeria* from an apple packinghouse with both persistent and persistent transient *Listeria* populations. RCA was found to be a useful strategy for identifying the initial cause of *Listeria* contamination in the packinghouse, in that 2 out of 4 site-specific interventions were effective at reducing *Listeria* populations from the respective sites. Specifically, the use of quat powder was effective at preventing *Listeria* isolation from (i) forklift stops and (ii) a floor crack. However, the use of a cleaner sanitizer in drains and the removal of a dead-end pipe did not eliminate persistent or transient persistent *Listeria* populations. These instances indicate, that while they were not initially successful, use of an iterative process to test several possible options to identify the true root cause or identify and implement effective interventions is often necessary to truly control *Listeria* in the packing

environment. Regardless, the RCA protocol tested in this study can be used by other packinghouses, produce processing facilities, or, with modifications, other food processing facilities to identify, implement, and test interventions for reducing or eliminating *Listeria* repeat isolation from within the operations.

ACKNOWLEDGMENTS

This work was funded by the US Department of Agriculture Specialty Crop Research Initiative (award number C00135GG). We would like to thank the packinghouse who participated in this study. We would also like to thank Samantha Bolten for her help with sample collection and processing.

Martin Wiedmann is a paid consultant for 3M. The authors have no other potential conflicts of interest to report.

REFERENCES

1. 21 C.F.R. § 117. 2021.
2. 21 C.F.R. § 1.227. 2021.
3. Belias, A., L. K. Strawn, M. Wiedmann, and D. Weller. 2021. Small produce farm environments can harbor diverse *Listeria monocytogenes* and *Listeria* spp. populations.

   J. Food Prot. 84:113-121.
4. Belias, A. and M. Wiedmann. 2021. Hazards, risks, and challenges of *Listeria* in the food supply. *Food Saf. Mgmt. Pract*. In press.
5. Beno, S. M., M. J. Stasiewicz, A. D. Andrus, R. D. Ralyea, D. J. Kent, N. H. Martin,

   M. Wiedmann, and K. J. Boor. Development and validation of pathogen environmental monitoring programs for small cheese processing facilities. *J. Food Prot*. 79:2095-2106.
6. Blatter, S., N. Giezendanner, R. Stephan, and C. Zweifel. 2010. Phenotypic and molecular typing of *Listeria monocytogenes* isolated from the processing environment and products of a sandwich-producing plant. *Food Control*. 21:1519–1523.
7. Cao, J., M. Clarke, R. Witkowsky, H. Lu, A. Sayedahaman, R. E. Levin, and L A. McLandsborough. 2006. Concentrations and tracking of *Listeria monocytogenes* strains in a seafood-processing environment using a most-probable-number enrichment procedure and randomly amplified polymorphic DNA analysis. *J. Food Prot*. 69:489–494.
8. Centers for Disease Control and Prevention. 27 August 2012. Multistate outbreak of listeriosis linked to whole cantaloupes from Jensen Farms, Colorado

(final update). Available at: https://www.cdc.gov/listeria/outbreaks/cantaloupes-jensen-farms/index.html. Accessed 29 February 2021.

9. Centers for Disease Control and Prevention. 27 January 2015. Wholesome Soy Products, Inc. sprouts and investigation of human listeriosis cases (final update). Available at: https://www.cdc.gov/listeria/outbreaks/bean-sprouts-11-14/index.html. Accessed 29 February 2021.
10. Centers for Disease Control and Prevention. 12 February 2015. Multistate outbreak of listeriosis linked to commercially produced, prepackaged caramel apples made from Bidart Bros. Apples (final update). Available at: https://www.cdc.gov/listeria/outbreaks/caramel-apples-12-14/index.html. Accessed 29 February 2021.
11. Centers for Disease Control and Prevention. 31 March 2016. Multistate outbreak of listeriosis linked to packaged salads produced at Springfield, Ohio Dole processing facility (final update). Available at: https://www.cdc.gov/listeria/outbreaks/bagged-salads-01-16/index.html. Accessed 29 February 2021.
12. Centers for Disease Control and Prevention. 15 July 2016. Multistate outbreak of listeriosis linked to frozen vegetables (final update). Available at: https://www.cdc.gov/listeria/outbreaks/frozen-vegetables-05-16/index.html. Accessed 29 February 2021.
13. Centers for Disease Control and Prevention. 9 June 2020. Outbreak of listeria infections linked to Enoki mushrooms. Available at: https://www.cdc.gov/listeria/outbreaks/enoki-mushrooms-03-20/index.html. Accessed 29 February 2021.

14. De Cesare, A., G. Manfreda, M. Macri, and C. Cantoni. 2007. Application of automated ribotyping to support the evaluation of *Listeria monocytogenes* sources in a Taleggio cheese producing plant. *J. Food Prot.* 70:1116–1121.
15. Estrada, E. M., A. M. Hamilton, G. B. Sullivan, M. Wiedmann, F. J. Critzer, and L. K. Strawn. 2020. Prevalence, persistence, and diversity of *Listeria monocytogenes* and *Listeria* species in produce packinghouses in three U.S. states. *J. Food Prot*. 83:277-286.
16. Ferreira, V., M. Wiedmann, P. Teixeira, and M. J. Stasiewicz. 2014. *Listeria monocytogenes* persistence in food-associated environments: epidemiology, strain characteristics, and implications for public health. *J. Food Prot*. 77:150-170.
17. Jorgensen, J., W. C. Joy, and K. Jovana. 2020. Prevalence of *Listeria* spp. in produce handling and processing facilities in the Pacific Northwest. *Food Microbiol*. 90:103468.
18. Hitchins, A. D., K. Jinneman, and Y. Chen. 31 October 2017. Detection of *Listeria monocytogenes* in foods and environmental samples, and enumeration of *Listeria monocytogenes* in foods, Chapter 10. In Bacteriological analytical manual (BAM). U.S. Food and Drug Administration, Washington, D.C.
19. Hoffman, A. D., K. L. Gall, D. M. Norton, and M. Wiedmann. 2003. *Listeria mono-cytogenes* contamination patterns for the smoked fish processing environment and for raw fish. *J. Food Prot*. 66:52-60.
20. Kabuki, D. Y., A. Y. Kuaye, M. Wiedmann, and K. J. Boor. 2004. Molecular subtyping and tracking of *Listeria monocytogenes* in Latin-style fresh-cheese processing plants.

    J. Dairy Sci. 87:2803–2812.

21. Keeratipibul, S., and P. Techaruwichit. 2012. Tracking sources of *Listeria contamination* in a cooked chicken meat factory by PCR-RAPD-based DNA fingerprinting. *Food Control*. 27:64–72.
22. Keto-Timonen, R., R. Tolvanen, J. Lunden, and H. Korkeala. 2007. An 8-year surveillance of the diversity and persistence of *Listeria monocytogenes* in a chilled food processing plant analyzed by amplified fragment length polymorphism. *J. Food Prot*. 70:1866-1873.
23. Liao, J., M. Wiedmann, and J. Kovac. 2017. Genetic stability and evolution of the sigB allele, used for *Listeria sensu stricto* subtyping and phylogenetic inference. *Appl. Environ. Microbiol*. 83:e00306-17.
24. McCollum, J. T., A. B. Cronquist, B. J. Silk, K. A. Jackson, K. A. O'Connor, S. Cosgrove, J. P. Gossack, S. S. Parachini, N. S. Jain, P. Ettestad, M. Ibraheem, V. Cantu, M. Joshi, T. DuVernoy, N. W. Fogg, J. R. Gorny, K. M. Mogen, C. Spires, P. Teitell,

    L. A. Joseph, C. L. Tarr, M. Imanishi, K. P. Neil, R. V. Tauxa, and B. E. Mahon. 2013. Multistate outbreak of listeriosis associated with cantaloupe. *N. Engl. J. Med*. 369:944-953.
25. Murugesan, L., Z. Kucerova, S. J. Knabel, and L. F. LaBorde. 2015. Predominance and distribution of a persistent *Listeria monocytogenes* clone in a commercial fresh mushroom processing environment. *J. Food Prot*. 78:1988-1998.
26. R Core Team. 2020. R: A language and environment for statistical computing. Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.

27. Ruiz-Llacsahuanga, B., A. Hamilton, R. Zaches, I. Hanrahan, and F. Critzer. 2021. Prevalence of *Listeria* species on food contact surfaces in Washington State apple packinghouses. *Appl. Environ. Microbiol*. In press.
28. Sauders, B. D., J. Overdevest, E. Fortes, K. Windham, Y. Schukken, A. Lembo, and M. Wiedmann. 2012. Diversity of *Listeria* species in urban and natural environments. *Appl. Environ Microbiol*. 78:4420-4433.
29. Stessl, B., K. Szakmary-Brandle, U. Vorberg, D. Schoder, and M. Wagner. 2020. Temporal analysis of the *Listeria monocytogenes* population structure in floor drains during reconstruction and expansion of a meat processing plant. *Int. J. Food Microbiol*. 314:108360.
30. Strawn, L. K., E. D. Fortes, E. A. Bihn, K. K. Nightingale, Y. T. Grohn, R. W. Worobo, M. Wiedmann, and P. W. Bergholz. 2013. Landscape and meteorological factors affecting prevalence of three food-borne pathogens in fruit and vegetable farms. *Appl. Environ. Microbiol*. 79:588-600.
31. Strawn, L. K., Y. T. Grohn, S. Warchocki, R. W. Worobo, E. A. Bihn, and M. Wiedmann. 2013. Risk factors associated with *Salmonella* and *Listeria monocytogenes* contamination of produce fields. *Appl. Environ. Microbiol*. 79:7618-7627.
32. Strydom, A., R. Vorster, P. A. Gouws, and R. C. Witthuhn. 2016. Successful management of *Listeria* spp. in an avocado processing facility. *Food Control*. 62:208-215.
33. Sullivan, G. and M. Wiedmann. 2020. Detection and prevalence of *Listeria* in U. S. produce packinghouses and fresh-cut facilities. *J. Food Prot*. 83:1656-1666.
34. Sullivan, G., R. H. Orsi, E. Estrada, L. Strawn, and M. Wiedmann. 2021. Whole genome sequencing-based characterization of *Listeria* isolates from produce

packinghouses and fresh-cut facilities suggests both persistence and re-introduction of fully virulent

L. monocytogenes. *In preparation*.

35. Tan, X., T. Chung, Y. Chen, D. Macarisin, L. LaBorde, and J. Kovac. 2019. The occurrence of *Listeria monocytogenes* is associated with built environment microbiota in three tree fruit processing facilities. *Microbiome*. 7:115.
36. The PEW Charitable Trust. 2020. A guide for conducting a food safety root cause analysis: Approaches for investigating contamination incidents and preventing recurrence. Available at: https://www.pewtrusts.org/en/research-and-analysis/reports/2020/03/a-guide-for-conducting-a-food-safety-root-cause-analysis. Accessed 3 March 2021.
37. U. S. Food and Drug Administration. 28 October 2019. North Bay Produce voluntarily recalls fresh apples because of possible health risk. Available at: https://www.fda.gov/safety/recalls-market-withdrawals-safety-alerts/north-bay-produce-voluntarily-recalls-fresh-apples-because-possible-health-risk. Accessed 3 March 2021.
38. U. S. Food and Drug Administration. 3 March 2021. Recalls, market withdrawals, and safety alerts. Available at: https://www.fda.gov/safety/recalls. Accessed 3 March 2021.
39. Weller, D., M. Wiedmann, and L. K. Strawn. 2015. Irrigation is significantly associated with an increased prevalence of *Listeria monocytogenes* in produce production environments in New York State. *J. Food Prot*. 78:1132-1141.
40. Weller, D., M. Wiedmann, and L. K. Strawn. 2015. Spatial and temporal factors associated with an increased prevalence of *Listeria monocytogenes* in spinach fields in New York State. *Appl. Environ. Microbiol*. 81:6059-6069.

CHAPTER 5

CROSS-VALIDATION INDICATES PREDICTIVE MODELS MAY PROVIDE AN ALTERNATIVE TO INDICATOR ORGANISM MONITORING FOR EVALUATING PATHOGEN PRESENCE IN SOUTHWESTERN US AGRICULTURAL WATER

Published in: Frontiers in Water, Environmental Water Quality Special Issue (2021), In Press

ABSTRACT

Pathogen contamination of agricultural water has been identified as a probable cause of recalls and outbreaks. However, variability in pathogen presence and concentration complicates the reliable identification of agricultural water at elevated risk of pathogen presence. In this study, we collected data on the presence of *Salmonella* and genetic markers for enterohemorrhagic *E. coli* (EHEC; PCR-based detection of *stx* and *eaeA*) in southwestern US canal water, which is used as agricultural water for produce. We developed and assessed the accuracy of models to predict the likelihood of pathogen contamination of southwestern US canal water. Based on 169 samples from 60 surface water canals (each sampled 1-3 times), 36% (60/169) and 21% (36/169) of samples were positive for *Salmonella* presence and EHEC markers, respectively. Water quality parameters (e.g., generic *E. coli* level, turbidity), surrounding land-use (e.g., natural cover, cropland cover), weather conditions (e.g., temperature), and sampling site characteristics (e.g., canal type) data were collected as predictor variables. Separate conditional forest models were trained for *Salmonella* isolation and EHEC marker detection, and cross-validated to assess predictive performance. For *Salmonella*, turbidity, day of year, generic *E. coli* level, and % natural cover in a 500-1,000ft (approx.

150-300m) buffer around the sampling site were the top 4 predictors identified by the conditional forest model. For EHEC markers, generic *E. coli* level, day of year, % natural cover in a 250-500ft (approx. 75-150m) buffer, and % natural cover in a 500-1,000ft (approx. 150-300m) buffer were the top 4 predictors. Predictive performance measures (e.g., area under the curve [AUC]) indicated predictive modeling shows potential as an alternative method for assessing the likelihood of pathogen presence in agricultural water. Secondary conditional forest models with generic *E. coli* level excluded as a predictor showed <0.01 difference in AUC as compared to the AUC values for the original models (i.e., with generic *E. coli* level included as a predictor) for both *Salmonella* (AUC=0.84) and EHEC markers (AUC=0.92). Our data suggests models that do not require the inclusion of microbiological data (e.g., indicator organism) show promise for real-time prediction of pathogen contamination of agricultural water (e.g., in surface water canals).

INTRODUCTION

Salmonella spp. and pathogenic *Escherichia coli* (such as enterohemorrhagic *E. coli*; EHEC) are common etiological agents of foodborne outbreaks and recalls linked to produce commodities. From 2004 to 2012, *Salmonella* has caused 71 and 40 outbreaks linked to produce in the United States (US) and European Union (EU), respectively *(12)*. During the same time frame, pathogenic *E. coli* have caused 46 and 7 outbreaks linked to produce in the US and EU, respectively *(12)*. Both livestock and wildlife have been identified as possible sources of *Salmonella* and EHEC in preharvest produce environments *(16, 25)*, with fecal matter from livestock operations being a known source of foodborne pathogens in surface water *(16, 25, 41)*. As such, application of surface waters to in-field produce has been identified as a potential route for pathogen contamination of produce *(13, 38, 43, 49)*. In fact, several outbreaks are thought to have been caused by application of contaminated water to preharvest produce *(23, 67, 68, 69)*, including a 2018 *E. coli* O157:H7 outbreak linked to romaine lettuce grown in Arizona, which caused 210 illnesses and 5 deaths *(69)*.

While enteric pathogens, such as EHEC and *Salmonella*, are known surface water contaminants, they are present sporadically and at low levels, complicating detection and limiting the value of testing surface water for pathogens *(30, 47)*. Instead, monitoring programs often test for indicator organisms, which are used to assess the hygienic quality of water and the likelihood of fecal contamination; indicator organisms include enterococci, fecal coliforms, and generic *E. coli (30, 47)*. *E. coli* is used as an indicator for monitoring fecal contamination in agricultural water by industry agreements and by government regulations *(11, 18, 21, 62, 70)*; see Table 5.1 for details.

Table 5.1. Microbial quality requirements or regulations established in different countries or regions for agricultural water applied to pre-harvest produce.

Region	Organization	Requirement	Reference
United States[a]	US FDA	20 water samples over a 2 to 4-year period must be collected: (i) geometric mean of <126 CFU generic *E. coli*/ 100 mL and (ii) a statistical threshold value (i.e., the 90th percentile) of <410 CFU generic *E. coli*/ 100 mL	*14*
Europe	European Union	<100 CFU generic *E. coli*/100 mL of water[c]	*18*
British Columbia, Canada	British Columbia Ministry of Agriculture	<77 CFU generic *E. coli*/100 mL of water[c]	*62*
Australia	Freshcare	<100 CFU generic *E. coli*/ 100 mL of water[c]	*21*
California and Arizona (leafy greens only)[b]	Leafy Greens Marketing Agreement (LGMA)	Furrow irrigation or overhead irrigation applied >21 days prior to harvest: 100 mL of water should be collected at least monthly and the rolling geometric mean of the generic *E. coli* levels in the 5 most recent samples must be <126 CFU/ 100 mL and no sample may have an *E. coli* level >576 CFU/ 100 mL. Overhead irrigation applied <21 days prior to harvest: generic *E. coli* should not be detected in the water.	*11*

[a]As of January 2021, this policy is not being enforced and is currently under review (*70*). This requirement is established under the Food Safety Modernization Act (FSMA)

[b]This is a voluntary agreement.

[c]No sampling scheme for water collection is specified in this regulation.

There are however several drawbacks to using generic *E. coli* as an indicator of fecal contamination for surface water *(48, 72).* While some studies have established associations between generic *E. coli* levels and pathogen presence *(10, 28, 39, 57, 65, 74)*, several studies have not shown similar associations *(7, 19, 24, 56). E. coli* has also been shown to survive for extended periods of time in the preharvest environment *(3, 20, 41)*; therefore, a high level of generic *E. coli* does not necessarily indicate recent fecal contamination. High levels of indicator organisms also do not necessarily indicate the presence of pathogens, and alternatively, the absence or low levels of indicator organisms do not necessarily indicate the absence of pathogens *(7, 19, 24, 56, 57).* Lastly, generic *E. coli* testing takes approx. 24h to complete and generic *E. coli* levels in surface waters can vary substantially over short time periods *(40, 74);* therefore, it is impossible to know the generic *E. coli* level in irrigation water at the time of its application.

Two previous studies have proposed the use of machine learning models for predicting pathogen presence in agricultural water; Weller et al. *(75)* utilized machine learning models to predict *Salmonella* presence and EHEC marker detection in New York streams and Polat et al. *(51)* utilized machine learning models to predict *Salmonella* presence in Florida ponds. While previous studies have explored the use of machine learning models for predicting pathogen contamination in specific produce growing areas, further model development is essential to verify that machine learning represents a viable approach in different locations and types of surface waters. We thus collected data on surface water quality and used several approaches, including machine learning, to model the presence enteric pathogens in southwestern US canal water. Regression was used as a preliminary assessment to determine which variables were associated with pathogen presence. Conditional forest models were used for prediction because they can utilize large numbers of predictors and better able to handle complex

and messy data than regression models *(35)*. While modeling alone will not improve the safety of produce, these models can indicate when corrective actions (e.g., water treatments) should be applied to reduce the risk of recalls and illnesses associated with produce *(2, 54)*. The southwestern US was selected for this study, because (i) it is a major produce growing region, (ii) there has been a high-profile outbreak associated with romaine lettuce contamination linked to irrigation water, and (iii) there is limited information on microbial quality of southwestern US canals *(40, 74)*. As such, the specific objectives of this study were to (i) identify land use, water quality, weather, and other sampling site specific variables associated with *Salmonella* presence and EHEC marker detection (i.e., *stx* and *eaeA* detection) in southwestern US canal water, (ii) determine the feasibility of predicting the likelihood of *Salmonella* presence and EHEC marker detection in southwestern US canal water, and (iii) determine if only real-time variables (i.e., no microbial testing) can be used to predict the likelihood of *Salmonella* presence and EHEC marker detection.

MATERIALS AND METHODS

Experimental design. A longitudinal study was conducted to assess agricultural water quality in the southwestern US. Water was sampled from 60 canals that provide water for irrigation from January 30thto November 19th, 2018; sampling was performed approx. twice a week every one to two weeks except in July when no sampling was performed (see Figure S5.1 for the exact sampling dates). Each canal was sampled 1 to 3 times for a total of 169 samplings; all samples from a given canal were collected at the same site. Sampling sites were randomly selected from irrigation districts where produce was grown and where permission was given using ArcGIS. Ground truthing was then performed to identify a location as close as possible to the randomly generated GPS coordinates for each site. A site survey was conducted to collect information on

features present at a given site (see Table S5.1).

Sample collection and processing. At each sampling, two 10 L water samples (one per pathogen) and 1 L of water (for enumeration of *E. coli* and turbidity levels) were collected. Dissolved oxygen, pH, conductivity, and temperature of the canal water were measured using a Hach HQ40d meter (Loveland, CO, United States). Water surface flow was measured using the float method as described by Gore and Banning *(22).* After collection, all samples were put on ice until processing. The 10 L samples were processed <18 h after collection and the 1 L sample for *E. coli* level and turbidity was processed <6 h after collection.

Laboratory testing of all samples was performed as described in Weller et al. *(74).* Briefly, generic *E. coli* enumeration was performed on a 100 mL aliquot of the 1 L sample using the Colilert Quanti-Tray 2000 kit (IDEXX, Westbrook, ME, United States), according to the manufacturer's instructions. Water turbidity was measured using the Hach 2100Q Portable Turbidimeter. The 10 L water samples were processed using the modified Moore swab (mMS) method *(55).* Each water sample was gravity-filtered through a separate mMS, placed in a separate Whirl-Pak bag (Nasco, Fort Atkinson, WI, United States), and processed as described below for either *Salmonella* presence or EHEC marker detection.

***Salmonella* isolation**. Briefly, 225 mL of buffered peptone water with 20 mg/L novobiocin (BPW + N) was added to the Whirl-Pak bag with the mMS, followed by incubation at 35°C for 24h. BPW+N inoculated with *Salmonella* Typhimurium (FSL F6-0826; http://www.foodmicrobetracker.com) and uninoculated BPW+N were included as positive and negative controls, respectively. Following incubation, a 1 mL aliquot of the enrichment was transferred to a sterile tube and was shipped overnight on ice to Cornell University for further processing. Upon arrival, all enrichments were used within 2 h for a *Salmonella* screen using the BAX real-time *Salmonella* assay (Hygiena,

Wilmington, DE, United States). BAX PCR-positive samples were culture confirmed. One mL and 0.1 mL of the BPW+N enrichment were added to 9 mL of tetrathionate broth (TT; Oxoid) supplemented with 200 µL of I2-KI and 100 µL of Brilliant Green and 9.9 mL of Rappaport Vassiliadis broth (RV; Acros Organic, Geel, Belgium), respectively. The TT and RV broth were incubated in a shaking water bath at 42°C for 24h. Following incubation, 50 µL of each broth were streaked onto *Salmonella* CHROMagar (DRG International, Springfield, NJ, United States) and xylose lysine deoxycholate agar (XLD; Neogen, Lansing, MI, United States) plates (i.e., 1 plate for TT on CHROMagar, 1 plate for TT on XLD, 1 plate for RV of CHROMagar, and 1 plate for RV on XLD). The CHROMagar and XLD plates were then incubated at 37°C and 35°C, respectively, for 24h. After incubation, PCR of the *invA* gene was performed on presumptive *Salmonella* colonies according to the protocol described by Kim et al. *(34).* If possible, 2 characteristic *Salmonella* colonies per media type (mauve colonies on CHROMagar and black colonies on XLD) were selected for PCR (4 colonies in total). If no characteristic colonies were present, up to 12 non-characteristic colonies were selected for PCR (blue colonies on CHROMagar and red colonies on XLD). All isolates were stored as 15% glycerol stocks at -80°C.

EHEC marker detection. PCR-based detection of *stx* and *eaeA* from the mMS was performed. 225 mL of tryptic soy broth with 10 g/L casamino acids and 8 mg/L of novobiocin (TSB+N) was added to the Whirl-Pak bag with the mMS, followed by incubation at 41°C for 24h. TSB+N inoculated with *E. coli* O157:H7 (FSL F6-0699; http://www.foodmicrobetracker.com) and uninoculated TSB+N were included as positive and negative controls, respectively. Following incubation, a 1 mL aliquot of the enrichment was shipped overnight on ice to Cornell University for further processing. All enrichments were used within 2 h of arrival to perform a PCR screen using the BAX real-time Shiga-toxin producing *E. coli* (STEC) assay (Hygiena) according to the

manufacturer's instructions to determine if the *eaeA* and/or *stx1/2* genes were present in the sample. If both *eaeA* and *stx1/2* were detected in a sample, the sample was classified as positive for "EHEC markers." However, such results could indicate either (i) both genes were present in a single organism (indicating presence of EHEC) or (ii) genes were present in separate organisms (e.g., *eaeA* indicates enteropathogenic *E. coli* presence, *stx1/2* indicates STEC presence).

Land use data collection. Land use data around the sampling sites were extracted from the 2016 National Land Cover Database (NLCD; https://www.mrlc.gov/) and quantified using ESRI ArcGIS Pro 2.4.0. The percentage of land under (i) developed open space, (ii) developed (combines low-, medium-, and high-intensity developed cover), (iii) barren, (iv) natural (combines forest and wetland), (v) pasture/ hay, and (vi) crop cover at various intervals around each sampling site were calculated (Yang et al., 2018). The intervals considered were: <250 ft (<approx. 75m), 250-500 ft (approx. 75-150m), 500-1,000 ft (approx. 150-300m), 1,000-5,000 ft (approx. 300-1,525m), and 5,00-10,000ft (approx. 1,525-3,050m). These buffer areas were selected, as they most closely represent the distances included in the California and Arizona Leafy Green Marketing Agreements Food Safety Practices *(11)* metrics; while it would have been useful to characterize land use directly adjacent to the canals, an accurate map of the canal networks was not available for the study area. The number of concentrated animal feeding operations (i.e., an animal feeding operation with >1,000 animal units confined on a site for more than 45 days of the year; CAFOs) within 10,000 ft of each site was also calculated.

Weather data collection. Temperature, solar radiation, precipitation, wind speed, and vapor pressure data were obtained from the University of Arizona (cals.arizona.edu/AZMET/). ESRI ArcGIS Pro 2.4.0 was used to identify the weather station closest to each of the sampling sites. Weather data were cleaned in R version

4.0.0 *(52)* and used to calculate weather at the time of sample collection and for the (i) 0-12 h, (ii) 12-24 h, (iii) 1-2 d, and (iv) 2-3 d prior to sample collection. Due to the small amount of precipitation during the study, all precipitation variables were converted to a binary factor to indicate if there was precipitation (>0 mm) or if there was no precipitation (=0 mm).

Regression analysis. All data cleaning, visualization, and analyses were performed in R *(52).* A description of all variables used in analyses are provided in Table S5.1. All analyses were performed separately for *Salmonella* presence and EHEC marker detection. Logistic regression was used as a preliminary assessment to characterize associations between site specific (i.e., data on features present at each site, see Table S5.1 for details), water quality, land-use, and weather variables and *Salmonella* presence and EHEC marker detection. Conditional forest analysis was used to determine if these variables could be used to predict *Salmonella* presence or EHEC marker detection.

For logistic regression, normalization and scaling of all numeric variables was performed using the "caret" package *(36).* Univariable logistic regression was performed, using the "lme4" package *(5),* to determine which of the explanatory variables listed in Table S5.1 were associated with *Salmonella* and EHEC marker presence. The day of year (number of days since Jan 1st) and irrigation district were included in each univariable model as random effects to account for temporal and spatial autocorrelation. Following univariable regression, continuous variables with $P<0.1$ were included in a principal component analysis (PCA) for variable reduction. PCA was performed using the prcomp function, such that the number of components retained must explain ≥90% of the variation in the data and each retained variable could only have major loading on one principal component. PCA was performed separately for *Salmonella* and EHEC marker presence. One representative continuous variable from

each principal component, as well as all categorical variables significant at P<0.1 by univariable analysis (categorical variables cannot be included in PCA) were included in the initial multivariable logistic regression models (implanted using the "lme4" package; *5*). Day of year and irrigation district were included each multivariable model as random effects. Backwards selection based on AIC (Akaike Information Criterion) was performed; the final selected model was the simplest model with an AIC value that was at least 2 less than the next simplest model. Model fit was assessed using the protocol described by Beauvais et al. *(6)* to determine if model assumptions were met. Variance inflation factors were also calculated to test for multicollinearity.

Conditional forest analysis. Conditional forest analysis was used to determine if sampling site, water quality, land-use, and weather variables could be used to predict *Salmonella* presence and EHEC marker detection, as it can handle missing data, skewed data, and is robust to small sample sizes. Imputation could not be performed because >10% of observations were missing for some variables, and so imputation could introduce bias into the results. No additional machine learning algorithms were tested, as a comprehensive comparison of 23 learners for predicting enteric pathogen presence in New York streams found that conditional forest models performed well for the type of data used in the study presented here *(75).* The "mlr" *(9)* and "party" *(29, 60, 61)* packages were used for model training and testing. Oversampling was performed to account for imbalanced training data. Repeated (5 iterations) 5-fold cross-validation was used to tune hyperparameters (i.e., mtry, minbucket, and mincriterion) to maximize AUC (area under the curve) and minimize overfitting. For each forest, 20,001 trees were fit. Following hyperparameter tuning, models were trained, and model testing was performed. While a separate testing data set would have been preferable to better evaluate the predictive performance of these models, one was not available. Instead, cross-validation was performed as part of model training to estimate performance

measures. Variable importance scores were calculated *(60, 61)* and partial dependence plots were fit for the 4 top-ranked variables for each pathogen. While using conditional importance scores would have been preferable to account for correlation between variables, it could not be calculated due to missing data (>10% of observations were missing for some variables); as a result, the variable importance scores reported here may be biased by this correlation. Even with this limitation, we determined conditional forest was a good option in this case due to its ability to handle a large number of predictors and small sample sizes.

We also evaluated if the inclusion of generic *E. coli* levels as an input variable would substantially improve the performance of the conditional forest models. To do so, separate conditional forest models were re-un (one per pathogen) as described above but excluding generic *E. coli* level as a feature. Performance measures were used to compare the models that included and excluded *E. coli* levels.

All models presented here, were developed as a proof of concept. As such, these models should not be used to guide on-farm decision making, and instead should be used as a starting point for the development of field-ready models (i.e., that can be used by stakeholders to predict pathogen presence in agricultural water) as part of future studies (e.g., using larger datasets, validated using an independent test dataset).

RESULTS

General water quality. In total, 169 samples were collected between January 30th and November 19th, 2018 from 60 canals; each canal is referred to as a "site." The sites were within an approx. 28,000 km^2 area representing 9 irrigation districts. On average, the majority of land in the 10,000 ft surrounding the sites was classified as cropland or natural cover (Table S5.3). The mean generic *E. coli* level in the water samples was 1.4 $\log_{10}$ MPN/ 100 mL (standard deviation= 0.7 $\log_{10}$ MPN/ 100 mL) and

the mean turbidity in the water samples was 32.7 NTU (standard deviation= 92.7 NTU); similar statistics for other variables can be found in Tables S5.2-S5.4.

Pathogen testing results. The overall prevalence of *Salmonella* was 36% (60/169). Of the 60 sites, 20 were *Salmonella* negative on all samplings, however, 2 of those sites were only sampled once. Conversely, 9 sites were positive on all samplings; (1, 5, and 3 of these sites were sampled 1, 2, and 3 times, respectively). The overall prevalence of EHEC markers (i.e., both *stx1*/*2* and *eaeA*) in the water samples was 21% (36/169). Thirty-three sites were negative for EHEC markers on all samplings, however, 3 of those sites were only sampled once. Conversely, 2 sites were positive on all samplings for EHEC markers (both sites were sampled 3 times).

Regression and conditional random forest modeling of *Salmonella* contamination. Multivariable regression was used to determine which variables were associated with *Salmonella* presence. The percent of developed open space (e.g., large-lot single family homes, golf courses, parks) in the 1,000-5,000 ft buffer area and precipitation 1-2d before sample collection were the only variables retained in the final model. The percent of developed open space in the 1,000-5,000 ft buffer area was negatively associated with the log-odds of a sample being *Salmonella*-positive (P=0.036, Table 5.2). Rain 1-2 d before sample collection was positively associated with the log-odds of a *Salmonella* positive sample (P=0.015, Table 5.2). Results of univariable regression and PCA are detailed in Table S5.5.

Table 5.2. Results of mixed effects regression models[a] that characterize the relationship between pathogen detection (*Salmonella* presence and EHEC marker detection) and environmental factors (e.g., land use, weather, sampling site characteristics, and water quality factors).

Outcome	Factor[b]	Log Odds	95% CI[c]
Salmonella presence	Intercept	-0.9	(-3.2, 0.8)
	% Developed (open) Cover, 1,000-5,000ft Buffer	-0.5	(-1.0, -0.1)
	Precipitation, 1-2d (Yes)[d]	3.5	(1.0, 7.0)
EHEC marker	Intercept	0.3	(-2.3, 1.1)
	E. coli level (log_{10} MPN/ 100 mL)	1.7	(1.0, 2.6)
	Point of Discharge Present (Yes)[e]	-1.8	(-4.0, -0.1)
	Road Crossing Present (Yes)[f]	-1.8	(-3.5, -0.3)
	Precipitation, 12-24h (Yes)[d]	3.7	(1.2, 6.7)

[a]The day of year and irrigation district were included in the models as random effects.

[b]For the *Salmonella* presence model, the residual variance and standard deviation for the day of year are 0.7302 and 0.8545, respectively, and the residual variance and standard deviation for the irrigation district are 2.9023 and 1.7036, respectively. For the EHEC maker model, the residual variance and standard deviation for the day of year are 0.8895 and 0.9431, respectively, and the residual variance and standard deviation for the irrigation district are 0.1269 and 0.3562, respectively.

[c]95% CI= 95% confidence interval

[d]Indicates if there was precipitation in the time frame specified before sample collection. Baseline is no precipitation.

[e]Indicates if a point of discharge (i.e., ground water well discharge into the canals) is present adjacent to the sampling site. Baseline is no point of discharge

[f]Indicates if a road crossing is present adjacent to the sampling site. Baseline is no road crossing.

In addition to multivariable regression, we performed conditional forest analysis to predict *Salmonella* presence. The 10 top-ranked predictors included (i) three predictors related to natural cover, (ii) three weather-related predictors, (iii) two water quality related predictors, (iv) one temporal predictor, and (v) one sampling site predictor (Figure 5.1). Only "precipitation 1-2d before sampling," was also retained in the final multivariable regression model and included in the 10 top-ranked predictors by the forest (Figure 5.1). While not retained in the multivariable regression analysis, several of the 10 top-ranked predictors in the conditional forest were significant according to univariable regression (Table S5.5). Given that conditional forest is better able to handle complex (e.g., interactions between features) and messy (e.g., missing data) data than regression, these differences are not unexpected (see Weller et al. [*75*] for more information). Partial dependence plots were fit to visualize the relationship between the 4 top-ranked predictors in the conditional forest model (Figure 5.2), which were, in order, turbidity, day of year, generic *E. coli* level, and percent natural cover in the 500-1,000 ft buffer area.

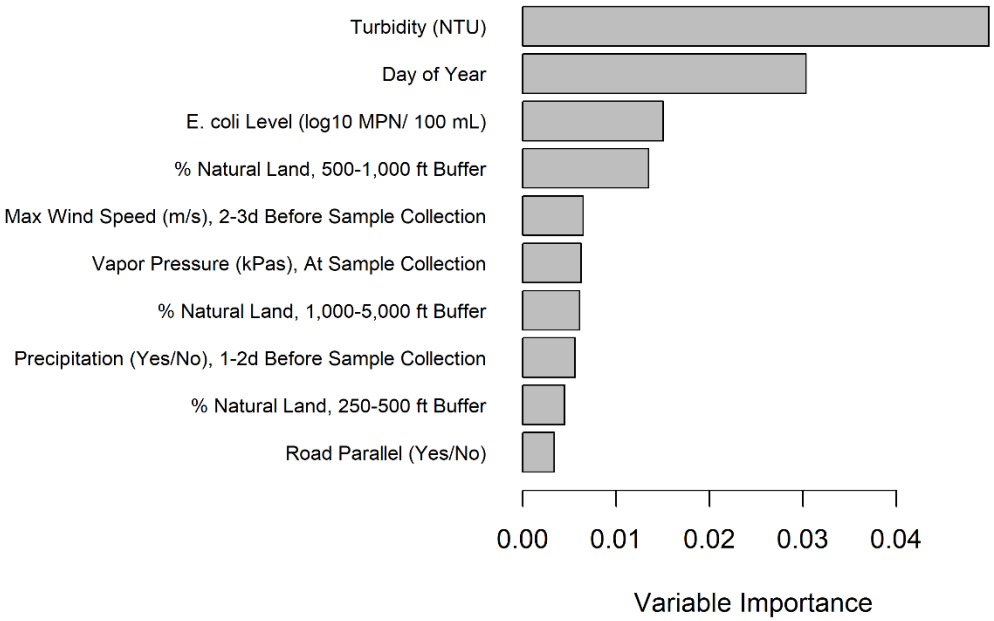


Figure 5.1. Variable importance values for the *Salmonella* conditional forest model with generic *E. coli* included as a predictor. Only the top 10 predictors are included and are listed from most to least important.

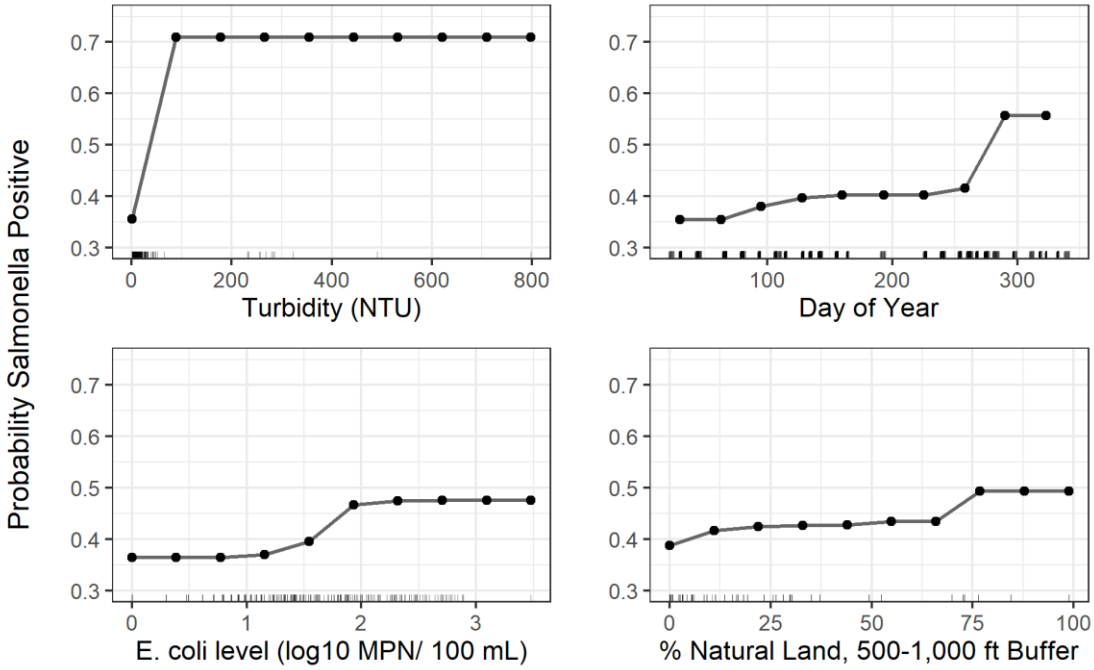


Figure 5.2. Partial dependence plots for the 4 top-ranked predictors according to variable importance in the *Salmonella* conditional forest model with generic *E. coli* included as a predictor; the plots indicate how the predicted probability of a water sample being positive for *Salmonella* presence changes as the x-axis variable (predictor) changes. The tick marks along the x-axis indicate values of the predictor variable in samples used to fit the conditional forest model.

The AUC (area under the curve) and kappa score for the *Salmonella* conditional forest model were 0.84 and 0.51, respectively (Table 5.3). When the probability threshold was set to 0.5 (i.e., to label a sample as positive, the predicted probability of that sample being positive for *Salmonella* must be 0.5 or greater), the sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio were 0.69, 0.82, 3.83, and 0.38, respectively. The sensitivity of 0.69 indicates there is a 0.31 false negative rate, or 31% of the time the model will predict a sample as being negative for *Salmonella* when it is truly positive. Since the model predicts the probability of *Salmonella* being present in a sample (i.e., a continuous outcome), a probability threshold is needed to dichotomize the predicted pathogen status as positive or negative (Table 5.2). If a binary outcome (as opposed to the continuous outcome generated by the forest algorithm) is needed when applying a predictive model, sensitivity and specificity can be adjusted by changing the threshold value (Table 5.4). For instance, if the threshold value was set at 0.4, the sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio would be 0.78, 0.71, 2.70, and 0.31, respectively (Table 5.4).

Table 5.3. Performance measures for the conditional random forest models displaying the relationship of the pathogen detection outcomes (*Salmonella* and EHEC marker presence) with the land use, weather, sampling site characteristics, and water quality predictors with generic *E. coli* level included and excluded as a predictor.

Outcome	*E. coli*[a]	AUC[b]	Kappa[c]	Sensitivity[d]	Specificity[d]	LR+[d,e]	LR-[d,f]	DOR[g] (95% CI[h])
Salmonella	Yes	0.84	0.51	0.69	0.82	3.83	0.38	10.06 (7.38, 13.70)
	No	0.83	0.49	0.69	0.80	3.44	0.39	8.82 (6.52, 11.94)
EHEC marker	Yes	0.92	0.66	0.78	0.88	6.27	0.25	25.19 (18.27, 34.72)
	No	0.91	0.63	0.78	0.85	5.25	0.26	20.54 (15.07, 27.98)

[a]Indicates if *E. coli* level (log_{10}MPN/100 mL water) was included (Yes) or not (No) as a variable in the forest.

[b]AUC=area under the receiver operating characteristic curve.

[c]Kappa score (a measure of agreement between the observed outcome and the predicted outcome; a value of 1 is indicative of perfect agreement and a value of 0 is indicative of an agreement no greater than that of chance).

[d]Measure is biased by the decision threshold used.

[e]LR+= positive likelihood ratio (the likelihood of a predicted pathogen presence when a pathogen is present compared to the likelihood of a predicted pathogen presence when a pathogen is absent.

[f]LR-= negative likelihood ratio (the likelihood of a predicted pathogen absence when a pathogen is present compared to the likelihood of a predicted pathogen absence when a pathogen is absent.

[g]DOR=diagnostic odds ratio (the ratio of the odds of a predicted pathogen presence if the pathogen is present to the odds of a predicted

pathogen presence if the pathogen is absent).

[h]95% CI= 95% confidence interval

Table 5.4. Differences in performance measures for the conditional forest models displaying the relationship of the pathogen detection outcomes (*Salmonella* and EHEC marker presence) with the land use, weather, sampling site characteristics, and water quality predictors with generic *E. coli* level included and excluded as a predictor[a].

Outcome	*E. coli*[b]	Threshold	Sensitivity	Specificity	LR+[c]	LR-[d]
Salmonella presence	Yes	0.5	0.69	0.82	3.83	0.38
		0.4	0.78	0.71	2.70	0.31
		0.3	0.86	0.58	2.07	0.23
	No	0.5	0.69	0.80	3.44	0.39
		0.4	0.78	0.72	2.82	0.30
		0.3	0.87	0.61	2.23	0.21
stx/ *eaeA* co-detection	Yes	0.5	0.78	0.88	6.27	0.25
		0.4	0.83	0.82	4.72	0.20
		0.3	0.93	0.73	3.40	0.10
	No	0.5	0.78	0.85	5.25	0.26
		0.4	0.88	0.75	3.50	0.16
		0.3	0.95	0.65	2.67	0.08

[a]Threshold value indicates the predicted probability a sample must be greater than to be labeled as a positive sample.

[b]Indicates if *E. coli* level (log_{10}MPN/100 mL water) was included (Yes) or excluded (No) in the model as a possible predictor.

[c]LR+= positive likelihood ratio (the likelihood of a predicted pathogen presence when a pathogen is present compared to the likelihood of a predicted pathogen presence when a pathogen is absent.

[d]LR-= negative likelihood ratio (the likelihood of a predicted pathogen absence when a pathogen is present compared to the likelihood

of a predicted pathogen absence when a pathogen is absent.

Regression and conditional random forest modeling of EHEC marker presence. Generic *E. coli* level, precipitation 12-24h before sample collection, if there was a point of discharge (i.e., there was ground water well discharge into the canal) visible from the site, and if there was a road crossing visible from the site were retained in the final EHEC regression model (Table 5.2). Generic *E. coli* level ($P<0.001$) and precipitation 12-24h before sample collection ($P=0.007$) were positively associated with the log-odds of EHEC marker detection ($P<0.001$). A point of discharge and a road crossing visible from the site were negatively associated with the log-odds of EHEC marker detection.

A conditional forest model was also fit to predict EHEC marker detection. The top 10 ranked predictors included (i) six land cover predictors (five related to natural cover), (ii) one weather predictor, (iii) two water quality predictors, and (iv) one temporal predictor (Figure 5.3). While generic *E. coli* level and precipitation 12-24h before sample collection were included in the 10 top-ranked predictors in the forest model, the presence of a point of discharge and road crossing adjacent to the sampling sites were not among the 10 top-ranked predictors in the forest (Figure 5.3). For the EHEC forest, partial dependence plots were fit for the 4 predictors, which were (i) generic *E. coli* level, (ii) day of year, (iii) percent of natural cover in the 250-500 ft buffer area, and (iv) percent of natural cover in the 500-1,000 ft buffer area (Figure 5.4).

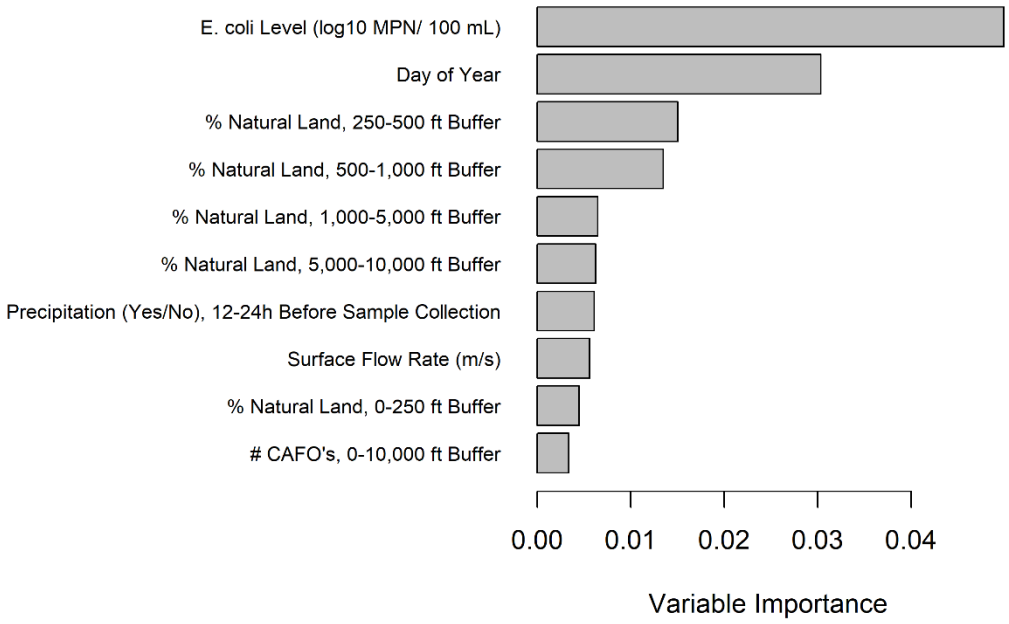


Figure 5.3. Variable importance values for the EHEC marker conditional forest model with generic *E. coli* included as a predictor. Only the top 10 predictors are included and are listed from most to least important.

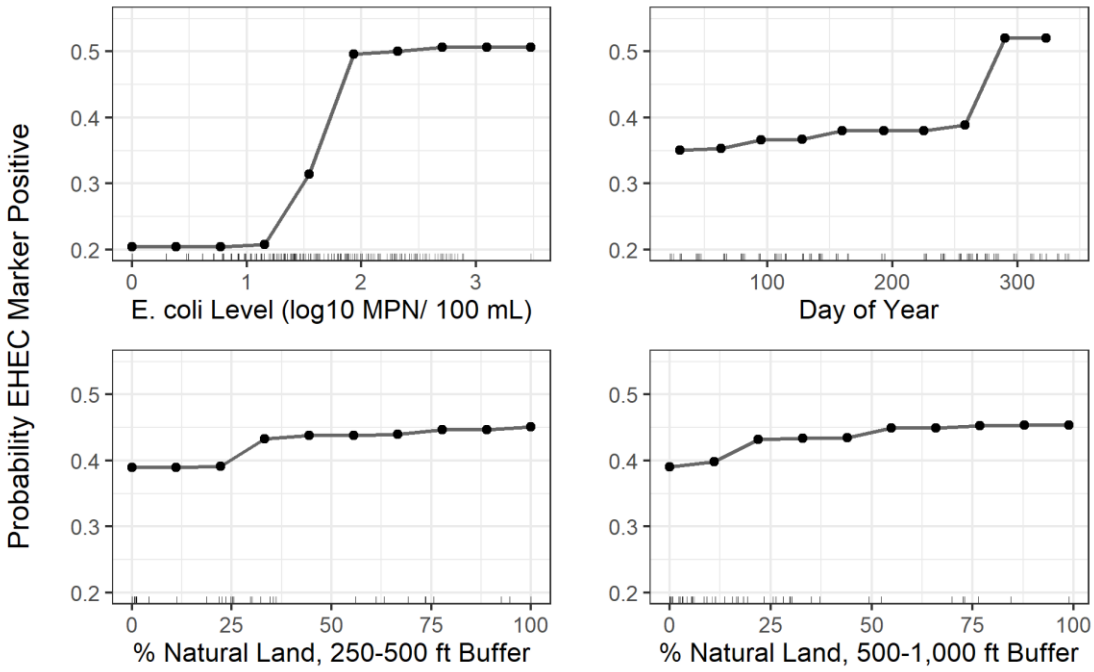


Figure 5.4. Partial dependence plots for the 4 top-ranked predictors according to variable importance in the EHEC marker conditional forest model with generic *E. coli* included as a predictor; the plots indicate how the predicted probability of a water sample being positive for EHEC marker detection changes as the x-axis variable (predictor) changes. The tick marks along the x-axis indicate values of the predictor variable in samples used to fit the conditional forest model.

The AUC and kappa score for the EHEC forest were 0.92 and 0.66, respectively (Table 5.3). The sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio were 0.78, 0.88, 6.27, and 0.25, respectively for the EHEC marker model at a threshold value of 0.5 (Table 5.4). As with the *Salmonella* forest, changing the threshold value could improve performance measures that rely on dichotomizing the predicted probability of EHEC marker detection (e.g., sensitivity; Table 5.4).

Performance of models that do not include generic *E. coli* levels as a predictor. To determine if including generic *E. coli* levels in the conditional forest

models substantially improved predictive performance, we re-ran the forest models without generic *E. coli* level as a predictor (see Figures S5.2-S5.5 for variable importance and partial dependence plots). There were no substantial differences in performance between the models with and without generic *E. coli* included as a predictor (Table 5.3). For instance, the AUC values were 0.84 and 0.83 for the *Salmonella* forests that included and excluded generic *E. coli* levels, respectively. Similarly, the AUC values were 0.92 and 0.91 for the EHEC forests that included and excluded generic *E. coli* levels, respectively.

DISCUSSION

The current study assessed *Salmonella* presence and EHEC marker detection in southwestern US canals used for agricultural water. Regression was used to identify associations between environmental conditions and *Salmonella* presence and EHEC marker detection. The data were also utilized to determine if forest-based models were a feasible approach for predicting *Salmonella* presence and EHEC marker detection in canals. While these models were developed as a proof of concept, they provide a conceptual framework on which future work (development of models that can be integrated into on-farm decision-making) can build. Our results can also be used to identify factors important for predicting pathogen presence in southwestern US canal water to guide future data collection to be used to provide maximum value for the refinement of predictive models that can be deployed for industry use.

***Salmonella* and EHEC marker prevalence.** *Salmonella* has been isolated from flowing surface water sources in both this and previous studies *(7, 10, 15, 17, 19, 24, 50, 57, 58, 59, 64, 65, 73, 74, 76, 77)*. While *Salmonella* prevalence varied widely between these studies, the *Salmonella* prevalence reported here (36%) falls within the range reported by these previous studies, which was between 3% (6/223) (British

Columbia, Canada; *19*) and 76% (80/105) (Georgia, USA; *10*).

While several studies have attempted to assess the prevalence of EHEC or different EHEC subgroups (e.g., *E. coli* O157) in surface water, the specific methodologies used can have a considerable impact on prevalence estimates. Some studies reported the percent of culture-confirmed EHEC or STEC positive samples *(7, 15, 19, 27, 45, 50, 58, 63, 64, 65, 76, 77)*, others solely relied on PCR screens for either O157 markers *(57)* or EHEC markers *(10, 56, 73, 74)*, such as the study reported here. Regardless, most of these previous studies that assessed EHEC in running surface waters have reported lower prevalence than found here (21%). For example, the EHEC prevalence ranged from <1% (5/818) (Ontario, Canada; *76*) to 19% (63/330) (British Columbia, Canada; *45*); both studies utilized culture confirmed EHEC results. While the higher EHEC prevalence in the current study could be due to a truly higher prevalence, the use of a PCR screen for *stx* and *eaeA* in a single sample without culture confirmation may overestimate the prevalence of EHEC, as this method may (i) detect *stx* and *eaeA* in different organisms and (ii) may detect genetic material from dead organisms. The lower prevalence of EHEC in previous studies that used culture confirmation is also likely explained by the lack of reliable EHEC culture-conformation methods which can underestimate its true prevalence *(4, 8, 44)*.

The complexity of relationships observed was dependent on the analytical approach. Two modeling approaches were utilized, including (i) multivariable regression to identify associations between environmental variables and *Salmonella* presence and EHEC marker detection, and (ii) conditional forest to develop models to predict *Salmonella* presence and EHEC marker detection. For both *Salmonella* presence and EHEC marker detection, there were several differences in the variables retained in the regression models and the variables ranked as important by the forests, even though there are some overlaps between important variables between the modeling strategies.

This highlights the fact that different modeling strategies are able to detect different relationships in the data *(35)*. For instance, regression relies on the assumption that there is a linear relationship between independent variables and the log-odds of the dependent variable being detected, and therefore, non-linear relationships cannot be detected. In addition, logistic regression cannot handle a large number of independent variables simultaneously (requiring variable selection before model development), missing data, or correlated variables, and can only account for a limited number of interactions. In comparison, conditional forest utilizes tree-based modeling which does not require the same linear assumption to be met, implicitly accounts for hierarchical relationships and interactions in the data, and can handle missing data, large numbers of independent variables, and correlation. For example, some variables were important by univariable regression but could not be included in final multivariable regression models because they loaded on more than 1 principal component (e.g., percent of natural cover around the sampling site for the *Salmonella* model). PCA is one strategy used for variable selection in regression analysis, as regression analysis is unable to handle overly complex models (see Kuhn and Johnson [*35*] for additional variable reduction strategies). On the other hand, variable selection is incorporated into the conditional forest algorithm and as such is better able to capture the complex relationships inherent to environmental data *(74)*. However, regression-based analysis does have its advantages, especially for hypothesis testing of the relationships between specific, independent variables and the outcome. Regression models are more interpretable than forests, making it easier to understand the relationships in a regression as opposed to forest models *(35)*.

Only precipitation 1-2 d before sampling was included among the 10 top-ranked predictors in the *Salmonella* forest and retained in the final regression model. Several other studies also found an increased likelihood of *Salmonella* detection following rain

events *(24, 37, 57, 73, 76)*. For instance, in a survey of surface water in Georgia, USA, Haley et al. *(24)* found significant (P<0.005) positive correlations between *Salmonella* levels, and rainfall 1 and 2 days before sample collection. This relationship may be driven by increases in run-off during rain events, which can transfer *Salmonella* from terrestrial sources to waterways. Unlike rain, developed open space was included in the final *Salmonella* regression model but was not highly ranked in the *Salmonella* forest. According to the regression model, as the percent of developed open space increased, the log-odds of detecting *Salmonella* decreased. Developed open space may act as a proxy for built-landscape features that prevent run-off and microbial contaminants from entering canals, such as vegetative buffers (e.g., in parks) or improved drainage systems. Consistent with these findings, a survey of Central California waterways found a significantly lower prevalence (P<0.05) of *Salmonella* in human-impact areas (47%) compared to animal-impacted areas (74%) *(64)*. However, several studies have found a positive association between *Salmonella* presence and variables linked to human presence human presence *(33, 73)*. Johnson et al. (33) speculated this inconsistency between studies could be due to the quality of wastewater removal infrastructure in the sampling area.

For the EHEC models, generic *E. coli* levels and precipitation before sampling were included in both the regression and forest models. The relationship between precipitation and an increased log-odds of EHEC detection was likely also driven by an increase in run-off during rain events, similar to the relationship between *Salmonella* and precipitation discussed above. The relationship between EHEC detection and precipitation is also consistent with past studies *(45, 57)*. Conversely, there is considerable variability between previous studies in the existence, direction, and strength of the relationship between EHEC detection and generic *E. coli* levels. For

example, some studies, like the study presented here, found evidence of a relationship *(10, 19, 28, 57, 65, 74)*, while others did not *(7, 19, 50, 56, 73)*.

While there were overall differences in the variables identified as being associated with *Salmonella* presence or EHEC marker detection by regression analysis and those identified as a top ranked predictor by conditional forest analysis, both modeling strategies used together can provide a more complete understanding of the processes that drive pathogen presence. For instance, the variables associated with pathogen presence via regression provide easy to interpret information on associations between a subset of factors and likelihood of pathogen contamination. On the other hand, the top ranked variables in the *Salmonella* presence or EHEC marker detection conditional forests may provide insight into what variables are important for inclusion in models that predict pathogen presence in agricultural water. This can be used to determine what additional information should be collected to improve the performance of these predictive models so they can be implemented by industry. However, the complex interactions between variables included in the conditional forest models can make it difficult to assess how a change in one variable alone (e.g., occurrence of rainfall) will impact the outcome (i.e., pathogen presence).

Machine-learning-based models have potential for prediction of pathogen contamination likelihood, including real time prediction that does not require microbiological data. Generic *E. coli* is traditionally used as a fecal indicator in agricultural water to indicate potentially unhygienic conditions. However, the high cost, slow turnaround time, uneven distribution of generic *E. coli* in surface waters, and inconsistent relationships between pathogen presence and generic *E. coli* level limit its value and feasibility of routine use *(48, 72)*. Our data here provide further support that conditional forest models are able to predict the presence of *Salmonella* and EHEC markers, as supported by AUC values of 0.84 for the *Salmonella* model and 0.92 for the

EHEC marker model (AUC values of 0.8 to 0.9 are generally indicative of excellent predictive performance; *42*). Previous studies by Polat et al. *(51)* and Weller et al. *(75)* also previously reported that machine learning models show potential as a strategy for identifying contaminated agricultural water in Florida ponds and New York streams, respectively. However, these AUC values from the current study were calculated using cross-validation, as opposed to an independent test dataset, and may be overfit. Regardless, overfitting concerns are mitigated by the fact that the models developed here were developed as a proof of concept and conceptual framework, and not intended for actual use on-farms. If predictive models are going to be developed as an alternative or supplement to indicator-based monitoring, sufficient data is needed, ideally spanning several years and regions, to allow for separate, independent training and test data. Furthermore, additional information is needed to determine if predictive models should be developed for individual waterways, specific regions, or if a standard model can be used across multiple regions.

Importantly, the removal of generic *E. coli* level as a predictor in the *Salmonella* presence and EHEC marker detection conditional forests did not substantially decrease predictive performance of either model. This provides evidence that it is possible to eliminate the use of generic *E. coli* water testing (or other microbial water testing strategies) and replace it with real-time predictive models with limited impact on the accuracy of identifying when water may be contaminated with pathogens and thus at an increased potential risk. These real-time models would be advantageous, as produce growers could estimate the likelihood of pathogen presence in their water sources at the time of water application.

Presenting predictive modeling outcomes as continuous risk measures and dichotomized outcomes have distinct advantages. The output of the conditional forest models is the predicted probability that a sample will be positive for *Salmonella*

presence or EHEC marker detection. One method for using this predicted probability for making decisions on how to utilize the water would be to dichotomize the outcome (i.e., pathogen is present or absent) based on if the predicted probability is greater than or less than some set threshold value. If this strategy is used, specificity, sensitivity positive likelihood ratio, and negative likelihood ratio can be calculated. While the specificity is adequate for our *Salmonella* model, the sensitivity is low (0.69). This is particularly concerning as this means the model often calls *Salmonella*-positive samples as negative, and thus may lead to instances where corrective actions (e.g., water treatment) were not performed when they should have been. However, by lowering the threshold value, the sensitivity of the model can be increased, minimizing this risk. A similar phenomenon was observed for the EHEC forest. If predictive models are to be used by produce growers to guide on-farm decision making (e.g., if corrective actions are needed before using water to irrigate crops), optimization of this threshold value is needed. A future quantitative microbial risk assessment would be helpful to identify the risk of illness associated with different threshold values *(53, 71)*. This information could then be used to optimize what threshold value should be used to balance the predicted number of illnesses versus the costs associated with different corrective actions (e.g., water treatment). While dichotomizing the outcome of the model, as described above, creates an easier to interpret model, it does cause a loss of information. As an alternative, the predicted probability of a sample being positive could instead be directly used for decision making; however, this would also require quantitative risk assessment to determine how predicted probabilities should be used.

Natural cover and the day of year are important for prediction of *Salmonella* presence and EHEC marker detection. As previously discussed, there were some differences in variables associated with pathogen presence by multivariable regression and the top ranked variables for predicting pathogen presence by conditional

forest. However, a specific discussion of the top ranked variables is important for informing what information should be collected for future refinement of models used to predict pathogen presence in agricultural water. Natural cover variables and the day of year were included in the ten top-ranked predictors in the *Salmonella* and EHEC forests. For both *Salmonella* and EHEC, there was a positive monotonic relationship between the percent of natural cover around the sampling site and pathogen presence. Since natural cover may function as habitat for wildlife, this may indicate wildlife is acting as a pathogen source in southwestern US canals; this is supported by the limited number of past studies that examined the prevalence of enteric pathogens in southwestern wildlife (e.g., *31, 32*). For example, an Arizona study found 32% (N=103 total samples) of coyote fecal samples were positive for *Salmonella*, while none were STEC-positive and 4.9% were enteropathogenic *E. coli*-positive *(32).* On the other hand, in a study investigating *E. coli* O157:H7 in feral swine in the central California coast, 14.9% (13/87) of samples were positive for *E. coli* O157:H7 *(31)*. Since, in the current study, the relationship between natural cover and *Salmonella* presence and EHEC marker detection is weak, additional research is needed to fully characterize the role wildlife plays as a source of enteric pathogen contamination for southwestern canals. Given the need for additional research, and the important ecosystem services provided by natural cover (e.g., water filtration) and wildlife (e.g., pest control, pollination), the authors want to emphasize we are not advocating the removal of natural cover or wildlife from growing areas *(1, 3, 46).*

Temporal trends in *Salmonella* presence and EHEC marker detection were accounted for by including the day of year each sample was collected on as a predictor in the forests. The day of year was used instead of season, as seasons are arbitrary periods of time; the end of a season is more similar to the beginning of the subsequent season than the beginning of the season itself. As such, using the day of year as a

continuous variable reduces bias in the final model by not forcing the data into arbitrary categories. For both models, the probability of a sample being pathogen-positive remained low until approx. September, after which the probability of a positive increased. This likely indicates some event occurs during early fall that leads to an increased likelihood of pathogen contamination of southwestern US canals. For instance, it is possible the canals are cleaned at this time of year, which causes the sediments at the bottom to re-distribute and re-contaminate the water. However, the current study only spanned one year so additional research is needed to determine if this relationship holds across time. Despite this limitation, our finding of intra-annual trends in microbial water quality is consistent with past studies that looked at *Salmonella (15, 37, 57, 64, 74, 76)* and EHEC *(45, 56, 57, 64)*. For instance, Cooley et al. *(15)* found a higher *Salmonella* prevalence in the spring and summer compared to the fall and winter in Central California surface water samples. Furthermore, consistent with our current study, Stea et al. *(57)* found a higher prevalence of STEC in the later summer and fall compared to all other seasons in Nova Scotia, Canada. Overall, the data collected to date appear to indicate that enteric pathogen contamination of surface water often shows some type of seasonality, although the specific trends appear to differ across locations and studies.

Limitations. While a large area (approx. 28,000 km^2) of the produce growing region in the southwestern US is represented here, stratification for certain land-use or sample site factors was not performed during sample site selection. As such, this could have biased the results (i.e., some potentially important factors could have been missed due to underrepresentation of certain variables). Additionally, since this is a proof of concept, only a small number of samples were collected (N=169) and each site was only visited few times (1-3 times), which could result in combinations of factors associated with an altered likelihood of pathogen presence being missed (e.g., if the greatest

likelihood of *Salmonella* contamination is after a rain event next to a dairy farm but no samples were collected after a rainfall from a site next to a dairy farm, this signal would have been missed). Therefore, further studies, with larger sampling efforts and spanning multiple years and geographical locations or growing regions, are needed to yield models appropriate for industry use and to answer key questions such as if a single model can be used from water source to water source, if a single model can be used from climate to climate (or region to region), and if a single model can be used over several years. In addition, there are other factors that may be important for pathogen presence in canals such as difference in elevation between surrounding land and the canal, livestock density surrounding the canals (instead of just CAFO presence), land use along canal flow paths, and relative humidity; future model building efforts should consider collecting these data to include in their models. Furthermore, there were several factors with missing data (e.g., flow rate could not be measured at all sites due to safety concerns). This could lead to information bias (i.e., bias caused by a lack of correct or complete information) in logistic regression, but we expect this to be non-differential (i.e., the bias direction is independent of the model outcome), indicating it should not have impacted the results of the study.

CONCLUSIONS

Machine learning-based predictive models, such as conditional forest models show promise for predicting *Salmonella* presence and EHEC marker detection in southwestern US canals used as sources for agricultural water. The use of machine learning models, in addition to regression analysis, provides a more complete assessment of the relationships between spatial and temporal factors and foodborne pathogen presence in agricultural water due to the complexity in the system. Furthermore, the use of predictive modeling, and real-time predictive models (using no

microbiological data), may provide an alternative or supplement to traditional generic *E. coli* testing for fine-tuning when and where food safety hazards may be present in agricultural water and corrective action is needed. The forests developed in the current study specifically indicate that use around the sampling site and day of the year are important predictors for both *Salmonella* presence and EHEC marker detection in southwestern US canal water. Despite the promising results in this and previous studies, these studies were proof of concept. Therefore, before predictive models can be deployed on farms and integrated into on-farm risk management plans additional research is needed to determine if models can predict pathogen presence accurately for regions, water types (e.g., canal, stream, pond), and years, other than the region(s), water type(s), and year(s) where the training data were collected.

FUNDING

This study was funded by the Center for Produce Safety (2017CPS09). Data analysis, manuscript preparation, and revision were also partially supported by the National Institute of Environmental Health Sciences, National Institutes of Health (NIH) under award number T32ES007271.

ACKNOWLEDGEMENTS

We would like to thank Maureen Gunderson, Kayla Ferris, Nathanael Henderson, Julia Muuse, and Sriya Sunil for their help on various aspects of this project. We would also like to acknowledge the various irrigation districts within the Southwest growing region for allowing our team to collect agricultural water samples and providing guidance on sampling locations appropriate for the study.

REFERENCES

1. Aarons, S. R., and Gourley, C. J. P. (2012). The role of riparian buffer management in reducing off-site impacts from grazed dairy systems. *Renew. Agr. Food Syst*. 28(1):1-16. doi: 10.1017/S1742170511000548

2. Allende, A., and Monaghan, J. (2015). Irrigation water quality for leafy crops: A perspective of risks and potential solutions. *Int. J. Environ. Res. Public. Health*. 12:7457-7477. doi: 10.3390/ijerph120707457

3. Allende, A., Datta, A. R., Smith, W. A., Adonis, R., MacKay, A., Adell, A. D. (2018). Implications of new legislation (US FSMA) and guidelines (EC) on the establishment of management systems for agricultural water. *Food Microbiol*. 75:119-125. doi: 10.1016/j.fm.2017.10.002

4. Baker, C. A., De, J., Bertoldi, B., Dunn, L., Chapin, T., Jay-Russell, M., Danyluk, M. D., Schneider, K. R. (2019). Prevalence and concentration of *stx+ E. coli* and *E. coli* O157 in bovine manure from Florida farms. *PLoS ONE*. 14(5): e0217445. doi: 10.1371/journal.pone.0217445

5. Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Stat*. 67(1):1-48. doi: 10.18637/jss.v067.i01.

6. Beauvais, W., Gart, E. V., Bean, M., Blanco, A., Wilsey, J., McWhinney, K., et al. (2018). The prevalence of *Escherichia coli* O157:H7 fecal shedding in feedlot pens is affected by the water-to-cattle ratio: A randomized controlled trial. *PLoS ONE*. 13(2): e0192149. doi: 10.1371/journal.pone.0192149

7. Benjamin, L., Atwill, E. R., Jay-Russell, M., Cooley, M., Carychao, D., Gorski, L., et al. (2013). Occurrence of generic *Escherichia coli*, *E. coli* O157 and *Salmonella* spp. in water and sediment from leafy green produce farms and streams of the Central California Coast. Int J Food Microbiol. 165:65-76. doi: 10.1016/j.ijfoodmicro.2013.04.003

8. Bettelheim, K. A. 2007. The Non-O157 Shiga-Toxigenic (Verotoxigenic) *Escherichia coli*; under-rated pathogens. *Crit. Rev. Microbiol*. 33:67-87. doi: 10.1080/10408410601172172

9. Bischl B., Lang M., Kotthoff L., Schiffner J., Richter J., Studerus E., et al. (2016). mlr: Machine Learning in R. *J. Mach. Learn. Res*. 17(170):1-5. http://jmlr.org/papers/v17/15-066.html.

10. Bradshaw, J. K., Snyder, B. J., Oladeinde, A., Spidle, D., Berrang, M. E., Meinersmann, R. J., et al. (2016). Characterizing relationships among fecal indicator bacteria, microbial source tracking markers, and associated waterborne pathogen occurrence in stream water and sediments in a mixed land use watershed. *Water Res*. 101:498-509. doi: 10.1016/j.watres.2016.05.014

11. California Leafy Greens Marketing Agreement (2020). Commodity specific food safety guidelines for production and harvest of lettuce and leafy greens. https://lgma-assets.sfo2.digitaloceanspaces.com/downloads/CA_LGMA_METRICS_FINAL_VERSION_Accessible_Jan2020.pdf [Accessed on July 2, 2020].

12. Callejon, R. M., Rodriguez-Naranjo, M. I., Ubeda, C., Hornedo-Ortega, R., Garcia-Parrilla, M. C., Troncoso, A. M. (2015). Reported foodborne outbreaks due to fresh produce in the United States and European Union: Trends and causes. *Foodborne Pathog. Dis*. 12(1):32-38. doi: 10.1089/fpd.2014.1821

13. Castro-Ibanez, I., Gil, M. I., Tudela, J. A., Ivanek, R., Allende, A. (2015). Assessment of microbial risk factors and impact of meteorological conditions during production of baby spinach in the Southeast of Spain. *Food Microbiol*. 49:173-181. doi: 10.1016/j.fm.2015.02.004

14. 21 C.F.R. § 112.44. 2019.

15. Cooley, M. B., Quinones, B., Oryang, D., Mandrell, R. E., Gorski, L. (2014). Prevalence of shiga toxin producing *Escherichia coli*, *Salmonella enterica*, and *Listeria monocytogenes* at public access watershed sites in a California Central Coast agricultural region. *Front. Cell. Infect. Microbiol*. 4(3):1-13. doi: 10.3389/fcimb.2014.00030

16. Delaquis, P., Bach, S., Dinu, L. D. (2007). Behavior of *Escherichia coli* O157:H7 in leafy vegetables. *J. Food Prot*. 70(8):1966-1974. doi: 10.4315/0362-028x-70.8.1966

17. Duffy, E. A., Lucia, L. M., Kells, J. M., Castillo, A., Pillai, S. D., Acuff, G. R. (2005). Concentrations of *Escherichia coli* and genetic diversity and antibiotic resistance profiling of *Salmonella* isolated from irrigation water, packing shed equipment, and fresh produce in Texas. *J. Food Prot.* 68(1):70-79. doi: 10.4315/0362-028x-68.1.70

18. European Commission (2017). Commission notice on guidance document on addressing microbiological risks in fresh fruits and vegetables at primary production through good hygiene. *Official Journal of the European Union*.

19. Falardeau J., Johnson, R. P., Pagotto, F., Wang, S. (2017). Occurrence, characterization, and potential predictors of verotoxigenic *Escherichia coli*, *Listeria monocytogenes*, and *Salmonella* in surface water used for produce irrigation in the Lower Mainland of British Columbia, Canada. *PLoS ONE*. 12(9): e0185437. doi: 10.1371/journal.pone.0185437

20. Franz, E., Schijven, J., de Roda Husman, A. M., Blaak, H. (2014). Meta-regression analysis of commensal and pathogenic *Escherichia coli* survival in soil and water. *Environ. Sci. Technol*. 48:6763-6771. doi: 10.1021/es501677c

21. Freshcare (2020). Food safety and quality standard edition 4.2. https://www.freshcare.com.au/wp-content/uploads/Freshcare-Food-Safety-Quality-Standard-Edition-4.2-JAN-2021.pdf [Accessed on May 22, 2021].

22. Gore, J. A., and Banning, J. (2017). Discharge measurements and streamflow analysis, p. 49-70. In F. R. Hauer and G. Lamberti (ed.), Methods in Stream Ecology, 3rd ed. Salt Lake City, UT: Academic Press

23. Greene, S. K., Daly, E. R., Talbot, E. A., Demma, L. J., Holzbauer, S., Patel, N. J., et al. (2008). Recurrent multistate outbreak of *Salmonella* Newport associated with tomatoes from contaminated fields, 2005. *Epidemiol. Infect*. 136(2):157-165. doi: 10.1017/S095026880700859X

24. Haley, B. J., Cole, D. J., Lipp, E. K., (2009). Distribution, diversity, and seasonality of waterborne *Salmonellae* in a rural watershed. *Appl. Environ. Microbiol*. 75(5):1248-1255. doi: 10.1128/AEM.01648-08

25. Hanning, I. B., Nutt, J. D., Ricke, S. C. (2009). Salmonellosis outbreaks in the United States due to fresh produce: Sources and potential intervention methods. *Foodborne Pathog. Dis*. 6(6):635-648. doi: 10.1089/fpd.2008.0232

26. Havelaar, A. H., Vazquez, K. M., Topalcengiz, Z., Munoz-Carpena, R., Danyluk, M. D. (2017). Evaluating the U.S. Food Safety Modernization Act produce safety rule standard for muicrobial quality of agricultural water for produce growing. *J. Food Prot*. 80(11):1832-1841. doi: 10.4315/0362-028X.JFP-17-122

27. Haymaker, J., Sharma, M., Parveen, S., Hashem, F., May, E. B., Handy, E. T., et al. (2019). Prevalence of shiga-toxigenic and atypical enteropathogenic *Escherichia coli* in untreated surface water and reclaimed water in the Mid-Atlantic U.S. *Environ. Res*. 172:630-636. doi: 10.1016/j.envres.2019.02.019

28. Holvoet, K., Sampers, I., Seynnaeve, M., Uyttendaele, M. (2014). Relationship among hygiene indicators and enteric pathogens in irrigation water, soil and lettuce and the impact of climatic conditions on contamination in the lettuce primary production. *Int. J. Food Microbiol*. 171:21-31. doi: 10.1016/j.ijfoodmicro.2013.11.009

29. Hothorn, T., Buehlmann, P., Dudoit, S., Molinaro, A., Van Der Laan, M. (2006). Survival Ensembles. *Biostatistics*, 7(3):355-373. doi: 10.1093/biostatistics/kxj011

30. Jamieson, R., Gordon, R., Joy, D., Lee, H. (2004). Assessing microbial pollution of rural surface waters: a review of current watershed scale modeling approaches. *Agric. Water Manag*. 70:1-17. doi: 10.1016/j.agwat.2004.05.006

31. Jay, M. T., Cooley, M., Carychao, D., Wiscomb, G. W., Sweitzer, R. A., Crawford-Miksza, L., et al. (2007). *Escherichia coli* O157:H7 in feral swine near spinach fields and cattle, central California coast. *Emerg. Inf. Dis*. 13(12):1908-1911. doi: 10.3201/eid1312.070763

32. Jay-Russell, M. T., Hake, A. F., Bengson, Y., Thiptara, A., Nguyen, T., (2014). Prevalence and characterization of *Escherichia coli* and *Salmonella* strains isolated from stray dog and coyote feces in a major leafy greens production

region at the United States- Mexico border. *PLoS ONE*. 9(11):e113433. doi: 10.1371/journal.pone.0113433

33. Johnson, J. Y. M., Thomas, J. E., Graham, T. A., Townshend, I., Byrne, J., Selinger, L. B., et al. (2003). Prevalence of *Escherichia coli* O157:H7 and *Salmonella* spp. in surface waters of southern Alberta and its relation to manure sources. *Can. J. Microbiol*. 49:326-335. doi: 10.1139/w03-046

34. Kim, J. S., Lee, G. G., Park, J. S., Jung, Y. H., Kwak, H. S., Kim, S. B., et al. (2007). A novel multiplex PCR assay for rapid and simultaneous detection of five pathogenic bacteria: *Escherichia coli* O157:H7, *Salmonella*, *Staphylococcus aureus*, *Listeria monocytogenes*, and *Vibrio parahaemolyticus*. *J. Food Prot*. 70(7):1656-1662. doi: 10.4315/0362-028x-70.7.1656

35. Kuhn, M., and Johnson, K. (2013). Applied Predictive Modeling. New York, USA: Springer.

36. Kuhn, M. (2020). caret: Classification and Regression Training. R package version 6.0-86. https://CRAN.R-project.org/package=caret

37. Liang, Z., He, Z., Zhou, X., Powell, C. A., Yang, Y., He, L. M., et al. (2013). Impact of mixed land-use practices on the microbial water quality in a subtropical coastal watershed. *Sci. Total Environ*. 449:426-433. doi: 10.1016/j.scitotenv.2013.01.087

38. Liu, H., Whitehouse, C. A., Li, B. (2018). Presence and Persistence of *Salmonella* in water: the impact on microbial quality of water and food safety. *Front. Public Health*. 6(159):1-13. doi: 10.3389/fpubh.2018.00159

39. Lopez-Galvez, F., Allende, A., Pedrero-Salcedo, F., Alarcon, J. J., Gil, M. I. (2014). Safety assessment of greenhouse hydroponic tomatoes irrigation with reclaimed and surface water. *Int. J. Food Microbiol*. 191:97-102. doi: 10.1016/j.ijfoodmicro.2014.09.004

40. Lothrop N., Bright, K. R., Sexton, J., Pearce-Walker, J., Reynolds, K. A., Verhougstraete, M. P. (2018). Optimal strategies for monitoring irrigation water quality. *Agric. Water Manag*. 199:86-92. doi: 10.1016/j.agwat.2017.12.018

41. Lu, L., Hume, M. E., Sternes, K. L., Pillai, S. D. (2004). Genetic diversity of *Escherichia coli* isolates in irrigation water and associated sediments: implications for source tracking. *Water Res*. 38:3899-3908. doi: 10.1016/j.watres.2004.06.020

42. Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol*. 5(9):1315-1316.

43. Mootian, G., Wu, W. H., Matthews, K R. (2009). Transfer of *Escherichia coli* O157:H7 from soil, water, and manure contaminated with low numbers of the pathogen to lettuce plants. *J. Food Prot*. 72(11):2308-2312. doi: 10.4315/0362-028x-72.11.2308

44. Muniesa, M., Jofre, J., Garcia-Aljaro, C., Blanch, A. R. (2006). Occurrence of *Escherichia coli* O157:H7 and other enterohemorrhagic *Escherichia coli* in the environment. *Environ. Sci. Technol*. 40:7141-7149. doi: 10.1021/es060927k

45. Nadya, S., Delaquis, P., Chen, J., Allen, K., Johnson, R. P., Ziebell, K., et al. (2016). Phenotypic and genotypic characteristics of shiga toxin-producing *Escherichia coli* isolated from surface waters and sediments in a Canadian urban-agricultural landscape. *Front. Cell. Infect. Microbiol*. 6(36):1-13. doi: 10.3389/fcimb.2016.00036

46. Navarro-Gonzalez, N., Wright, S., Aminabadi, P., Gwinn, A., Suslow, T. V., Jay-Russell, M. T. (2020). Carriage and subtypes of foodborne pathogens identified in wild birds residing near agricultural lands in California: a repeated cross-sectional study. *Appl. Environ. Microbiol*. 86(3):e01678-19. doi: 10.1128/AEM.01678-19

47. Pachepsky, Y., Shelton, D. R., McLain, J. E. T., Patel, J., Mandrell, R. E. (2011). Irrigation waters as a source of pathogenic microorganisms in produce: a review. *Adv. Agron*. 113:73-138.

48. Pachepsky, Y., Shelton, D., Dorner, S., Whelan, G. (2016). Can *E. coli* or thermotolerant coliform concentrations predict pathogen presence or prevalence in irrigation waters? *Crit. Rev. Microbiol*. 42(3):384-393. doi: 10.3109/1040841X.2014.954524

49. Park, S., Szonyi, B., Gautam, R., Nightingale, K., Aniciso, J., Ivanek, R. (2012). Risk factors for microbial contamination in fruits and vegetables at the preharvest level: a systematic review. *J. Food Prot*. 75(11):2055-2081. doi: 10.4315/0362-028X.JFP-12-160

50. Partyka, M. L., Bond, R. F., Chase, J. A., Atwill, E. R. (2018). Spatiotemporal variability in microbial quality of western U.S. agricultural water supplies: a multistate study. *J. Environ. Qual*. 47L939-948. doi: 10.2134/jeq2017.12.0501

51. Polat, H., Topalcengiz, Z., Danyluk, M. D. (2020). Prediction of *Salmonella* presence and absence in agricultural surface waters by artificial intelligence approaches. *J. Food Saf*. 40:e12733. doi: 10.1111/jfs.12733

52. R Core Team. (2021). R: A language and environment for statistical computing. Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.

53. Rock, C. M., Brassill, N., Dery, J.L., Carr, D., McLain, J. E., Bright, K. R., Gerba, C. P. (2019). Review of water quality criteria for water reuse and risk-based implications for irrigated produce under the FDA Food Safety Modernization Act, produce safety rule. *Environ. Res*. 172:616-629. doi: 10.1016/j.envres.2018.12.050.

54. 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. doi: 10.1016/j.watres.2007.03.028

55. Sbodio, A., Maeda, S., Lopez-Velasco, G., Suslow, T. V. (2013). Modified Moore swab optimization and validation in capturing *E. coli* O157:H7 and *Salmonella enterica* in large volume field samples of irrigation water. *Food Res. Int.* 51:654-662. doi: 10.1016/j.foodres.2013.01.011

56. Shelton, D. R., Karns, J. S., Coppock, C., Patel, J., Sharma, M., Pachepsky, Y. A. (2011). Relationship between *eae* and *stx* virulence genes and *Escherichia coli* in an agricultural watershed: implications for irrigation water standards and leafy green commodities. *J. Food Prot*. 74(1):18-23. doi: 10.4315/0362-028X.JFP-10-241

57. Stea, E. C., Truelstrup Hansen, L., Jamison, R. C., Yost, C. K. (2015). Fecal contamination in the surface waters of a rural and an urban-source watershed. *J. Environ. Qual*. 44:1556-1567. doi: 10.2134/jeq2014.11.0459

58. Strawn, L. K., Fortes, E. D., Bihn, E. A., Nightingale, K. K., Grohn, Y. T., Worobo, R. W., et al. (2013a). Landscape and meteorological factors affecting prevalence of three food-borne pathogens in fruit and vegetable farms. *Appl. Environ. Microbiol*. 79:588-600. doi: 10.1128/AEM.02491-12

59. Strawn, L. K., Grohn, Y. T., Warchocki, S., Worobo, R. W., Bihn, E. A., Wiedmann, M. (2013b). Risk factors associated with *Salmonella* and *Listeria monocytogenes* contamination of produce fields. *Appl. Environ. Microbiol*. 79:7618-7627. doi: 10.1128/AEM.02831-13

60. Strobl, C., Boulesteix, A. L., Zeileis, A., Hothorn, T. (2007). Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution. *BMC Bioinform*. 8(25). http://www.biomedcentral.com/1471-2105/8/25.

61. Strobl, C., Boulesteix, A, L., Kneib, T., Augustin, T., Zeileis, A. (2008). Conditional Variable Importance for Random Forests. BMC Bioinform. 9(307). http://www.biomedcentral.com/1471-2105/9/307.

62. Tam, S., and Petersen, A. (2014). B.C. Sprinkler irrigation manual. British Columbia Ministry of Agriculture. https://www2.gov.bc.ca/assets/gov/farming-natural-resources-and-industry/agriculture-and-seafood/agricultural-land-and-environment/water/irrigation/sprinkler-irrigation-manual/563000-1_web_titlepage_preface_tofc.pdf [Accessed July 2, 2020]

63. Tanaro, J. D., Piaggio, M. C., Galli, L., Gasparovic, A. M. C., Procura, F., Molina, D. A., et al. (2014). Prevalence of *Escherichia coli* O157:H7 in surface water near cattle feedlots. *Foodborne Pathog. Dis*. 11(12):960-965.

64. Tian, P., Yang, D., Shan, L., Wang, D., Li, Q., Gorski, L., et al., (2017). Concurrent detection of human norovirus and bacterial pathogens in water samples from an agricultural region in Central California Coast. *Front. Microbiol*. 8(1560):1-11. doi: 10.3389/fmicb.2017.01560

65. Truchado, P., Hernandez, N., Gil, M. I., Ivanek, R., Allende, A. (2018). Correlation between *E. coli* levels and the presence of foodborne pathogens in surface water: establishment of a sampling program. *Water Res*. 128:226-233. doi: 10.1016/j.watres.2017.10.041

66. Truitt, L. N., Vazquez, K. M., Pfuntner, R. C., Rideout, S. L., Havelaar, A. H., Strawn, L. K. (2018). Microbial quality of agricultural water used in produce preharvest production on the eastern shore of Virginia. *J. Food Prot*. 81(1):1661-1672. doi: 10.4315/0362-028X.JFP-18-185

67. US Centers for Disease Control and Prevention. (2006). Multistate outbreak of *E. coli* O157:H7 infections linked to fresh spinach (final update). https://www.cdc.gov/ecoli/2006/spinach-10-2006.html [Accessed on July 2, 2020]

68. US Centers for Disease Control and Prevention. (2018a). Multistate outbreak of Shiga toxin- producing *Escherichia coli* O157:H7 infections linked to leafy greens (final update). https://www.cdc.gov/ecoli/2017/o157h7-12-c17/index.html [Accessed on July 2, 2020]

69. US Centers for Disease Control and Prevention. (2018b). Multistate outbreak of *E. coli* O157:H7 infections linked to romaine lettuce (final update). https://www.cdc.gov/ecoli/2018/o157h7-04-18/index.html [Accessed on July 2, 2020]

70. U.S. Food and Drug Administration. (2015). Standards for the Growing, Harvesting, Packing, and Holding of Produce for Human Consumption. Final Rule, Docket No. FDA– 2011–N–0921, https://www.federalregister.gov/documents/2015/11/27/2015-28159/standards-for-the-growing-harvesting-packing-and-holding-of-produce-for-human-consumption [Accessed on July 2, 2019]

71. Uyttendaele, M. Jaykus, L. A., Amoah, P., Chiodini, A., Cunliffe, D., Jacxsens, L., et al. (2015). Microbial hazards in irrigation water: standards, norms, and testing to manage use of water in fresh produce primary production. *Compr. Rev. Food Sci. Food Saf*. 14:336-356. doi: 10.1111/1541-4337.12133

72. Wall, G. L., Clements, D. P., Fisk, C. L., Stoeckel, D. M., Woods, K. L., Bihn, E. A. (2019). Meeting report: key outcomes from a collaborative summit on agricultural water standards for fresh produce. *Compr. Rev. Food Sci. Food Saf*. 18:723-737. doi: 10.1111/1541-4337.12434

73. Weller, D., Belias, A., Green, H., Roof, S., Wiedmann, M. (2020a). Landscape, water quality, and weather factors associated with an increased likelihood of foodborne pathogen contamination of New York streams used to source water for produce production. *Front. Sustain. Food Syst*. 3(124). doi: 10.3389/fsufs.2019.00124

74. Weller, D., Brassill, N., Rock, C., Ivanek, R., Mudrak, E., Roof, S., et al. (2020b). Complex Interactions Between Weather, and Microbial and Physicochemical Water Quality Impact the Likelihood of Detecting Foodborne Pathogens in Agricultural Water. *Front. Microbiol*. 11(134):1-20. doi: 10.3389/fmicb.2020.00134

75. Weller, D., Love, T., Belias, A., Wiedmann, M. (2020c). Predictive models may complement existing *E. coli* water quality testing for managing enteric pathogen contamination of Northeastern streams used for produce production. *Front. Sustain. Food Syst*. 4:561517. doi: 10.3389/fsufs.2020.561517

76. Wilkes, G., Edge, T., Gannon, V., Jokinen, C., Lyautey, E., Medeiros, D., et al. (2009). Seasonal relationships among indicator bacteria, pathogenic bacteria, *Cryptosporidium* oocysts, *Giardia* cysts, and hydrological indices for surface waters within an agricultural landscape. *Water Res*. 43:2209-2223. doi: 10.1016/j.watres.2009.01.033

77. Wilkes, G., Edge, T. A., Gannon, V. P. J., Jokinen, C., Lyautey, E., Neumann, N. F., et al. (2011). Associations among pathogenic bacteria, parasites, and environmental and land use factors in multiple mixed-use watersheds. *Water Res*. 45:5807-5825. doi: 10.1016/j.watres.2011.06.021

78. Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S. M., et al. (2018). A new generation of the United States National Land Cover Database: requirements, research priorities, design, and implementation strategies. *ISPRS J. Photogramm. Remote Sens*. 146:108-123. doi: 10.1016/j.isprsjprs.2018.09.006

CHAPTER 6

EFFECT OF WEATHER ON THE DIE-OFF OF *ESCHERICHIA COLI* AND ATTENUATED *SALMONELLA ENTERICA* SEROVAR TYPHIMURIUM ON PREHARVEST LEAFY GREENS FOLLOWING IRRIGATION WITH CONTAMINATED WATER

Published in: Applied and Environmental Microbiology (2020), 86:e00899-20

ABSTRACT

The Food Safety Modernization Act (FSMA) includes a time-to-harvest interval following the application of non-compliant water to pre-harvest produce to allow for microbial die-off. However, additional scientific evidence is needed to support this rule. This study aimed to determine the impact of weather on the die-off rate of *E. coli* and *Salmonella* on spinach and lettuce under field conditions. Standardized, replicated field trials were conducted in California, New York, and Spain over two years. Baby spinach and lettuce were grown and inoculated with a ~10^4 CFU/mL cocktail of *E. coli* and attenuated *Salmonella*. Leaf samples were collected at 7 timepoints (0-96h) following inoculation; *E. coli* and *Salmonella* were enumerated. The associations of die-off with study design factors (location, produce type, and bacteria) and weather were assessed using log-linear and biphasic segmented log-linear regression. A segmented log-linear model best fit die-off on inoculated leaves in most cases, with a greater variation in the segment 1 die-off rate across trials [-0.46 (95% confidence interval (95% CI): -0.52, -0.41) to -6.99 (95% CI: -7.38, -6.59) $\log_{10}$ die-off/day] compared to the segment 2 die-off rate [0.28 (95% CI: -0.20, 0.77) to -1.00 (95% CI: -1.16, -0.85) $\log_{10}$ die-off/day]. A lower relative humidity was associated with a faster segment 1 die-off and an earlier breakpoint (the time when segment 1 die-off rate switches to the segment 2 rate).

Relative humidity was also found to be associated with whether die-off would comply with FSMA's specified die-off rate of -0.5 $\log_{10}$ die-off/day.

IMPORTANCE

The log-linear die-off rate proposed by FSMA is not always appropriate, as the die-off of foodborne bacterial pathogens and specified agricultural water quality indicator organisms appear to commonly follow a biphasic die-off pattern with an initial rapid decline followed by a period of tailing. While we observed substantial variation in the net culturable population levels of *Salmonella* and *E. coli* at each time point, die-off rate and FSMA compliance (i.e., at least a 2 $\log_{10}$ die-off over 4 days) appear to be impacted by produce type, bacteria, and weather; die-off on lettuce tended to be faster than that on spinach, die-off of *E. coli* tended to be faster than that of attenuated *Salmonella*, and die-off tended to become faster as relative humidity decreased. As such, the use of a single die-off rate for estimating time-to-harvest intervals across different weather conditions, produce types, and bacteria should be revised.

INTRODUCTION

As consumers increase consumption of fresh produce and as detection of illness cases and contaminated products improve, there has been an increase in the number of foodborne illnesses and recalls linked to produce (*31*). In particular, 51.7% (N=571) of produce related outbreaks in developed countries have been linked to leafy greens (*46*). In the United States, there have been 6 multistate *E. coli* O157:H7 outbreaks linked to leafy greens between 2017 and 2019, which have caused a total of 497 illnesses and 6 deaths (*10, 11, 12, 13, 14, 64*). There have also been 3 recalls in 2019 caused by *Salmonella* contamination of leafy greens (*65*). As such, there is a shared pressure among industry, academia, and public health agencies to develop better risk

management strategies for contamination of leafy greens by foodborne pathogens, including *E. coli* and *Salmonella*. Equally, there is a need to challenge the underlying assumptions of existing one-size-fits-all interventions and standards for well-known hazards, such as contaminated agricultural water.

A common route for transfer of pathogenic *E. coli* and *Salmonella* in the preharvest produce environment is surface water (*8, 22, 29, 32, 35, 54, 55, 59, 66*); surface water has also been identified as a potential cause of several outbreaks (*9, 11, 12*). Surface water can be applied to preharvest produce as irrigation water, through agrichemical applications, and frost protection, among others. As such, as part of the Food Safety Modernization Act (FSMA), the FDA proposed an agricultural water standard to define the microbial quality required for any source of surface water applied to the harvestable potion of produce (*63*). The standard states 20 water samples must be collected from each water source over a two to four-year period prior to water application and tested for generic *E. coli* level. The geometric mean *E. coli* level must be <126 CFU/100mL and the statistical threshold value (i.e. the 90^{th} percentile) must be <410 CFU/100mL for those samples (*15*). If the water does not meet this standard, growers can choose to (i) not use the water source, (ii) treat the water prior to use, or (iii) wait up to 4 days from water application to harvest to achieve microbial die-off to a level compliant with the agricultural water standard. This assumes a -0.5 $\log_{10}$ die-off per day (*16*, a negative die-off rate indicates a reduction in bacterial counts and a positive die-off rate indicates an increase in bacterial counts). However, the die-off rate within the agricultural water standard has been challenged due to lack of agreement in literature (*71*).

Several studies have investigated the survival and die-off of attenuated EHEC and non-pathogenic *E. coli* (*3, 6, 17, 18, 24, 26, 27, 28, 33, 37, 40, 45, 47, 49, 50, 52, 72, 73*) and *Salmonella* (*33, 42, 45*) on in-field leafy greens. For instance, Moyne et al.

(*49*) irrigated in-field lettuce with water contaminated with attenuated *E. coli* O157:H7 in 3 replicated trials in Salinas, California and observed a 2-3 $\log_{10}$ reduction within 2h of inoculation. In comparison, Chase et al. (*18*), inoculated in-field lettuce with the same strains of attenuated *E. coli* O157:H7 in 2 replicated trials in Salinas, California and observed a net reduction of 2.6 and 3.2 $\log_{10}$ over 10 days. This indicates substantial variation in observed die-off between studies. As such, additional research is needed to better understand the drivers (e.g., weather) of this variability in foodborne pathogen population dynamics on produce under field conditions. Additionally, only a small number of studies have investigated die-off in multiple climatic regions using a standardized protocol (*6, 27*). Without multiple climatic regions there is insufficient variability in weather conditions, making it difficult to identify associations between weather and die-off. Furthermore, several studies indicate the log-linear die-off pattern included in the FSMA agricultural water rule may not be appropriate to model in-field pathogen die-off (*47, 72*). The use of an inappropriate die-off pattern can lead to an over- or under-estimation of the actual net die-off, which can lead to an under-estimation of the bacterial counts on produce at the time of harvest. Alternatively, it can require produce growers to extend the time between water application and harvest more than is actually required. Lastly, more information is needed on the difference between die-off of foodborne pathogens, indicator organisms, and surrogate organisms used to conduct such field studies, as the majority of studies were conducted on *E. coli* die-off on lettuce; additional information on die-off variability on different leafy green varieties is also needed.

As such, the current study aimed to help fill these knowledge gaps and develop a better understanding of the population dynamics of surrogate organisms across three climatic regions to investigate the impact of weather on net viable bacteria recovery over the proposed die-off interval. In particular, the objectives of this study were to (i)

quantify and compare the die-off rates and die-off patterns of *E. coli* and attenuated *Salmonella* on in-field baby spinach and lettuce in replicated controlled trials in California, New York, and Spain and (ii) identify weather factors associated with the die-off rate and die-off pattern.

MATERIALS AND METHODS

Field set-up. Replicated controlled field trials were conducted in 2017 and 2018 in three locations: Davis, California (University of CA, Plant Sciences Field Research Facility); Freeville, New York (Homer C. Thompson Research Farm); and Murcia, Spain (La Matanza Research Farm). In each trial, 6 rows of lettuce or spinach seed were sown into approx. 1.5 m wide by 4 m long plots; the seeding rate was between 1.25 cm and 4 cm. Seeding in California was performed by creating multirow seed line slots with tractor-mounted shallow shanks and hand-distributing pre-calibrated seed masses blended with sterile horticultural sand to assist in uniform placement. Seeding in New York was performed using a JANG Seeder, model JPH (Mechanical Transplanter Company, Holland, Michigan). Seeding in Spain was performed by hand, using foam trays with evenly spaced holes to push the seeds through and into the soil. Due to uncontrollable factors (e.g., animal intrusion, poor stand germination, and adverse weather), the number of plots, spinach and lettuce varieties, and time between planting and harvest varied between trials (Table 6.1). For instance, in trial NY3 two varieties of spinach were compared: the variety used in the other two locations and a variety well suited for New York conditions. The new variety was used in the remaining New York trials (i.e., NY1, NY2, and NY4) due to the poor stand germination of the variety used in California and Spain. Fertilization, pesticide application, and overhead irrigation during produce growing were applied as deemed necessary for the conditions experienced in each location and according to industry standards.

Table 6.1. Description of the experimental setup for each trial.

Location	Trial	Produce Type	Produce Variety[a,b]	Number of Plots[c]	Date of Inoculation	Time from Planting to Inoculation (days)	Sample Collection Times (h)[d]	No. Samples Collected per Plot per Timepoint	Data included in analysis[e]
California	CAp	Lettuce	Tamarindo	4	7/19/2018	44	0, 24, 96	3	No
		Spinach	Acadia F1	4	7/12/2018	37	0, 24, 96	3	No
	CA1	Lettuce	Tamarindo	4	11/12/2018	38	0, 4, 8, 24, 48, 72, 96	5	Yes
		Spinach	Acadia F1	4	11/12/2018	38	0, 4, 8, 24, 48, 72, 96	5	Yes
	CA2	Lettuce	Tamarindo	4	12/18/2018	57	0, 4, 8, 24, 48, 72, 96	5	Yes
		Spinach	Acadia F1	4	12/18/2018	57	0, 4, 8, 24, 48, 72, 96	5	Yes
	CA3	Lettuce	Tamarindo	4	7/1/2019	59	0, 4, 8, 24, 48, 72, 96	5	Yes
		Spinach	Acadia F1	4	7/1/2019	59	0, 4, 8, 24, 48, 72, 96	5	Yes
New York	NY1	Lettuce	Tamarindo	3	8/27/2018	28	0, 4, 8, 24, 48, 72, 96	5	Yes
		Spinach	Seaside F1	3	8/27/2018	48	0, 4, 8, 24, 48, 72, 96	5	Yes
	NY2	Lettuce	Tamarindo	3	10/1/2018	38	0, 4, 8, 24, 48, 72, 96	5	Yes
		Spinach	Seaside F1	3	10/1/2018	38	0, 4, 8, 24, 48, 72, 96	5	Yes
	NY3	Spinach	Acadia F1	2	7/16/2018	31	0, 4, 8, 24, 48, 72, 96	5	Yes
		Spinach	Seaside F1	2	7/16/2018	31	0, 4, 8, 24, 48, 72, 96	5	Yes
	NY4	Lettuce	Tamarindo	4	7/1/2019	40	0, 4, 8, 24, 48, 72, 96	5	Yes
		Spinach	Seaside F1	4	7/1/2019	40	0, 4, 8, 24, 48, 72, 96	5	Yes
Spain	SP1	Lettuce	Tamarindo	4	5/29/2018	47	0, 4, 8, 24, 48, 72, 96	5	Yes
		Spinach	Acadia F1	4	5/29/2018	47	0, 4, 8, 24, 48, 72, 96	5	Yes
	SP2	Lettuce	Tamarindo	4	1/8/2019	91	0, 4, 8, 24, 48, 72, 96	5	Yes
		Spinach	Acadia F1	4	1/8/2019	91	0, 4, 8, 24	5	No
	SP3	Lettuce	Tamarindo	4	4/29/2019	77	0, 4, 8, 24, 48, 72, 96	5	Yes
		Spinach	Acadia F1	4	4/29/2019	77	0, 4, 8, 24, 48, 72, 96	5	Yes
	SP4	Lettuce	Tamarindo	2	7/1/2019	26	0, 4, 8, 24, 48, 72, 96	3	Yes
		Spinach	Acadia F1	4	7/1/2019	26	0, 4, 8, 24	3	No

[a]Tamarindo lettuce and Acadia F1 spinach were supplied by Enza Zaden (Enkhuizen, Netherlands). Tamarindo is red leaf lettuce. Tamarindo and Acadia F1 are ideal for baby leaf.

[b]Seaside F1 spinach was supplied by Harris Seeds (Rochester, NY) and is ideal for baby leaf.

[c]The number of plots planted per trial varied to balance available resources and the need for additional trials in each location.

[d]Sample collection times are reported in hours past inoculation. Trials where samples were not collected at all seven time points was caused by crop loss.

[e]Trials or produce types in a trial where samples were not collected at all time points were excluded from data analysis. This was due to crop loss.

Testing for naturally occurring rifampicin resistant *E. coli* and attenuated *Salmonella*. The following control samples were collected prior to each trial to test for naturally occurring rifampicin resistant *E. coli* and *Salmonella*: soil samples from each experimental plot, one spinach sample, one lettuce sample, and 1L of water used for irrigation. The samples were stored on ice, transferred to the lab, and processed within 24h of collection. For produce and soil samples, 25g was weighed out in separate sterile bags (Nasco, Fort Atkinson, WI). The samples were then diluted 1:10 with tryptic soy broth (Becton Dickson, Franklin Lakes, NJ) supplemented with 0.1 g/L Rifampicin (TSB+R; MilliporeSigma, Burlington, MA). The 1L water sample was run through a 0.45 µm NEO-GRID filter (Neogen Corp., Lansing, MI). If the water contained large amounts of sediment, multiple filters were used. The filters were then transferred to a sterile bag and 90 mL of TSB+R was added. All control samples were incubated at 37°C for 18 to 24h. After incubation, 50 µL of each enriched sample was streaked onto an *E.coli* ChromAgar (DRG International, Springfield, NJ) plate supplemented with 0.1 g/L Rifampicin (ECC+R) and onto a *Salmonella* ChromAgar (DRG International) plate supplemented with 0.1 g/L Rifampicin (SC+R). The plates were incubated at 37°C for 18-24h. The presence of blue colonies on ECC+R plates indicates a positive result for rifampicin resistant *E. coli* and the presence of mauve colonies on SC+R plates indicates a positive result for rifampicin resistant *Salmonella*.

Inoculum preparation and inoculation. Three strains of rifampicin resistant *E. coli* (TVS 353, TVS 354, and TVS 355), provided by the Suslow lab at the University of California, Davis (*62*), and two strains of rifampicin resistant attenuated *Salmonella* [*Salmonella enterica* sv. Typhimurium strain MHM112 (*23*) and *Salmonella enterica* sv. Typhimurium UK-χ3985 (*21*)] were used to inoculate the lettuce and spinach plots. The rifampicin resistance in the two *Salmonella* strains was developed by the Suslow lab using the procedure described by Lopez-Velasco et al. (*44*); the rifampicin resistant

strains derived from MHM112 and UK-χ3985 were named attPTVS 355 and attPTVS 337, respectively. Each strain was streaked onto separate tryptic soy agar plates (Becton Dickson) supplemented with 0.1 g/L Rifampicin to form a confluent lawn. The plates were incubated at 37°C for 18-24h. After incubation, each plate was flooded with 5 mL of phosphate buffered saline (PBS), and the cells were suspended using a sterile loop or L-spreader. Each strain's suspension was then transferred to a separate sterile bottle. If there were visible cell masses remaining on the plates, the washing step was repeated. Enough PBS was then added to each sterile bottle containing the cell suspensions to reach final volumes of 100 mL. Each bottle was vortexed, and 10 mL was transferred to a separate, sterile 15 mL tube. The cell suspensions for each strain were spun down at 2,000 x g for 10 min in New York and 2,500 x g for 5 min in Spain and California; the differences in centrifuge conditions were due to differences in the equipment available in each lab. Following centrifugation, the supernatant was pipetted off and the cells were washed twice in PBS using the same centrifuge conditions listed above. The cells were then re-suspended in 6 mL of PBS and the optical density at 600 nm was measured to confirm the suspensions were at approx. 9 $\log_{10}$ CFU/mL. The cell suspensions were stored at 4°C overnight; since die-off was calculated relative to the 0h timepoint, the 12h hold at 4°C should not bias the results. After approx. 12h, 4 mL of each *E. coli* strain's suspension were mixed into a sterile 15 mL tube and 6 mL of each *Salmonella* strain's suspension were mixed into a second sterile 15 mL tube. Serial dilutions were then performed on each bacteria cocktail (i.e., *E. coli* and *Salmonella*) to reach approx. 6 $\log_{10}$ CFU/mL. Four sanitized 2 L bottles were then obtained, and 1.8 L of PBS, 100 mL of the *E. coli* cocktail, and 100 mL of the *Salmonella* cocktail were added to each bottle to reach a final concentration of approx. 5 $\log_{10}$ CFU/mL for each bacterium (i.e., *Salmonella* and *E. coli*). A 5 mL aliquot from each bottle was transferred to separate, sterile 15 mL tubes and stored at 4°C to be used to confirm the inoculum

concentration.

Confirmation of the inoculum concentration was performed within 24 h of inoculum preparation. The aliquot from each bottle was diluted by 10^{-1} and 10^{-2}, and 100 µL of each dilution was plated on ECC+R. The plates were incubated at 37°C for 18-24h. After incubation, the number of blue colonies (*E. coli*) and white colonies (*Salmonella*) were counted and recorded.

Inoculation was targeted to be performed when the lettuce plants developed 6 true leaves, or 30 to 40 days following seeding. At this time, the spinach plants were expected to have 10 to 12 true leaves. However, due to weather conditions, some trials could not be conducted at this targeted time (Table 6.1). In all three experimental locations, inoculation was performed using the same make of a CO_2 powered backpack sprayer with the pressure between 27 and 30 psi and 2 Turbo TeeJet (tip #8) nozzles spaced 38 inches apart (R&D Sprayers, Opelousas, LA). The inoculum was applied to the plots at 2 L per approx. 45 m^2.

Sample collection. Following inoculation, samples were collected at the following time points: 0, 4, 8, 24, 48, 72, and 96h. For trials CAp (preliminary trial in California), SP2 spinach, and SP4 spinach, samples were collected at a reduced number of time points due to crop loss and were excluded from all data analyses (Table 6.1). At each time point, 5 samples were collected per plot, with the exceptions of trials CAp and SP4 (Table 6.1). Each sample consisted of 6 adjacent plants from a single row. The locations of these samples were randomly selected. No samples were harvested from the outer plot rows. The samples were harvested by cutting each plant in the identified sample approx. 2 cm above the soil line with scissors and transferring them to a sterile bag. The scissors were wiped with a 20% bleach wipe, followed by a 70% ethanol wipe between each sample. Gloves were changed and sprayed with 70% ethanol between each sample. The samples were stored on ice and transferred back to the lab for microbial testing. All microbial testing was performed within 24h of sample collection.

Die-off of *Salmonella* and *E. coli* in the soil of spinach and lettuce plots in trials CA1 and CA2 was also assessed; details are provided in Supplementary Information Appendix A.

Microbial testing. The weight of each sample was measured and recorded. Samples collected at 0, 4, or 8h past inoculation were diluted 1:5 with PBS. Samples collected at 24, 48, 72, or 96h past inoculation were diluted 1:10 with PBS. Larger dilution factors were used at the later time points to ensure enough washate was available to plate larger volumes (e.g., 100 mL). All samples were massaged by hand for 1 min.

All samples were enumerated for rifampicin resistant *E. coli* and *Salmonella*. Based on the discretion of each location, between 10 µL and 100 mL were plated for each bacterium per sample to increase the likelihood of observing the countable range. For plating volumes less than or equal to 250 µL, the samples were spread plated on ECC+R and incubated at 37°C for 18-24h. The blue and white colonies were counted and recorded as *E. coli* and *Salmonella*, respectively. For plating volumes of 1 mL or greater, the samples were filtered through 0.45 µm pore-size NEO-GRID units. When using the filters, *Salmonella* can no longer be reliably counted on ECC+R plates, so filtering of each sample was performed in duplicate; the first filter was aseptically transferred to an ECC+R plate and the second filter was aseptically transferred to a SC+R plate. The ECC+R plates were incubated at 37°C for 18-24h and the blue colonies were counted and recorded as *E. coli*. The SC+R plates were incubated at 37°C for 42-48h and the mauve colonies were counted and recorded as *Salmonella*. Following plating, the samples were stored at 4°C. The number of colonies per plate where converted to CFU/100g of produce, which is referred to as the "population level" of *Salmonella* or *E. coli* present on the sample; all counts were reported in $\log_{10}$ CFU/100g of produce to avoid negative $\log_{10}$ counts.

Enrichment. Enrichment was performed on any sample negative by enumeration. To do so, the sample was diluted 1:2 with 2X TSB supplemented with 200 g/L of Rifampicin (2X TSB+R), based on the volume of PBS remaining in the sample (the sample refers to the plant material and the PBS remaining in the sample). For example, if there was 50 mL of PBS remaining in the sample bag, 50 mL of 2X TSB+R was added. The enrichments were then incubated at 37°C for 18-24h. Following incubation, 50 µL of the enrichment was streaked onto an ECC+R plate to test for the presence of *E. coli* and/ or onto a SC+R plate to test for the presence of *Salmonella*. All plates were incubated at 37°C for 18-24h. The presence of blue colonies on ECC+R were recorded as *E. coli* positive and the presence of mauve colonies of SC+R were recorded as *Salmonella* positive.

Strain identification. A PCR protocol was developed to differentiate the *E. coli* inoculum strains to determine if there was a difference in strain survival. The protocol utilized 4 primer sets: 1 specific for each strain and 1 that amplified all 3 strains (Table 6.2). Dirty lysates were prepared for each isolate by transferring a portion of a colony into 100 µL of dH_2O in a 0.2 mL tube. These suspensions were placed in the thermocycler and heated to 95°C for 15 minutes. Fifty µL reactions were performed, with the reagents at the following final concentrations: 1X Green GoTaq Flexi Reaction Buffer (Promega, Madison, WI), 1.5 mM $MgCl_2$ (Promega), 0.2 mM of each dNTP (Thermo Scientific, Waltham, MA), 0.2 µM 353F, 0.2 µM 353R, 0.2 µM 354F, 0.2 µM 354R, 0.2 µM 355F, 0.2 µM 355R, 0.2 µM 35XF, 0.2 µM 35XR (Integrate DNA Technologies, Coralville, IA), 1.25 U GoTaq Polymerase (Promega), <0.5 µg/50 mL DNA template. The thermocycler conditions were as follows: 94°C for 2 minutes; 30 cycles of 94°C for 30 seconds, 57°C for 1 minute, 72°C for 1.5 minutes; 72°C for 7 minutes, 4°C for ∞. Gel electrophoresis was performed using a 1.5% agarose gel at 120 volts. The primers amplified a 240 bp fragment from TVS 353, a 521 bp fragment from

TVS 354, a 960 bp fragment from TVS 355, and an 1835 bp fragment with the universal primers. In California, PCR was performed as follows for *E. coli* positive samples: up to 6 isolates per sample in CAp; 1-29 isolates per 24, 72, and 96h samples in CA1; 2-18 isolates per 24 and 96h samples in CA2; and 1-89 isolates per 24, 72, and 96h samples in CA3. Strain characterization of *E. coli* in the soil samples collected in trials CA1 and CA2 was also performed; details can be found in Supplementary Information Appendix A. In New York, PCR was performed on up to 6 isolates for each *E. coli* positive sample (by either enumeration or enrichment) for each trial. In Spain, PCR was performed as follows on *E. coli* positive samples: up to 1 isolate per 48, 72, and 96h samples in SP1; up to 6 isolates per 24 and 48h lettuce samples in SP2; up to 6 isolates per 8, 24, and 48h samples in SP3; and up to 6 isolates per 8 and 24h spinach samples and up to 6 isolates per 8, 24, and 48h lettuce samples in SP4.

Table 6.2. Primer sequences used to differentiate between the 3 *E. coli* inoculum strains.

Primer Name	Target Strain	Sequence
353F	TVS353	TGACGGACAGGGACTCTATCTG
353R	TVS353	CAGCGTTCGCTCACTGAGAG
354F	TVS354	TAGGTTTGTTCACATTAGGTGATGTCG
354R	TVS354	AAATGTGGGTATGGCATATGGCAG
355F	TVS355	GTGACACCAATGACATCTGATGTTATCC
355R	TVS355	CGTCCTTATCCTGTTGGCTTGTG
35XF	All 3 strains	TTCGACAACGGTATTATTCTCTGCC
35XR	All 3 strains	TATCAATGACCCGAATCTGATCCTCG

To evaluate differential survival of the 2 attenuated *Salmonella* inoculum strains, *Salmonella* colonies were streaked on Xylose Lysine Deoxycholate agar supplemented with 0.1 g/L Rifampicin (XLD+R). The XLD+R plates were incubated at 35°C for 18-24h and the resulting color of the colonies was recorded. PTVS 337 does not produce H_2S and forms pink colonies on XLD+R, while PTVS 355 does produce H_2S and

produces black colonies on XLD+R. In CA1, CA2, and CA3, 1-121 colonies per plot were tested from two time points (24 and 96h). In NY4, 40 characteristic *Salmonella* colonies per plot were tested from two time points (24 and 96h). In SP2 and SP3, 25 characteristic *Salmonella* colonies per plot were tested from two time points (24 and 96h).

Monitoring environmental conditions. The following weather conditions were recorded in all locations: Temperature (°C), relative humidity (%), solar radiation (kilowatts per meter square, kW/m^2), precipitation (mm), and wind speed (m/s). The geographic coordinates (WGS 84 Web Mercator) of the weather stations were: Lat: 38.53, Lon: -121.79 (elevation: 18.3 m) in California; Lat: 42.52, Lon: -76.33 (elevation: 335.3 m) in New York; and Lat: 38.11, Lon: -1.03 (elevation: 135 m) in Spain. All weather data was cleaned in R version 3.5.3 (R Core Team, Vienna, Austria). Hourly dew point was also calculated using humidity.to.dewpoint function in the "weathermetrics" package (*1*). Leaf wetness (min) was collected in New York and Spain, however, no associations between leaf wetness and die-off were identified (and so no data are shown). The weather variables derived from recorded weather data are listed in Table 6.3. All weather variables were calculated over the 96, 24, and 8h following inoculation. The summary variables for the three time periods were developed to serve as three practical ways a grower could use the study findings (at the end of a 4-day wait period, 24h, or 8h after irrigation) to plan harvest. No solar radiation variables were calculated for the 8h following inoculation because the weather station malfunctioned during this time in trial NY3. The weather station during trial SP4 also malfunctioned 55h following inoculation, so 96h weather variables for this trial were only calculated with available data. However, the weather appears to be relatively stable from day to day, so we did not expect this to affect the results. A standard operating procedure (SOP) document was developed to provide the basis for the above described

standardized field and microbial testing components of study design across the three experimental locations. The SOP is provided in Supplementary Information Appendix B.

Table 6.3. Definitions of die-off outcomes and predictor variables (study design and weather), considered in statistical analyses at the plot-level. Weather variable sets were created for three specified time periods of interest (i.e., the 8h[a], 24h, and 96h) following inoculation and one set was used at a time in analysis.

Variable type[b]	Notation	Definition	Unit
O	seg1	Segment 1 die-off rate	$\log_{10}$ die-off/day
O	se1	Segment 1 die-off rate standard error	$\log_{10}$ die-off/day
O	seg2	Segment 2 die-off rate	$\log_{10}$ die-off/day
O	se2	Segment 2 die-off rate standard error	$\log_{10}$ die-off/day
O	Bp	Breakpoint between segment 1 and segment 2	Days
O	Die-off pattern	Indicates if a biphasic segmented log-linear fit is superior to a log-linear fit for each plot and bacteria combination. The superior model fit for each plot and bacteria subset was determined such that for the segmented fit to be superior, its Bayesian information criteria (BIC) value must be 10 or more than the BIC value of the log-linear model	log-linear (baseline)/ segmented
O	FSMA compliance	If the observed segmented die-off in each bacterium, plot combination is compliant with the Food Safety Modernization Act (FSMA) (i.e., $\geq$2 $\log_{10}$ overall die-off from 0h to 96h)	Not compliant (baseline)/ compliant
S	Produce type	Designates spinach or lettuce. Lettuce is the baseline in all regression models	Lettuce (baseline)/ spinach
S	Bacteria	Designates *Salmonella* or *E. coli*. *E. coli* is the baseline in all regression models.	*E. coli* (baseline)/ *Salmonella*
S	Location	Geographic location where the experiment was conducted	California (baseline)/ New York/ Spain
W	Min. Temperature	Minimum temperature during a time period of interest	°C
W	Max. Temperature	Maximum temperature during a time period of interest	°C
W	Avg. Temperature	Average temperature during a time period of interest	°C
W	Temperature Range	Maximum minus minimum temperature during a time period of interest	°C

W	Max. Temperature Change Rate	Maximum change in temperature from one hour to the next during a time period of interest	°C/h
W	Min. Relative Humidity	Minimum relative humidity during a time period of interest	%
W	Max. Relative Humidity	Maximum relative humidity during a time period of interest	%
W	Avg. Relative Humidity	Average relative humidity during a time period of interest	%
W	Relative Humidity Range	Maximum minus minimum relative humidity during a time period of interest	%
W	Max. Relative Humidity Change Rate	Maximum change in relative humidity from one hour to the next during a time period of interest	%/h
W	Max. Solar Radiation[a]	Maximum solar radiation during a time period of interest	kW/m^2
W	Avg. Solar Radiation[a]	Average solar radiation during a time period of interest	kW/m^2
W	Max. Solar Radiation Change Rate[a]	Indicates to the maximum change in solar radiation from one hour to the next during a time period of interest	$kW/m^2 \cdot h$
W	Precipitation	Designates if there was precipitation during a time period of interest	yes/ no
W	Min. Wind Speed	Minimum wind speed during a time period of interest	m/s
W	Max. Wind Speed	Maximum wind speed during a time period of interest	m/s
W	Avg. Wind Speed	Average wind speed during a time period of interest	m/s
W	Wind Speed Range	Maximum minus minimum wind speed during a time period of interest	m/s
W	Max. Wind Speed Change Rate	Maximum change in wind speed from one hour to the next during a time period of interest	$m/s \cdot h$
W	Min. Dew Point	Minimum dew point during a time period of interest	°C
W	Max. Dew Point	Maximum dew point during a time period of interest	°C

W	Avg. Dew Point	Average dew point during a time period of interest	°C
W	Dew Point Range	Maximum minus minimum dew point during a time period of interest	°C
W	Max. Dew Point Change Rate	Maximum change in dew point from one hour to the next during a time period of interest	°C/h

[a]Solar radiation variables were not created for the 8h weather variables due to missing data.

[b]O = outcome, S = study design, W = weather.

Statistical Analyses.

Data processing. All data cleaning and analyses were performed in R version 3.5.3. Due to the microbial testing strategy, samples could be (A) positive with microbial counts but above the countable range [i.e., too numerous to count, N=7], (B) positive with counts within the countable range [i.e., positive by enumeration, N=4,237], (C) negative by enumeration and positive by enrichment [N=452], and (D) negative by enumeration and enrichment [N=185]. There were also positive samples with an unknown population level due to (E) sample loss after enumeration [i.e., enumeration was performed but no enrichment was performed (N=12)] or (F) inability to read plates because of excessive mud on the sample [N=7]. To account for unknown population levels on positive samples, previous studies have imputed a single value (e.g., lower-detection limit, the mid-point between lower- and upper-detection limits, or the upper detection limit). However, this strategy is not optimal, as it can lead to biased results (*30*). As such, in the current study, multiple imputations were performed for the samples with unknown population levels (in categories A, C, D, E and F), as this has been shown to limit bias (*30*). To do so, the lower and upper limits of detection were calculated for each of these samples (based on the sample dilution) and a population level between those two limits of detection was randomly selected (i.e., imputed) using a uniform distribution to reflect uncertainty; ten imputation rounds were performed per sample. For samples that were lost after enumeration (category E), the concentration was imputed such that the lower bound for imputation was 0 and the upper bound for imputation was the limit of quantification for enumeration. For samples with un-readable plates (category F), it was known if the bacterium of interest was present, however, the exact concentration was unknown. As such, the lower bound of imputation was set at the lower limit of quantification by enumeration and the upper bound was set at 10^7 CFU/100g (i.e., ~1 $\log_{10}$ above the highest counted population level with filter

plating). For samples above the countable range (category A), the upper limit was set at 10^7 CFU/100g of produce. The microbial counts recorded for samples in category B were used as is, in the analysis except for a correction due to the filter vs spread plating methodology explained below. Samples in the category D were imputed with a value between 0 and the limit of detection for enrichment, as each sample was used for testing *E. coli* and *Salmonella*. This means the entire sample could not be tested for each respective bacterium, and it is unknown if the samples negative by enrichment were truly negative; the consequence is a possible overestimation by imputation in these samples. However, considering only 185 samples required this treatment we do not expect a measurable effect on the findings. The raw and imputed microbial counts data were compiled into a "raw imputed" dataset.

Through data visualization, it was determined there were systematically higher population levels of *Salmonella* on the spinach and lettuce when samples were tested using spread plating as compared to filtering. We hypothesized 2 potential reasons for this difference: (i) there were false positive *Salmonella* colonies on ECC+R plates used for spread plating, since different plating media were used for enumerating *Salmonella* by spread plating (i.e., ECC+R) and filter plating (SC+R) and (ii) the NEO-GRID filters restricted the diffusion of nutrients from the SC+R plates to allow for *Salmonella* colony growth. To assess the potential for false positive *Salmonella* colony growth on ECC+R, 3,588 colonies were streaked from ECC+R plates onto SC+R plates and incubated at 37°C for 18-24h; the development of mauve colonies indicates a *Salmonella* positive. After incubation, 99.6% (3,572/ 3,588) of colonies were mauve, indicating the difference between spread and filter plates was not due to false positives on ECC+R. As such, we concluded the difference was due to the reduced ability of nutrients to pass from SC+R media through the NEO-GRID filter to allow for *Salmonella* colony growth. To correct for this, the percentage difference between spread and filter plates was

calculated separately for spinach and lettuce using samples for which we had available counts from both spread and filter plates (N=368 for spinach, N=300 for lettuce). The percentage difference in spinach and lettuce were used to adjust the population levels of spread plates samples down to the expected concentration if filter plating had been used, because the difference in concentrations between spread and filter plates was thought to be caused by the underestimation of the number of colonies by filter plates. This methodology was chosen because there were fewer spread plated samples than filtered samples over the whole length of the experiment, therefore requiring correcting the concentration of fewer samples. Also, the main interest was in correctly capturing the difference in counts from time point to time point to allow estimation of the microbial die-off rate. An acceptable side-effect of this approach was that the true microbial concentration at each time point was likely slightly underestimated.

After the above correction for the enumeration method (filter vs plate spreading) was applied, the "raw imputed" microbial count data was subset by each combination of plot and bacteria (Figure 6.1), resulting in a total of 70 plot level subsets across all trials for *E. coli* and an equal number for *Salmonella* (since each plot was inoculated with a mixture of strains of both bacteria). For instance, an individual subset of data was created for *E. coli* on the first spinach plot in trial NY3 comprising of "raw imputed" counts at each of the 7 timepoints of sample collection. A log-linear and a biphasic, segmented log-linear regression model were fit using the lm function for each of the ten imputed data sets for each subset (i.e., for each plot-commodity-bacteria combination). These are referred to as plots. The breakpoint in each of the segmented models was identified by selecting the breakpoint that minimized the deviance of the model. Additional biphasic models (e.g., Weibull) were not evaluated due to the difficulty with practical interpretation. Each of the ten models for a given plot were then combined using the pool function (based on the principles explained by Rubin, *57*) in the "mice"

package (*67*) into a single model. For each of plot models, Table 6.3 and Figure 6.1 show the statistics that were recorded and subsequently used as outcome variables in the plot-level analysis of predictors: (i) segment 1 die-off rate coefficient (seg1), (ii) segment 1 die-off rate standard error (se1), (iii) segment 2 die-off rate coefficient (seg2), (iv) segment 2 die-off rate standard error (se2), (v) breakpoint between segment 1 and segment 2 (bp), (vi) if a segmented log-linear or log-linear model had a superior fit for the subset (die-off pattern), and (vii) if the observed die-off in the subset is faster than the FSMA die-off rate with FSMA (FSMA compliance). The 7 outcomes and predictors [i.e., study design factors (i.e., bacteria, produce type, and location) and weather factors] from individual plots were compiled into a "processed" dataset for further plot-level analyses of microbial die-off (Table 6.3). This approach was taken because we hypothesized there were different association between weather factors and die-off for each of the segments. Additionally, it is more practical for produce growers, public health agencies, and academia to utilize the relationship between weather, study design factors, and die-off rates compared to the relationship with microbial concentration on individual samples. This analysis method also readily takes data from multiple imputations in "raw imputer" dataset into the "processed" dataset. For the "die-off pattern" outcome (Table 6.3), to assess if a segmented log-linear or log-linear model had a superior fit for each plot, the Bayesian information criterion (BIC) was estimated for each model type. For the segmented fit to be considered preferred, its BIC value must be 10 or more than the BIC value of the log-linear model. A cut-off value of 10 was selected according to the findings of Kass and Wasserman (*41*). For the "FSMA compliance" outcome (Table 6.3), to assess if the observed die-off in a plot is compliant with FSMA (i.e., at least a 0.5 $\log_{10}$ reduction/day), it was determined if the segmented die-off calculated for each plot would achieve at least a 2 $\log_{10}$ reduction in 4 days. This assumes, starting at a population level of 4 $\log_{10}$ CFU/100g of produce and 100 mL is

applied to each 100g of produce, at least a 2 $\log_{10}$ reduction would need to be achieved in 4 days for the produce to be considered compliant (i.e., <126 CFU/100mL) with the FSMA agricultural water standard.

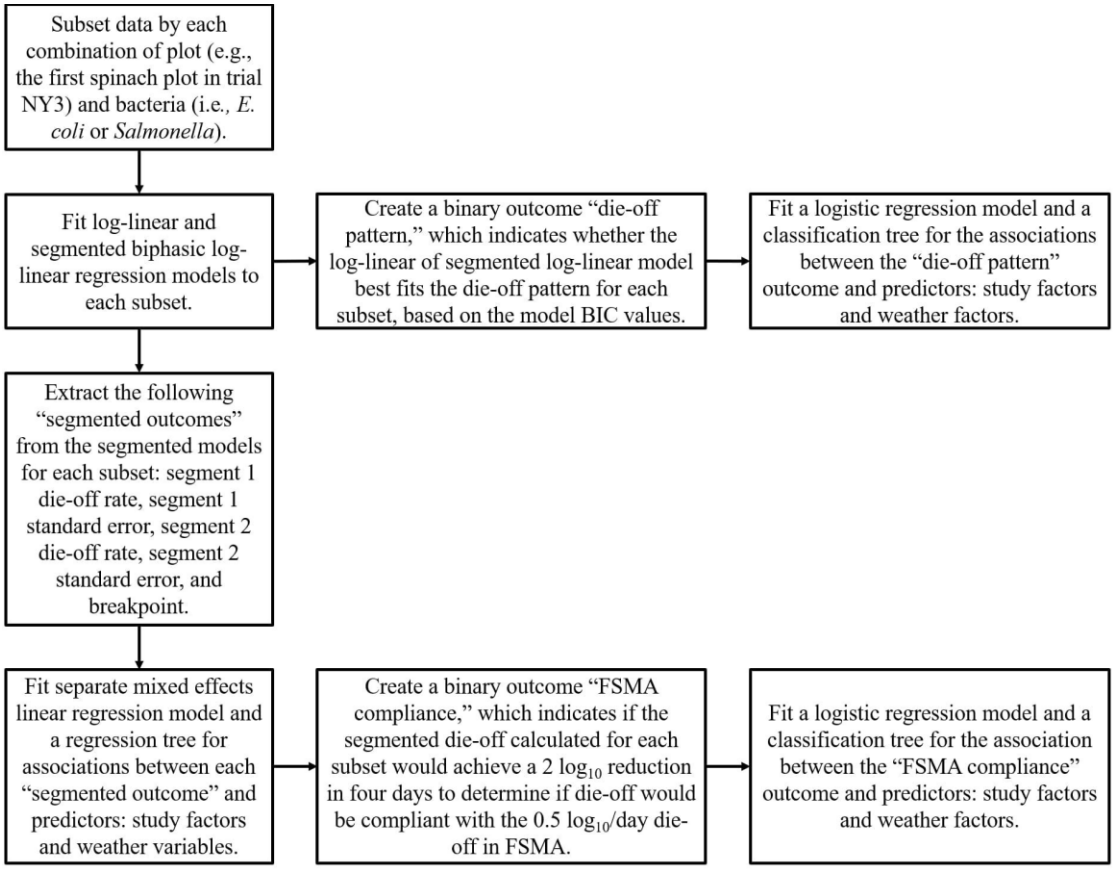


Figure 6.1. Overview of the statistical analysis plan.

Univariable analyses. Mixed effects linear regression was performed on the "processed" dataset using the lmer function in the "lme4" package (*4*) to determine univariable associations between each of the 7 continuous outcomes [(i) seg1, (ii) se1, (iii) cov1, (iv) seg2, (v) se2, (vi) cov2, and (vii) bp] and explanatory variables (study design factors and weather factors). Trial was included in the models as a random effect. Each study design factor and weather factor were tested separately as fixed effects. A

F-test was used to compare the model fit with the fixed predictor to the model fit with only random effects. Similarly, mixed effects logistic regression was performed on the "processed" dataset using the glmer function in the "lme4" package (*4*) to determine the univariable associations between each of the binary outcomes [(i) die-off pattern, and (ii) FSMA compliance] and explanatory variables (study design factors and weather factors). Trial was included in the models as a random effect. Each study design factor and weather factor were tested separately as fixed effects.

Multivariable analyses. Following both the linear and logistic regression analyses, weather factors with a p-value of less than 0.1 in univariable analysis were included in principle component analysis (PCA) for variable reduction. PCA was performed using the prcomp function, such that the number of components retained must explain ≥90% of the variation in the data and each retained variable can only have a major loading on 1 principle component. PCA was performed separately for each outcome in Table 5.3 (because each outcome is associated with different predictors in univariable analysis). If an outcome had 4 or fewer weather predictors with $P<0.1$, PCA was not performed. One representative variable at a time for each principle component was included in multivariable analysis.

Mixed effects linear regression was performed using the lmer function in the "lme4" package (*4*) for multivariable analysis for each of the 5 continuous outcomes: seg1, se1, seg2, se2, and bp (Table 6.3). Trial was included in the models as a random effect. All study design factors significant at a 10% level in univariable analysis were first included in the multivariable models. Backwards selection was performed by removing the variable with the lowest t-value. The final base model was selected as the simplest model that was significantly different ($p<0.05$) from the next simplest nested model using a F-test. Once the base model was developed, representative weather variables for each principle component were included in the base model and backwards

selection was performed as described above to determine the final model. Separate models were fit with 96h, 24h, and 8h weather variables due to the strong correlation between weather variables across the time points. The corresponding regression models are referred to as the 96h regression model, 24h regression model, and 8h regression model to distinguish the time frames over which the weather variables were calculated. Model fit was also assessed using the protocol described by Beauvais et al. (*5*) to check for normality of the residuals, heteroskedasticity, and a linear relationship between the exposure and outcome.

Similarly, mixed effects logistic regression was performed using the glmer function in the "lme4" package (*4*) for multivariable analysis of each of the following binary outcomes: (i) die-off pattern and (ii) FSMA compliance (Table 6.3). Trial was included in the models as a random effect. All study design factors significant at a 10% level in univariable analysis were first included in the multivariable models. Backwards selection was performed by removing variables with the highest p-value. The final base model was selected as the simplest model significantly different ($p<0.05$) from the next simplest nested model using a likelihood ratio test. Once the base model was developed, representative weather predictors for each principle component were included in the base model and backwards selection was performed as described above to determine the final model. Model fit was also assessed using the protocol described by Beauvais et al. (*5*) to determine if the assumption of linearity between the exposure and log odds of the outcome was met.

Classification and regression trees (CART) were fit for each of the seven outcomes (Table 6.3) using the rpart function in the "rpart" package (*61*) to visualize possible interactions and aid interpretation of regression analysis. Tree pruning was performed such that the complexity parameter was set to minimize the cross-validation relative error to prevent overfitting in the trees; a 10-fold cross-validation was used.

Classification and regression trees cannot control for clustering of plots within a trial, however, subsetting as part of the performed 10-fold cross-validation is expected to have reduced the effect of pseudo-replication at the trial level. As an internal validation step, the predictions from the classification trees are described in terms of sensitivity, specificity and negative and positive predictive values. Only 96h weather variables and study factors important by univariable regression analysis (Table S4) were tested for inclusion in the classification and regression trees. These are referred to as the 96h classification trees and 96h regression trees.

Identification of inoculation strains. To assess differential survival of inoculation strains over the 4-day experiment, the following analyses were conducted at the sample level. To determine the effect of trial, time, and produce type on the survival of the 3 *E. coli* inoculum strains, multinomial regression was performed using the multinom function in the "nnet" package (*69*). To determine the effect of trial, time, and produce type on the survival of the 2 *Salmonella* inoculum strains, mixed effects logistic regression was performed using the glmer function (*4*); sample id (i.e., the sample the isolate was taken from) was included in the model as a random effect. Plot id was included as a random effect instead of sample id in the Spain *Salmonella* model, as the sample each isolate came from was not recorded. For both the *E. coli* and *Salmonella* strain identification protocols, isolates were taken from both enumerated samples and enriched samples. However, all isolates from an enriched sample are likely to have propagated from the same cell or same couple of cells. To account for this, all models were weighted such that all isolates from non-enriched samples were given a weight of 1 and all isolates from enriched samples were given a weight of 1 divided by the number of isolates tested from that sample. Weighting was not performed in Spain *Salmonella* models because it was unknown which isolates came from enumerated vs. enriched plates; as such, these results may be biased. Additionally, all strain

identification analyses were performed separately for each location to account for differences in sample sizes and sampling times. For *E. coli*, N=1,920 isolates from California, N=4,700 isolates from New York, and N=1,313 isolates from Spain were tested. For *Salmonella*, N=1,910 isolates from California, N=640 isolates from New York, and N=640 isolates from Spain were tested.

Unless otherwise stated, statistical significance was evaluated at the 5% level. Correction for multiple testing was not applied due to the foundational nature of the study.

Data availability. All data used in this study can be found at https://github.com/abelias/die_off.

RESULTS

Descriptions of samples and weather. In total, 5,252 sample-level data points were collected. However, only 4,900 of these sample-level data points were used in statistical analyses due to crop loss in some trials (Table 6.1). Of those 4,900 data points used in analyses, 1,260 were *E. coli* on lettuce, 1,260 were *Salmonella* on lettuce, 1,190 were *E. coli* on spinach, and 1,190 were *Salmonella* on spinach. Of the 4,900 data points, 1,680, 1,680, and 1,540 data points were from California, New York, and Spain trials, respectively. Summary statistics of the weather variables across all locations can be found in Tables S1-S3.

Descriptions of microbial counts from field data. The levels of *Salmonella* and *E. coli* in the inoculum ranged from 3.68- 5.84 $\log_{10}$ CFU/mL and 3.77- 5.84 $\log_{10}$ CFU/mL in all trials, respectively. Due to the variations in inoculum microbial population levels across trials, the association between 0h population levels on the produce and die-off rate was assessed, as the 0h population level described the effectively applied inoculum concentration levels. The inoculum population levels were

not used directly due to the possibility of slight differences in inoculation rate achieved with the backpack sprayer across locations or trials, even though a standard protocol was used. To do so, each 0h population level was categorized as low (i.e., <4.8 $\log_{10}$ CFU/100g produce) or high (i.e., ≥4.8 $\log_{10}$ CFU/100g produce); the cut-off value was set so there was an approx. equal number of data points in the low and high starting population level categories. *Salmonella* and *E. coli* were modeled separately. There was no significant difference in die-off rates between the starting population level groups for *Salmonella* (*P*=0.374) or *E. coli* (*P*=0.678). The range in total *E. coli* reduction over 4 days was 3.48- 4.40 $\log_{10}$ CFU/100g produce in California trials, 2.29- 4.21 $\log_{10}$ CFU/100g produce in New York trials, and 2.63- 4.97 $\log_{10}$ CFU/100g produce in Spain trials (Figure S1).

In trial NY3, microbial die-off on two varieties of spinach were compared: Harris Seeds Seaside F1 (new variety) and Enza Zaden Acadia F1 (original variety); the original variety, which was used in California and Spain trials, showed poor stand germination under New York conditions. No significant differences in the *E. coli* and *Salmonella* counts via a Wilcoxon rank sum test (p=0.801) or log-linear microbial die-off (p=0.988) were observed between the two spinach varieties. As such, the new variety of spinach was utilized in the remainder of New York trials and data from the new variety of spinach was combined with that of the original spinach variety for all analyses.

Several control samples (i.e., water used for irrigation, soil, or produce prior to inoculation) were presumptively positive (i.e., blue colonies on ECC+R or mauve colonies on SC+R for *E. coli* and *Salmonella*, respectively) for the presence of naturally occurring rifampicin resistant *E. coli* and *Salmonella* in New York trials. In trial NY3, naturally occurring rifampicin resistant *E. coli* and *Salmonella* were isolated from 1/1 irrigation water samples used in spinach growth. However, no further testing was

performed, as no colonies indicative of *E. coli* or *Salmonella* were isolated from soil or produce samples. In trial NY1, naturally occurring rifampicin resistant *E. coli* was identified in 3/6 soil samples and naturally occurring rifampicin resistant *Salmonella* was found in 2/6 soil samples and in 1/1 irrigation water samples. No further testing was performed, as *Salmonella* was not isolated from the produce samples. However, 1/1 spinach sample was positive for rifampicin resistant *E. coli*. To determine if the contamination of spinach samples altered the population levels of *E. coli* on experimental spinach samples, the *E. coli* strain identification PCR protocol, followed by PCR and subsequent sequencing of the *clpX* gene, was performed on 24 randomly selected (i.e., using a random number generator) *E. coli* isolates from produce samples in this trial. PCR and subsequent sequencing on the *clpX* gene were performed as previously described by Walk et al. (*70*) and Weller et al. (*72*). If the *E. coli* strain identification PCR banding pattern and *clpX* allelic type of the 24 isolates did not match one of the inoculum strains, there was 95% confidence the percentage of colonies of naturally occurring rifampicin resistant *E. coli* was below 12%. This threshold was selected as there is indication that the uncertainty threshold of a typical plate count is approx. 12% (*39*). Zero of the 24 isolates tested from spinach samples in trial NY1 matched the *E. coli* strain identification PCR banding pattern and *clpX* allelic type of the inoculum strains. This indicated the presence of naturally occurring rifampicin resistant *E. coli* on spinach did not impact the results of the plate counts by more than would be expected in the typical uncertainty of the method at a 95% confidence level. In trial NY2, naturally occurring rifampicin resistant *E. coli* and *Salmonella* were isolated from the 1 irrigation water sample. Naturally occurring rifampicin resistant *Salmonella* was also isolated from 6/6 soil samples. To account for the presence of rifampicin resistant *E. coli* on the spinach sample collected in trial NY1, enumeration of naturally occurring *E. coli* and *Salmonella* in 1 spinach and 1 lettuce sample was also

performed in trial NY3 and NY4. Both produce samples were negative for naturally occurring rifampicin resistant *E. coli* and *Salmonella* by enumeration, with a limit of detection of 0.04 CFU per gram of produce. No control samples were positive for rifampicin resistant *E. coli* or *Salmonella* in trial NY4, California, or Spain.

Variation in microbial die-off across trials. While the FSMA die-off rate (i.e., -0.5 $\log_{10}$ die-off/day) assumes a log- linear die-off pattern of foodborne pathogens and indicator organisms on in-field fresh produce, visual examination of the data and results from previous studies (*47, 72*) indicated a biphasic pattern may be more appropriate in most cases. As such, initially both segmented log-linear and log-linear models were fit to the data from each trial (data for both produce types and bacteria were combined in these trial level models, Figures 6.2-6.3). The overall log-linear die-off rate across all trials, produce types, and bacteria was -0.60 (95% C1: -0.63, -0.58) $\log_{10}$/day. However, when comparing the log-linear die-off among trials, there was substantial variation in the die-off rate, which ranged from -0.20 (95% CI: -0.27, -0.13) $\log_{10}$/day in trial NY1 to -1.01 (95% CI: -1.06, -0.95) $\log_{10}$/day in trial SP1 (Figure 6.2). The overall segmented log-linear die-off rate across all trials was -4.41 (95% CI: -4.69, -4.12) $\log_{10}$/day for 0-10h and -0.3 (95% CI: -0.33, -0.26) $\log_{10}$/day for 10-96h. When comparing die-off among trials, the segment 1 die-off rate ranged from -0.46 (95% CI: -0.52, -0.41) $\log_{10}$/day in trial CA2 to -6.99 (95% CI: -7.38, -6.59) $\log_{10}$/ day in trial CA1. The breakpoints ranged from 2.5h in trial NY1 to 3.48d in trial NY2, with 9 out of 11 trials having a breakpoint at 12h or earlier. The segment 2 die-off rate was less variable than the segment 1 die-off rate, in that it ranged from 0.28 (95% CI: -0.20, 0.77) $\log_{10}$/day in trial NY2 to -1.00 (95% CI: -1.16, -0.85) $\log_{10}$/day in trial CA2 (Figure 6.3). The die-off rates on the plot level followed the same overall trend as those on the trial level, however, there was substantial variation in die-off between plots from a single trial in some cases (Table 6.4). As such, the remainder of the analyses were conducted on the

plot level to account for this variation; trial was included as a random affect.

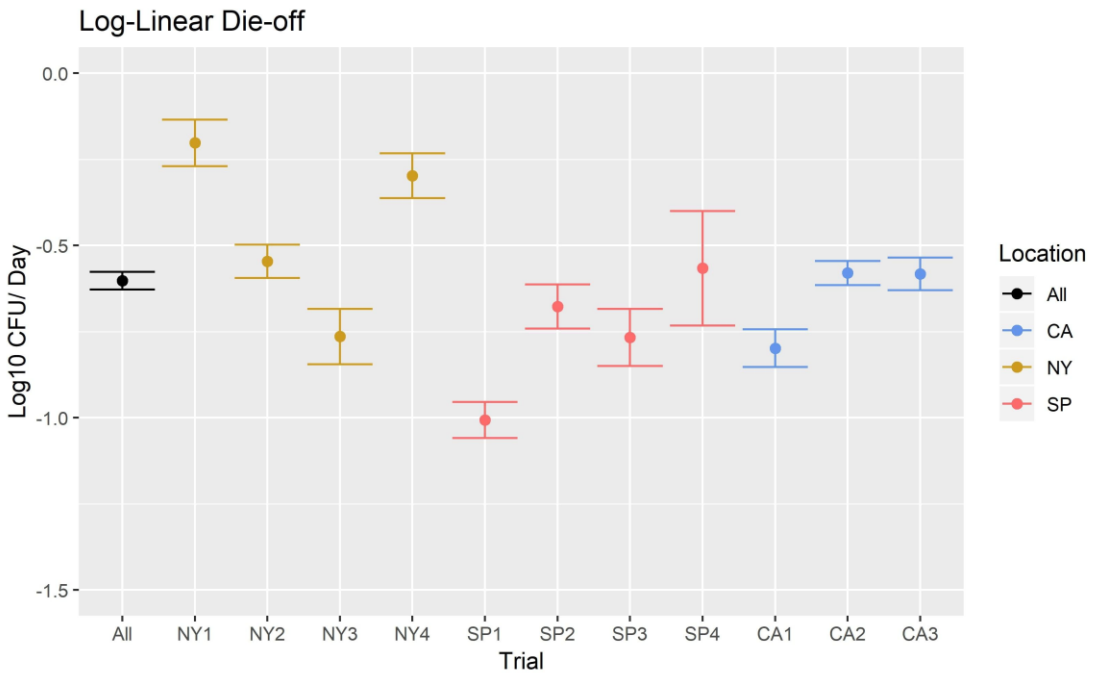


Figure 6.2. Log-linear die-off by trial expressed as $\log_{10}$ CFU die-off/day. Black point indicates the mean die-off for data across all trials, yellow points indicate die-off for New York trials, blue points indicate die-off for California trials, and pink points indicate die-off for Spain trials. Error bars represent the 95% confidence interval for the mean die-off rate for the corresponding trial(s). Calculation of the die-off rates at the trial-level shown here was conducted on all data from a trial (i.e., both produce types and bacteria and across all plots combined) to allow visual examination of data; all further analyses were performed on the plot-level to better represent variations in die-off rates within trials.

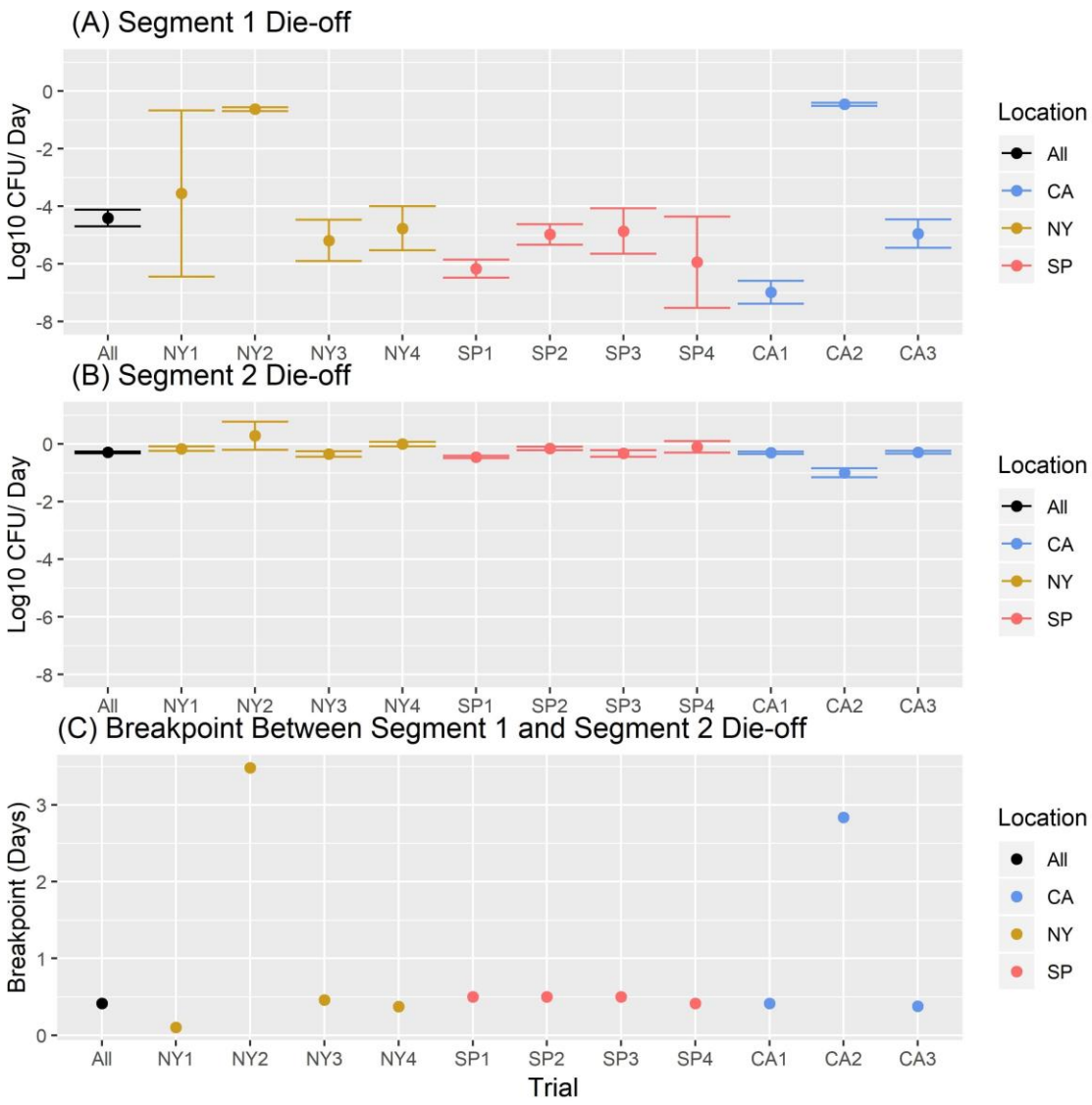


Figure 6.3. Segmented log-linear die-off by trial. (A) Segment 1 die-off rate in $\log_{10}$ CFU die-off/day, (B) Segment 2 die-off rate in $\log_{10}$ CFU die-off/day, and (C) breakpoint between segment 1 and segment 2 die-off. In plots A and B, black points indicate the mean die-off rate for data across all trials, yellow points indicate die-off for New York trials, blue points indicate die-off for California trials, and pink points indicate die-off for Spain trials. Error bars for segment 1 and segment 2 die-off rate represent the 95% confidence intervals for the mean die-off rates from the corresponding trial(s). Calculation of the die-off rates at the trial-level shown here was conducted on all data from a trial (i.e., both produce types and bacteria and across all

plots combined) to allow visual examination of data; all further analyses were performed on the plot-level to better represent variations in die-off rates within trials.

Table 6.4. Summary statistics for die-off outcome variables from the log-linear and segmented log-linear regression, separated by produce type (i.e., spinach and lettuce) and bacteria (i.e., *Salmonella* and *E. coli*). Die-off rates were calculated on the plot level.

Category	Variable[a]	Minimum	Q1[b]	Median	Q3[b]	Maximum	Mean	SD[c]
E. coli, spinach	Linear Die-off	-1.16	-0.93	-0.80	-0.62	-0.07	-0.72	0.29
	Linear SE	0.04	0.06	0.09	0.12	0.14	0.09	0.03
	seg1	-10.42	-7.32	-4.93	-3.45	-0.14	-5.07	2.80
	se1	0.07	0.48	0.64	0.89	2.48	0.69	0.46
	seg2	-0.75	-0.56	-0.44	-0.17	15.80	0.13	2.79
	se2	0.03	0.06	0.09	0.12	3.45	0.21	0.58
	Bp	0.17	0.38	0.45	0.54	3.92	0.77	0.95
E. coli, lettuce	Linear Die-off	-1.04	-0.94	-0.79	-0.67	-0.33	-0.77	0.21
	Linear SE	0.04	0.09	0.11	0.13	0.20	0.11	0.04
	seg1	-16.52	-9.04	-6.66	-5.52	-0.47	-7.07	3.41
	se1	0.06	0.48	0.66	1.57	4.12	1.07	0.94
	seg2	-1.94	-0.55	-0.22	-0.13	3.04	-0.24	0.70
	se2	0.03	0.06	0.09	0.13	0.95	0.14	0.18
	Bp	0.11	0.25	0.38	0.48	3.71	0.68	0.98
Salmonella, spinach	Linear Die-off	-1.04	-0.72	-0.56	-0.15	0.40	-0.45	0.38
	Linear SE	0.04	0.07	0.08	0.09	0.14	0.08	0.02
	seg1	-7.52	-3.88	-2.77	-0.18	0.97	-2.37	2.10
	se1	0.04	0.25	0.36	0.54	1.41	0.41	0.29
	seg2	-1.16	-0.40	-0.28	-0.07	1.81	-0.20	0.55
	se2	0.05	0.08	0.10	0.13	0.62	0.12	0.09
	Bp	0.21	0.50	0.87	1.20	3.56	1.09	0.86
Salmonella, lettuce	Linear Die-off	-1.00	-0.70	-0.56	-0.42	0.07	-0.55	0.24
	Linear SE	0.03	0.06	0.09	0.11	0.13	0.08	0.03
	seg1	-9.70	-5.32	-4.63	-0.93	0.33	-3.71	2.62

se1	0.05	0.24	0.43	0.72	4.32	0.64	0.78
seg2	-7.30	-0.43	-0.25	-0.08	0.45	-0.48	1.21
seg2	0.04	0.06	0.08	0.11	2.56	0.16	0.42
Bp	0.11	0.45	0.52	1.00	3.89	0.86	0.84

[a]seg1 = segment 1 die-off rate (log_{10} die-off/ day), se1 = segment 1 die-off rate standard error (log_{10} die-off/ day), seg2 = segment 2 die-off rate (log_{10} die-off/ day), se2 = segment 2 die-off rate standard error (log_{10} die-off/ day), bp= breakpoint between segment 1 and segment 2 (days). Linear Die-off and Linear SE denote the log-linear die-off rate (log_{10} die-off/ day) and the log-linear die-off rate standard error (log_{10} die-off/ day).

[b]Q1: 1st quartile, or 25% of observations are below and 75% of observations are above this value; Q3: 3rd quartile, or 75% of observations are below and 25% of observations are above this value.

[c]SD: standard deviation in the respective variable across all plots.

Associations between the die-off pattern outcome and explanatory variables: study design factors (i.e., produce type, bacteria, and location) and weather. Log-linear and segmented biphasic log-linear regression models were fit to the data from each plot (e.g., *Salmonella* population levels from the second lettuce plot in the first trial in Spain or *E. coli* population levels from the fourth spinach plot in trial NY4). Model fits were compared for each plot to determine superior fit. Based on the threshold of ≥10 larger BIC score for a segmented model to be considered superior, there was log-linear die-off on 33 plots (24%) and a segmented log-linear die-off on 107 plots (76%). Results of 96h univariable and PCA analyses can be found in Table S4. For the 96h multivariable regression, the only predictive factors retained in the model were average dew point (°C) and relative humidity range (%). A 1°C increase in average dew point was associated with a -0.35 (95% CI: -0.37, -0.32) change in log odds of segmented vs. log linear (baseline) die-off pattern and a 1% increase in relative humidity range was associated with a 0.09 (95% CI: 0.08, 0.09) change in log odds of following a segmented die-off distribution. No study design factors were retained in the model (Table 6.5). For the 96h die-off pattern classification tree, only the maximum change in dew point from one hour to the next (i.e., maximum dew point change rate, °C/h) was retained (Figure 6.4). Based on internal validation of this classification tree, the sensitivity, the specificity, positive predictive value, and negative predictive value were 0.64, 0.97, 0.88, and 0.90, respectively.

Table 6.5. Final mixed effects multivariable logistic regression models displaying the relationship of the categorical die-off outcomes (die-off pattern and FSMA compliance) with the study design factors (i.e., produce type, location, and bacteria) and 96h weather factors (i.e., weather factors calculated over the 96h following inoculation). Trial was included in the models as a random effect[a].

Outcome[b]	Factor	Log Odds	95% CI
Die-off Pattern[b]	Intercept	2.02	(1.63, 2.41)
	Avg. Dew Point (^{o}C)	-0.35	(-0.37, -0.32)
	Relative Humidity Range (%)	0.09	(0.08, 0.09)
FSMA Compliance[b]	Intercept	17.88	(17.06, 18.70)
	Produce Type (Spinach)[c]	-1.63	(-1.74, -1.51)
	Bacteria (*Salmonella*)[c]	-3.02	(-3.14, -2.89)
	Avg. Relative Humidity (%/10)	-1.88	(-1.98, -1.77)

[a]For the die-off pattern model, the variance and standard deviation for the trial random effect were 1.291 and 1.136, respectively. For the FSMA compliance model, the variance and standard deviation for the trial random effect were 2.343 and 1.531, respectively.

[b]Die-off Pattern = Indicates if a biphasic segmented log-linear fit is superior to a log-linear fit for each plot and bacteria combination. The superior model fit for each plot and bacteria subset was determined such that for the segmented fit to be superior, its Bayesian information criteria (BIC) value must be 10 or more than the BIC value of the log-linear model. FSMA Compliance = If the observed segmented die-off in each bacterium, plot combination is compliant with the Food Safety Modernization Act (FSMA) (i.e., $\geq$2 $\log_{10}$ overall die-off from 0h to 96h).

[c]The baseline for produce type is lettuce and the baseline for Bacteria is *E. coli*.

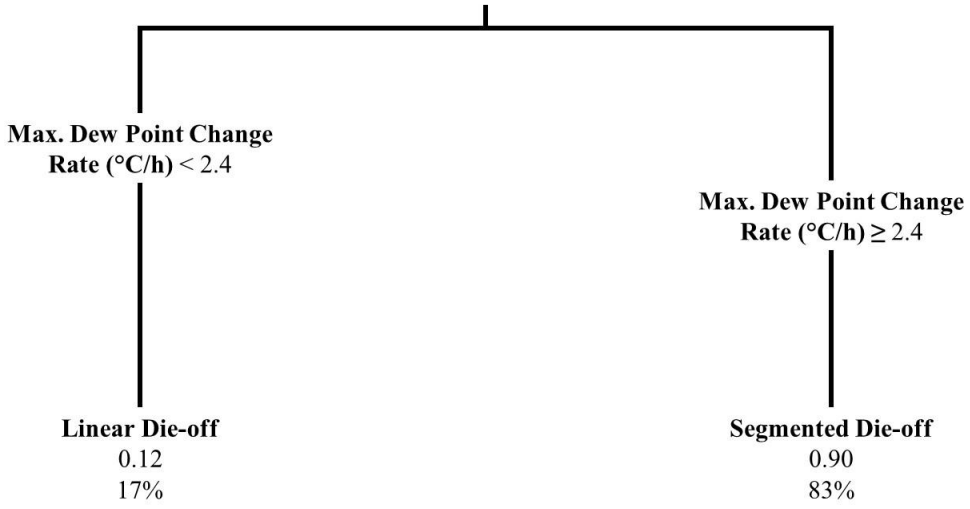


Figure 6.4. Classification tree displaying the relationship between the die-off pattern outcome (i.e., "Linear Die-off" vs. "Segmented Die-off" denoting the best fit of the log-linear vs. segmented log-linear model, respectively) and max. dew point change rate (i.e., maximum change in dew point from one hour to the next °C/h) for the experimental plots (N=140 representing both produce types and bacteria). The superior model fit for each plot and bacteria subset was determined such that for the segmented fit to be superior, its BIC (Bayesian Information Criterion) value must be 10 or more than the BIC value of the log-linear model. The classification tree was fit using the rpart function in R; tree pruning was performed to avoid overfitting. At the end of each terminal node, the superior die-off pattern is designated. The first number below the designated die-off pattern indicates the probability the segmented log-linear model is superior, and the second number indicates the percentage of plots that fall in that node.

Associations between outcomes describing segmented die-off, study design factors and weather. Associations between segmented model outcomes and predictors describing study design and weather were assessed. We hypothesized the associations between study design factors, weather, and individual segmented die-off outcomes would differ, so each of the following outcomes were modeled separately: (i) segment 1 die-off rate (seg1), (ii) segment 1 die-off rate standard error (se1), (iii) segment 2 die-off rate (seg2), (iv) segment 2 die-off rate standard error (se2), (v) breakpoint between segment 1 and segment 2 (bp; Table 6.3). For each of these statistics, we report the outcomes as interquartile ranges (IQR; i.e., the middle 50% of observations), because the interquartile ranges were less impacted by outliers compared to the standard deviation; the mean and standard deviation are reported in Table 6.4. Linear die-off rate and linear die-off rate standard error are also reported in Table 6.4. Similarly, we report all die-off rates per day rather than per hour to allow comparison of results within the study and to available literature. However, for plots where a steep seg1 die-off occurred over a few hours before the bp (breakpoint) and leveling off in seg2, the die-off rate represents the rate that would occur over a whole day. A die-off rate expressed per day can easily be converted to a die-off per hour to further aid interpretation. For instance, a seg1 of -10 $\log_{10}$/day is equivalent to -10 $\log_{10}$/ 24h = -0.41 $\log_{10}$/h. The results of 96h, 24h, and 8h univariate analysis can be found in Tables S4, S5, and S6, respectively.

The IQR of seg1 ($\log_{10}$ change/day) was from -7.32 to -3.45 for *E. coli* on spinach, -9.04 to -5.52 for *E. coli* on lettuce, -3.88 to -0.18 for *Salmonella* on spinach, and -5.32 to -0.93 for *Salmonella* on lettuce (Table 6.4). While there was seg1 die-off in the majority of plots, there was seg1 growth in 7.9% (11/140) of plots. For 96h multivariable regression of seg1, produce type, bacteria, and relative humidity range (%) were retained in the model, such that *Salmonella* had slower seg1 die-off rates compared to *E. coli*, seg1 die-off on spinach was slower than on lettuce, and as the

relative humidity range increased seg1 die-off was faster (Table 6.6, Figure S2). In addition to 96h regression models, 24h and 8h models were also fit (Tables S7-S8). In the 8h and 24h models, produce type and bacteria were retained, similar to the 96h model. However, minimum relative humidity (%) and average relative humidity (%) were retained in the 24h and 8h models, respectively, in place of relative humidity range. Relative humidity range, minimum relative humidity, and average relative humidity are strongly correlated, indicating the 96h, 24h and 8h models are similar, and that humidity during the first 8 or 24 hours after irrigation may be able to predict die-off and the necessary irrigation-to-harvest interval. Additionally, bacteria and minimum relative humidity were retained in the 96h seg1 regression tree (Figure 6.5). Therefore, a similar pattern was seen between the results of multivariable regression and the regression tree for seg1 die-off rate.

Table 6.6. Final mixed effects multivariable linear regression models displaying the relationship of the continuous segmented die-off outcomes (seg1, se1, seg2, se2, and bp) with the study design factors (i.e., produce type, location, and bacteria) and 96h weather factors (i.e., weather factors calculated over the 96h following inoculation). Trial was included in the models as a random effect[a].

Outcome[b]	Factor	Coefficient[c]	95% CI[d]
seg1	Intercept	-1.25	(-3.34, 0.87)
	Produce Type (Spinach)[e]	1.77	(1.09, 2.43)
	Bacteria (*Salmonella*)[e]	3.04	(2.40, 3.68)
	Relative Humidity Range (%)	-0.11	(-0.15, -0.07)
se1	Intercept	-0.28	(-0.36, -0.19)
	Produce Type (Spinach)[e]	-0.37	(-0.39, -0.36)
	Bacteria (*Salmonella*)[e]	-0.35	(-0.37, -0.34)
	Max. Temperature (^{o}C)	0.05	(0.05, 0.06)
seg2	Intercept	6.22	(5.88, 6.56)
	Max. Relative Humidity (%)	-0.06	(-0.06, -0.05)
	Max. Relative Humidity Change Rate (%/h)[f]	-0.06	(-0.06, -0.05)
se2	Intercept	0.43	(0.41, 0.45)
	Relative Humidity Range (%)	-0.01	(-0.01, 0.00)
bp	Intercept	2.22	(2.16, 2.28)
	Bacteria (*Salmonella*)[e]	0.25	(0.23, 0.27)
	Relative Humidity Range (%)	-0.03	(-0.03, -0.03)

[a]For the segment 1 die-off rate model, the residual variance and intercept for the random effects are 3.713 and 0.899, respectively. For the segment 1 die-off rate standard error model, the residual variance and intercept for the random effects are 0.331 and 0.093, respectively. For the segment 2 die-off rate model, the residual variance and intercept for the random effects are 2.245 and 0.015, respectively. For the segment 2 die-off rate standard error model, the residual variance and intercept for the random effects are 0.123 and 0.006, respectively. For the breakpoint model, the

residual variance and intercept for the random effects are 0.543 and 0.078, respectively.

[b]seg1 = segment 1 die-off rate (log_{10} die-off/ day), se1 = segment 1 die-off rate standard error (log_{10} die-off/ day), seg2 = segment 2 die-off rate (log_{10} die-off/ day), se2 = segment 2 die-off rate standard error (log_{10} die-off/ day), bp= breakpoint between segment 1 and segment 2 (days).

[c]Coefficients were estimated using multivariable mixed effects linear regression via the lmer function in R.

[d]95% CI indicates a 95% confidence interval.

[e]The baseline produce type is lettuce and the baseline bacteria is *E. coli*.

[f] Maximum change in relative humidity from one hour to the next

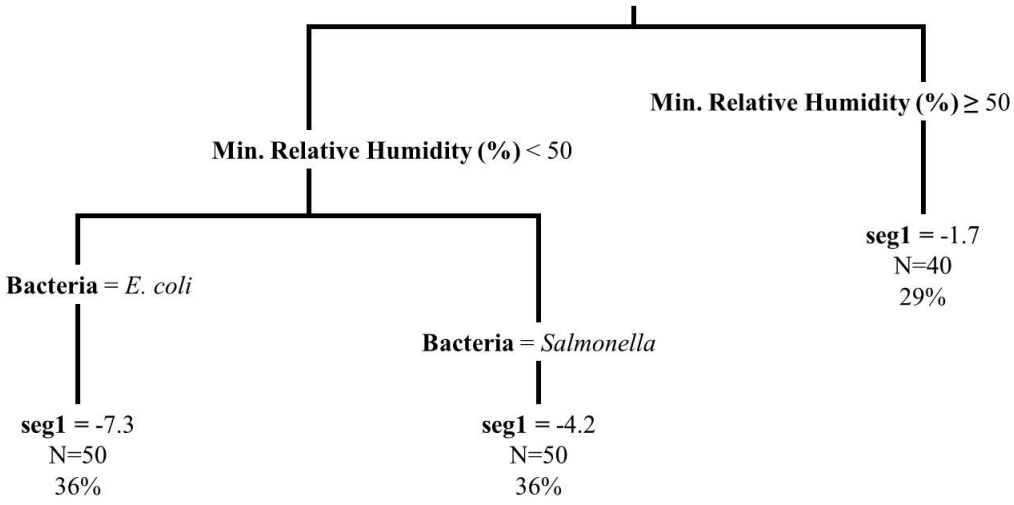


Figure 6.5. Regression tree displaying the relationship of the segment 1 die-off rate (seg1, $\log_{10}$ CFU/day) outcome with bacteria and min. relative humidity (%) for the experimental plots (N=140 representing both produce types and bacteria). The regression tree was fit using the rpart function in R; tree pruning was performed to avoid overfitting. The first number listed at each terminal node is the mean segment 1 die-off rate ($\log_{10}$ CFU/day) for that node, the next number (i.e., N=) designates the number of plots that fall in that node, and the final number designates the percentage of plots that fall in that node.

The IQR of se1 (log_{10} change/day) was from 0.48 to 0.89 for *E. coli* on spinach, 0.48 to 1.57 for *E. coli* on lettuce, 0.25 to 0.54 for *Salmonella* on spinach and 0.24 to 0.72 for *Salmonella* on lettuce (Table 6.4). For 96h multivariable regression of the se1, produce type, bacteria, and maximum temperature (°C) were retained in the model (Table 6.6, Figure S3). The model showed there was a greater se1 for *E. coli* compared to *Salmonella*, there was a greater se1 for lettuce compared to spinach, and as the maximum temperature increased the se1 increased. The same variables (i.e., produce type, bacteria, and maximum temperature) were retained in the 8h and 24h multivariable models (Tables S7-S8). For the 96h se1 regression tree, maximum dew point (°C) and produce type were retained in the model (Figure 6.6). However, maximum dew point and maximum temperature were strongly correlated (Spearman's rank coefficient = 0.80), again providing support for similarity between results of the two modeling approaches. We hypothesized the increase in variation in die-off was due to more stressful conditions that occur at higher temperatures or dew points. In light of such conditions, an increased variability among die-off rates seems logical, assuming variable resistance across sub-populations of *E. coli* or *Salmonella* present on the spinach or lettuce.

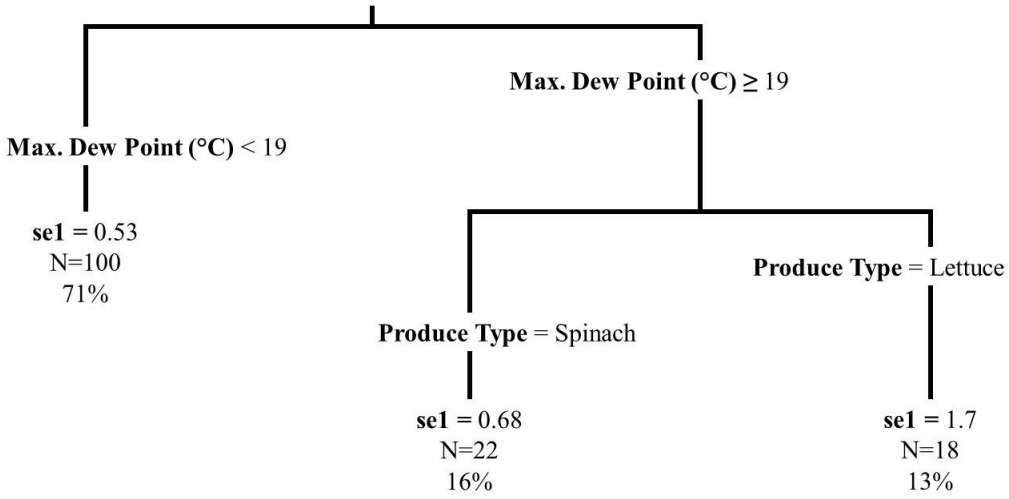


Figure 6.6. Regression tree displaying the relationship of the segment 1 die-off rate standard error (se1, $\log_{10}$ CFU/day) outcome with max. dew point (°C) and produce type for the experimental plots (N=140 representing both produce types and bacteria). The regression tree was fit using the rpart function in R; tree pruning was performed to avoid overfitting. The first number listed at each terminal node is the mean segment 1 die-off rate standard error ($\log_{10}$ CFU/day) for that node, the next number (i.e., N=) designates the number of plots that fall in that node, and the final number designates the percentage of plots that fall in that node.

The IQR of seg2 (log_{10} change/day) was from -0.56 to -0.17 for *E. coli* on spinach, -0.55 to -0.13 for *E. coli* on lettuce, -0.40 to -0.07 for *Salmonella* on spinach, -0.43 to -0.08 for *Salmonella* on lettuce (Table 6.4). While there was seg2 die-off in the majority of plots, there was seg2 growth in 16.4% (23/140) of plots. For 96h seg2 multivariable regression, maximum relative humidity (%) and maximum relative humidity change rate (%/h) were retained in the model (Table 6.6). The 24h and 8h seg2 regression models do not match that of the 96h regression model. For the 24h seg2 model, temperature range (°C) and maximum wind speed (m/s) were retained in the model (Table S7). For the 8h seg2 model, maximum relative humidity change rate (%/h) and average wind speed (m/s) were retained in the model (Table S8). However, due to the small variation in seg2 die-off rates across plots, the effect size of these variables was small. Additionally, no variables were retained in the 96h seg2 regression tree.

The IQR of se2 (log_{10} change/day) was from 0.06 to 0.12 for *E. coli* on spinach, 0.06 to 0.13 for *E. coli* on lettuce, 0.08 to 0.13 for *Salmonella* on spinach, and 0.06 to 0.11 for *Salmonella* on lettuce (Table 6.4). For the 96h multivariable se2 regression model, relative humidity range (%) was retained (Table 6.6). For the 8h and 24h multivariable se2 models, maximum relative humidity change rate (%/h) was retained (Tables S7-S8). However, the effect sizes of the weather variables on se2 were substantially smaller than those for se1. Additionally, no variables were retained in the 96h se2 regression tree.

The IQR of bp (days) was from 0.38 to 0.54 day for *E. coli* on spinach, 0.25 to 0.48 day for *E. coli* on lettuce, 0.50 to 1.20 day for *Salmonella* on spinach, 0.45 to 1.00 day for *Salmonella* on lettuce (Table 6.4). For the 96h bp multivariable regression model, bacteria and relative humidity range were retained in the model, such that *Salmonella* was associated with a later bp compared to *E. coli* and an increase in relative humidity range was associated with an earlier bp (Table 6.6, Figure S4). Minimum

relative humidity and average relative humidity were retained in the 24h (Table S7) and 8h (Table S8) models instead of relative humidity range, respectively, which were the same variables important for seg1 for each timeframe (i.e., 8h, 24h, and 96h). This may indicate that, as conditions become more stressful, the more sensitive sub-population die-off more rapidly and the underlying slow die-off of the more resistant sub-population becomes apparent at an earlier time following inoculation. However, the fit for the bp regression models was poor due to a non-linear relationship (i.e., at low relative humidity ranges the breakpoint appears to follow no pattern, and at higher relative humidity ranges the breakpoint tends to occur earlier; Figure S3). This indicates there were likely additional variables impacting bp. For the 96h bp regression tree, minimum relative humidity, average relative humidity, and bacteria were retained (Figure 6.7). The segment 1 and segment 2 coefficients of variation were also calculated for each plot, and regression models and regression trees were fit. However, they did not differ across plots, so the model results are not discussed.

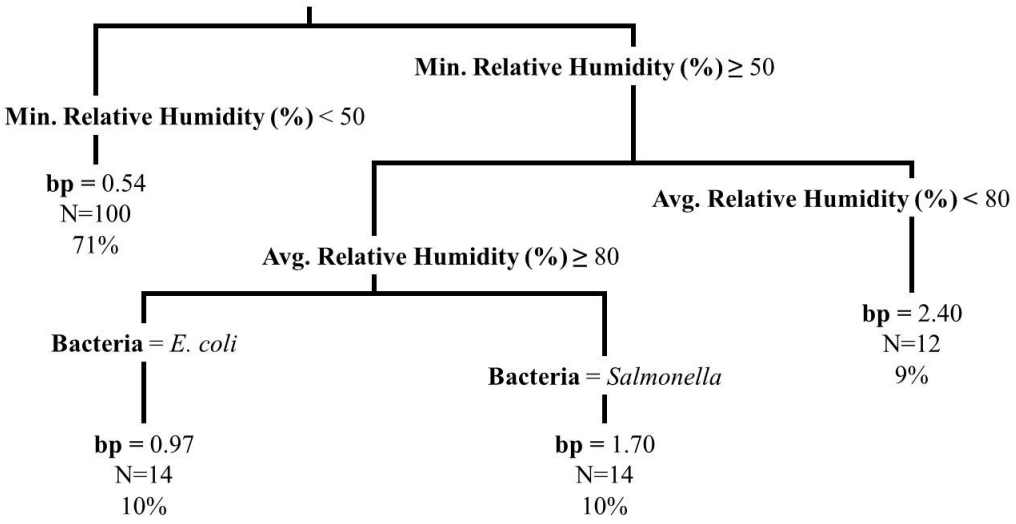


Figure 6.7. Regression tree displaying the relationship of the outcome denoting the breakpoint (bp, days) between segment 1 and segment 2 with min. relative humidity (%), avg. relative humidity, and bacteria for the experimental plots (N=140 representing both produce types and bacteria). The regression tree was fit using the rpart function in R; tree pruning was performed to avoid overfitting. The first number listed at each terminal node is the mean breakpoint (days) for that node, the next number (i.e., N=) designates the number of plots that fall in that node, and the final number designates the percentage of plots that fall in that node.

Additionally, the similarity in weather variables retained and the similar regression coefficients in the 8h, 24h, and 96h regression models for seg1, se1, and bp indicate 8h or 24h weather variables can be used instead of 96h weather variables. This will allow produce growers to use weather data in real-time to plan time-to-harvest intervals.

Associations between the FSMA compliance outcome, study design factors, and weather. FSMA compliance was designated if the segmented die-off rate calculated for each plot would achieve at least a 2 $\log_{10}$ reduction in 4 days (i.e., assumes a 0.5 $\log_{10}$/day reduction as specified in FSMA; Table 6.3). In total, 75% (105/140) of plots were compliant with FSMA. Additionally, 79% (27/34) of *E. coli* on spinach plots, 97% (35/36) of *E. coli* on lettuce plots, 56% of *Salmonella* on spinach plots, and 67% of *Salmonella* on lettuce plots were compliant with FSMA. According to 96h multivariable logistic regression, produce type, bacteria, and average relative humidity (%) were retained (Table 6.5). Spinach was associated with a decrease in log odds of compliance compared to lettuce (P<0.001), *Salmonella* was associated with a decrease in log odds of compliance compared to *E. coli* (P=0.017), and a decrease in average relative humidity was associated with an increase in log odds of compliance (P=0.002). Minimum relative humidity (%) and bacteria were retained in the 96h FSMA compliance classification tree (Figure 6.8). Internal validation indicated the sensitivity, specificity, positive predictive value, and negative predictive value for this classification tree were 0.69, 0.96, 0.86, and 0.90, respectively.

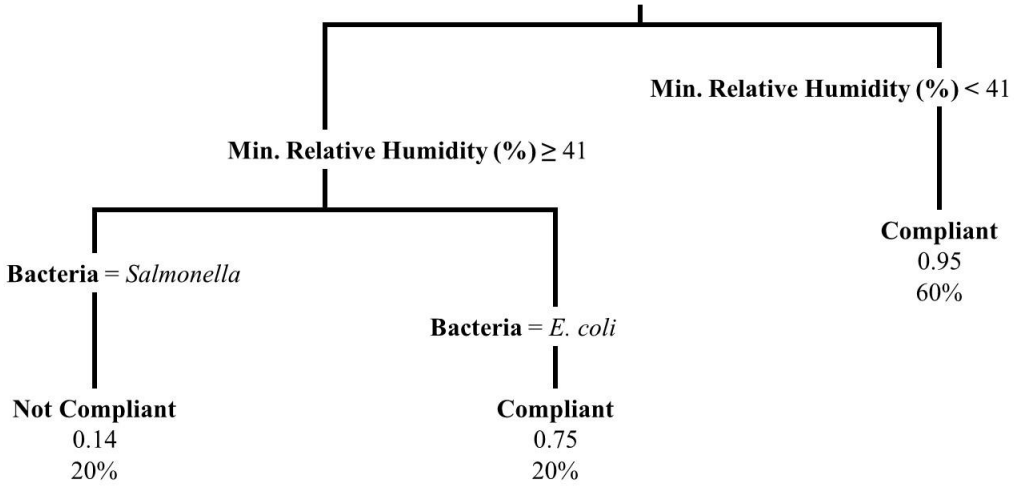


Figure 6.8. Classification tree displaying the relationship of the compliance with FSMA outcome with min. relative humidity (%) and bacteria for the experimental plots (N=140 representing both produce types and bacteria). Compliance was designated if the segmented die-off calculated for an experimental plot would achieve at least a 2 $\log_{10}$ reduction in 4 days (i.e., assumes a 0.5 $\log_{10}$ die-off/day as specified in FSMA). The classification tree was using the rpart function in R; tree pruning was performed to avoid overfitting. At the end of each terminal node, whether the experimental die-off was FSMA compliance is designated. The first number below the FSMA compliance designation is the probability of being compliant and the second number indicates the percentage of plots that fall in that node.

Comparison in the survival of *E. coli* and *Salmonella* inoculum strains. PCR of *E. coli* isolates was performed in each study location to investigate differential strain survival. For *E. coli*, N=1,920 isolates from California, N=4,700 isolates from New York, and N=1,313 isolates from Spain were tested. According to multinomial regression, there were no obvious trends regarding the associations between time, produce type and the survival of the 3 inoculum strains across trials in the 3 study

locations (Tables S9-S11). For instance, in New York, there was a significantly higher odds of isolating TVS 354 compared to TVS 353 as time increased, however, the effect size was small [OR=1.004 (95% CI: 1.002, 1.006)]. There was also a significantly higher odds of isolating strain TVS 355 compared to TVS 353 in trial NY2 (P<0.001), with an OR of 1.858 (95% CI: 1.417, 2.435). There were no other significant differences between strains. Produce type was not retained in the model (Table S9). Similar inconsistencies were seen in California (Table S10) and Spain (Table S11).

Identification of *Salmonella* strains was performed in each study location. For *Salmonella*, N=1,910 isolates from California, N=640 isolates from New York, and N=640 isolates from Spain were tested. According to mixed effects logistic regression for New York, there was an interaction between produce type and time (P<0.001); as such, separate models were fit for spinach and lettuce for interpretation of results. For New York spinach, the odds of isolating attPTVS 355 compared to attPTVS 337 significantly decreased with time [OR=0.973 (95% CI: 0.957, 0.987); P<0.001; Table S12]. For lettuce, the odds of isolating attPTVS 355 compared to attPTVS 337 appeared to be stable with time (Table S12). A similar trend was seen in California (Table S13). In comparison, in Spain, there was no significant change in odds of isolating attPTVS 355 compared to attPTVS 337 from spinach samples with time (Table S14). Additionally, on lettuce samples, there was a 17.664 (95% CI: 7.398, 54.225; P<0.001) times greater odds of isolating attPTVS 355 compared to attPTVS 337 regardless of time (Table S14). However, it should be noted that this difference in *Salmonella* strain survival in Spain may be due to the bias from not accounting for isolates coming from enumerated vs. enriched samples in the analysis.

DISCUSSION

The current study quantified the die-off rates of *E. coli* and *Salmonella* on baby

spinach and lettuce under field conditions in three distinct climatic regions: California, New York, and Spain. The differences in die-off between *Salmonella* and *E. coli* and differences between die-off on spinach and lettuce were evaluated. The replication of a standard protocol across the three locations also allowed for a more comprehensive assessment of the associations between weather and microbial die-off. It was found that, in most cases, a log-linear die-off pattern did not fit the data as well as a segmented, biphasic die-off pattern, with a rapid initial decline followed by a period of tailing. Additionally, the initial period of decline (seg1) is associated with relative humidity, produce type, and bacteria, and the same factors also predict if experimental die-off is compliant with the FSMA regulation. In comparison, the period of tailing (seg2) is less affected by weather, produce type, and bacteria; however, the bacterial population levels were still highly variable, indicating caution should be taken when implementing a wait period between non-compliant (according to FSMA) water application and harvest until more data is available to explain the variability. These results can be used by industry to inform food safety programs, by academia to develop risk assessment models, and by government to update regulations.

The assumption of a log-linear die-off pattern is not appropriate in most cases. The FDA FSMA agricultural water standard specifies the use of a wait period between water application and harvest as a possible intervention if agricultural water applied to the harvestable portion of a crop does not meet the standard. The allowed corrective measure assumes a log-linear die-off (0.5 $\log_{10}$ die-off/day) of *E. coli* for a maximum of 4 days. However, the results of the current study showed a biphasic, segmented log-linear die-off pattern was more appropriate than a log-linear die-off pattern in 76% (107/140) of plots. Additionally, the produce plots in 6 out of the 11 trials did not follow the same die-off pattern (i.e., some of the plots followed a segmented log-linear pattern and some of the plots followed a log-linear die-off pattern).

Previous studies have also indicated a log-linear die-off pattern is not appropriate for foodborne pathogens, indicator organisms, and surrogates on in-field produce (*2, 6, 17, 18, 24, 25, 28, 47, 49, 50, 72*). For instance, in a study conducted by McKellar et al. (*47*), which fit log-linear, Cerf, and Weibull regression models to *E. coli* O157:H7 die-off data from several different experimental trials, found the biphasic models (i.e., Cerf and Weibull) best fit the data in most cases with an initial period of rapid decline followed by a period of tailing. Similar patterns were also seen in several other studies (*6, 24, 25, 28, 49, 50, 53, 72*). Brouwer et al. (*7*), summarized possible mechanisms of this biphasic pattern, which include heterogeneity in the hardiness of the inoculum population, hardening off (i.e., differential gene expression following exposure to harsh environmental conditions), cells entering the viable but not culturable state, and the effect of cell density. However, additional information is needed to better understand the underlying mechanism which caused the biphasic die-off pattern observed in the current study.

In the current study, there were also some plots with periods of growth (7.9% for segment 1, 16.4% for segment 2; N=140). Chase et al. (*17, 18*) also saw periods of growth in their studies, which investigated the population of *E. coli* O157:H7 on lettuce; they hypothesized this growth was caused by an increase in moisture on the plants. If the possibility of growth is not accounted for when using time-to-harvest intervals as an intervention, food safety incidents may occur. This growth can be accounted for using predictive models that are able to identify periods of growth based on weather conditions, rather than a one-size-fits-all strategy that only accounts for net microbial death.

The effect of weather on die-off distribution was also assessed. According to 96h regression analysis, a decrease in average dew point and an increase in relative humidity range were associated with occurrence of a segmented compared to a log-

linear pattern. Furthermore, according to the 96h classification tree, at higher maximum dew point change rates (i.e., describing a rapid change in dew point over an hour and thus presumably more stressful conditions for bacteria) there is an increased probability die-off will follow a segmented pattern. The difference in identified risk factors in the two modeling approaches is likely to be due to the different strengths and weaknesses the two methods have. Regardless, both sets of risk factors are plausible, particularly, as dew point was found to play a role in both modeling strategies and the two methods may be identifying different mechanisms of the effect of dew point on die-off. As such, these results support the need to revise the die-off pattern proposed in the FSMA agricultural water standard to account for biphasic decay, the possibility of growth rather than decay following water application, and the effect of weather.

Weather, produce type, and bacteria are differentially associated with segment 1 and 2 die-off. Based on examination of the interquartile ranges of seg1 and seg2 on the plot level, there is substantially greater variability in seg1 (Table 6.3). This indicates seg1 has a greater impact on the plot to plot variability in overall die-off compared to seg2. In addition, the substantially smaller degree of variation in the seg2 from plot to plot indicates it is less impacted by weather and environmental conditions compared to seg1.

According to regression analyses and the 96h regression tree, relative humidity was associated with seg1. Several previous studies have also shown that an increase relative humidity was associated with a decrease in microbial die-off (*19, 40, 43, 48, 72*). Additionally, Wood et al. (*73*) and Moyne et al. (*50*), saw slower initial die-off after inoculation of spinach and lettuce, respectively, with *E. coli* O157:H7 when inoculation was performed at night as compared to in the morning. It was suggested this phenomenon was caused by the higher moisture/ relative humidity typically experienced at night compared to in the morning. The mechanism by which relative

humidity is related to die-off in this study and in previous studies has been referred to as desiccation; the lack of moisture in the air causes a drying effect that is associated with a reduction in the number of bacteria.

Produce variety and bacteria were also retained in the seg1 model, such that the die-off of *Salmonella* and die-off on spinach were slower than that of *E. coli* and lettuce, respectively. While no studies looked at the difference in die-off or survival of these two microorganisms or produce types during the initial decay period (i.e., segment 1), several have assessed the overall survival of these organisms on these produce varieties. For instance, Erikson et al. (*25*) showed *Salmonella* had better survival compared to *E. coli* O157:H7 under growth chamber and field conditions. Lopez-Velasco et al. (*45*), also saw better survival of *Salmonella* compared to *E. coli* on a variety of lettuce cultivars. Additionally, Hutchison et al. (*33*) saw better survival of *Salmonella* on lettuce and spinach compared to *E. coli* O157:H7 (i.e., greater number of positive samples at 2 weeks following inoculation); survival of *Salmonella* on spinach was better than survival of *Salmonella* on lettuce in this study. Stine et al. (*60*), saw greater inactivation rates for *E. coli* and *E. coli* O157:H7 compared to *Salmonella enterica* when sprayed on the surface of lettuce under both dry and humid conditions. However, when investigating this relationship between *E. coli* and *Salmonella* on other produce varieties (i.e., cantaloupe and bell pepper), the results were inconsistent. This indicates there may be interactions between the effects of bacteria and produce type on die-off. As such, additional studies are needed to further understand these relationships. Regardless, the improved survival or slower die-off of *Salmonella* compared to *E. coli* reported in this and previous studies demonstrates, while *E. coli* is used as an indicator of fecal contamination, it likely cannot be used as a surrogate for in field pathogen die-off. However, it is possible to use *E. coli* as a surrogate if the findings are appropriately adjusted for the expected reduction in die-off of *Salmonella* from the data collected in

this and in other studies.

While seg2 regression analysis included weather in the final model, no weather was retained in the final regression tree. Additionally, the small effect sizes of weather in seg2 regression analysis indicate it is not a strong risk factor. The lack of variation in seg2 from plot to plot further supports this point. As such, it appears after seg1 is complete, the bacteria behave similarly regardless of weather or other environmental conditions not explored in the current study (e.g., native microflora). Some potential explanations for this include a phenotypic switch occurs in the bacteria after exposure to the stressors of the farm environment or a more sensitive sub-population has died off during the first segment and the remaining resistant sub-population is less impacted by the stressors of the environment (*2, 7, 51*). As such, this indicates regulations should focus on die-off that occurs during the first segment. However, it should also be noted that, while seg2 was relatively consistent from plot to plot, the microbial counts at each time point within a plot or trial were still highly variable (Figure S1). This may be because the variation in seg1 leads to a highly variable microbial concentration at the start of segment 2. Additionally, it is possible some of this variation is because the majority of the data imputations fell in segment 2. Therefore, if the use of wait time as an intervention strategy is implemented, it should be noted that it is not reliable under all conditions.

As such, due to the associations between weather variables and segmented die-off outcomes, it is likely not appropriate to use a single die-off rate across different locations and seasons (characterized by different weather conditions). Rather, it may be more appropriate to develop predictive models for identifying the optimal wait period between water application and harvest for specific weather conditions to reduce produce contamination at harvest to an acceptable level. Furthermore, these predictive models need to account for both the variability in count data, in addition to the variability of

die-off rates to capture the true risk associated with pathogen die-off.

The FSMA die-off rate should be updated to account for weather. In the majority of cases (75%, N=140), experimental die-off appeared to be compliant with the FSMA die-off rate. According to 96h regression analysis the log odds that *Salmonella* and *E. coli* die-off by at least 0.5 $\log_{10}$/day for 4 days (i.e., complying with FSMA) increased as the average relative humidity decreased. Additionally, *Salmonella* and spinach were associated with a lower log-odds of complying with FSMA, compared to *E. coli* and lettuce, respectively. The variables retained in the FSMA compliance classification tree followed a similar trend. As such, it appears FSMA compliance is related to segment 1 die-off, which aligns with our result that suggests most die-off variation occurs during this segment. This information suggests if a wait period is to be used as an effective intervention strategy for irrigation with contaminated water, weather (i.e., relative humidity), bacteria, and produce type must be accounted for. This can be done using predictive modeling. However, further research is needed to confirm these predictions, as the methods used to model this relationship required extrapolation of die-off to a lower starting inoculum concentration. Additionally, it should be noted that this analysis does not consider the scenario where the rate is too conservative and can lead to decreases in profit for produce growers due to product loss.

Differential survival of the inoculum strains used in this study supports the use of multiple inoculum strains in future pre-harvest studies. According to the results observed in the current study, it appears there were no consistent trends in survival among the *E. coli* strains across trials and locations in the current study. Previous studies have also investigated the survival of these strains in the pre-harvest environment, such that Gutierrez-Rodriguez et al. (*27*) found the greatest survival by TVS 355 and Tomas-Callejas et al. (*62*) found the greatest survival by TVS 353. The lack of consistency in survival of these strains across and within studies could indicate

random variation or that each of the inoculum strains is better equipped for survival under different environmental conditions. Alternatively, these differences could be due to chance (i.e., related to which colonies are picked). However, due to the large number of isolates characterized in the current study, the second hypothesis is less likely. In particular, this indicates all three *E. coli* inoculum strains (TVS 353, TVS 354, and TVS 355) and even other or additional strains should continue to be used in tandem in future preharvest produce studies to better capture the diversity in survival abilities among the strains under varying environmental conditions.

The survival of the *Salmonella* inoculum strains appeared to be different on each of the produce types. This suggests each of these strains is likely more equipped to survive under different conditions. It is likely the difference in survival was due to genetic differences between the wild type parent strains of each attenuated mutant. Additionally, the interaction between time and produce type on the probability of survival of each of the *Salmonella* inoculum strains could be due to differences in the microbiota present on each of the produce varieties, as was indicated by Lopez-Galvez et al. (*43*). As with *E. coli*, the observed differential survival of these *Salmonella* inoculum strains supports their use in tandem in future pre-harvest produce studies.

Limitations. The current study investigated the impact of bacteria, produce type, location, and weather on in-field microbial die-off. However, these statistical models are not able to completely explain microbial die-off. Previous studies have indicated that epiphyte populations (*20, 56*), soil composition (*34*), and composition of irrigation medium (*36, 68, 58*), among others, could represent unmeasured risk factors and lead to confounding bias. While this study assessed the impact of the widest range of weather conditions to date, only 11 weather patterns were included in the analysis to assess the effects of weather die-off rate (i.e., each trial only experiences 1 set of weather conditions). As such, assessment of generalizability of the associations identified in this

study is needed. In addition, it was decided to use a lower starting inoculum concentration to be more representative of real-world conditions. However, as a result, some samples were below the limit of quantification and the true bacterial concentration on the plants was not determined. While some degree of measurement error was possible in appraising microbial counts and weather variables, if true, the resulting information bias would have been non-differential, meaning it would have underestimated the measures of association. The differences in *Salmonella* concentrations detected on paired spread and filter plates indicate the *Salmonella* counts for samples within the countable range may be of poor accuracy. However, this systematic error was detected and corrected by calculating the percent difference between samples with paired spread and filter plates and using the percent difference to adjust the spread plate counts down to where they would be if filter plating was used. There are also several advantages and disadvantages associated with the regression and classification and regression trees used to model the die-off data, which are reviewed by Ivanek et al. (*38*). Regardless, while each method has several limitations associated with their parametric and non-parametric nature, their limitations are complimentary to one another and make them a good pair to use in tandem.

Conclusion. The current study indicates die-off of *Salmonella* and *E. coli* on baby spinach and lettuce follows a segmented log-linear pattern. The die-off rate in the first segment is variable and appears to be associated with relative humidity, produce type, and bacteria. After the breakpoint, the die-off rate is less variable, however, there is still a large variation in the microbial counts at each time point in this segment across experimental plots and trials. Additionally, this study provides evidence that relative humidity can be used to estimate when experimental die-off is compliant with FSMA. As such, the use of a single die-off rate, as proposed by FSMA, is likely not appropriate, and rather the regulation should consider the effect of weather, bacteria, and produce

type on microbial die-off. Furthermore, additional information is needed to evaluate the effectiveness of the wait period intervention in reducing the risk of recalls or illness.

ACKNOWLEDGEMENTS

This study was funded by the Center for Produce Safety (2018CPS04) awarded to RI. We would like to thank Enza Zaden for proving seeds used in this project. We would like to thank Renato Orsi for developing the primers used in *E. coli* strain identification. We would also like to thank Sherry Roof, Steve McKay, Sriya Sunil, Sean Guo, Maureen Gunderson, Alan Bitar, and Wendy Beauvais for their help on various aspects of the project. Technical and scientific support provided by Maria I. Gil, Juan Antonio Tudela, Jenifer Cascales, Silvia Andújar, Macarena Moreno, Natalia Acosta and Vicente Martínez from the CEBAS-CSIC is appreciated. MW served as paid consultant or scientific advisor for Neogen, ThermoFisher, 3M, and Yum Brands.

SUPPLEMENTAL MATERIAL

Supplemental material can be found at: https://aem.asm.org/content/aem/suppl/2020/08/07/AEM.00899-20.DCSupplemental/AEM.00899-20-s0001.pdf

REFERENCES

1. Anderson GB, Bell ML, Peng RD. 2013. Methods to calculate the heat index as an exposure metric in environmental health research. Environ Health Perspect 121(10):1111-1119.

2. Balaban NQ, Merrin J, Chait R, Kowalik L, Leibler S. 2004. Bacterial persistence as a phenotypic switch. Science 305(5690):1622-1625.

3. Barker-Reid F, Harapas D, Engleitner S, Kreidl S, Holmes R, Faggian R. 2009. Persistence of *Escherichia coli* on injured iceberg lettuce in the field, overhead irrigated with contaminated water. J Food Prot 72(3):458-464.

4. Bates D, Maechler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. J Stat Softw 67(1):1-48.

5. Beauvais W, Gart EV, Bean M, Blanco A, Wilsey J, McWhinney K, Bryan L, Krath M, Yang CY, Alvarez DM, Paudyal S, Bryan K, Stewart S, Cook PW, Lahodny G, Baumgarten K, Gautam R, Nightingale K, Lawhon SD, Pinedo P, Ivanek R. 2018. The prevalence of *Escherichia coli* O157:H7 fecal shedding in feedlot pens is affected by the water to-cattle ratio: A randomized controlled trial. PLoS ONE 13(2):e0192149.

6. Bezanson G, Delaquis P, Bach S, McKellar R, Topp E, Gill A, Blais B, Gilmour M. 2012. Comparative examination of *Escherichia coli* O157:H7 survival on romaine lettuce and in soula and two independent experimental sites. J Food Prot 75(3):480-487.

7. Brouwer AF, Eisenberg MC, Remais JV, Collender PA, Meza R, Eisenberg JNS. 2017. Modeling biphasic environmental decay of pathogens and implications for risk analysis. Environ Sci Technol 51:2186-2196.

8. Castro-Ibanez I, Gil MI, Tudela JA, Ivanek R, Allende A. 2015. Assessment of microbial risk factors and impact of meteorological conditions during production of baby spinach in the southeast of Spain. Food Microbiol 49:173-181.

9. Centers for Disease Control and Prevention. 2006. Multistate outbreak of *E. coli* O157:H7 infections linked to fresh spinach (final update). Accessed on December 15, 2019. https://www.cdc.gov/ecoli/2006/spinach-10-2006.html

10. Centers for Disease Control and Prevention. 2018. Multistate outbreak of Shiga toxin-producing *Escherichia coli* O157:H7 infections linked to leafy greens (final update). Accessed on December 15, 2019. https://www.cdc.gov/ecoli/2017/o157h7-12-17/index.html

11. Centers for Disease Control and Prevention. 2018. Multistate outbreak of *E. coli* O157:H7 infections linked to romaine lettuce (final update). Accessed on December 15, 2019. https://www.cdc.gov/ecoli/2018/o157h7-04-18/index.html

12. Centers for Disease Control and Prevention. 2019. Outbreak of *E. coli* infections linked to romaine lettuce. Accessed on December 15, 2019. https://www.cdc.gov/ecoli/2018/o157h7-11-18/index.html

13. Centers for Disease Control and Prevention. 2020. Outbreak of *E. coli* infections linked to Fresh Express sunflower crisp chopped salad kits. Accessed on February 12, 2020. https://www.cdc.gov/ecoli/2019/o157h7-12-19/index.html.

14. Centers for Disease Control and Prevention. 2020. Outbreak of *E. coli* infections linked to romaine lettuce. Accessed on February 12, 2020. https://www.cdc.gov/ecoli/2019/o157h7-11-19/index.html

15. 21 CFR § 112.44. 2019.

16. 21 CFR § 112.45. 2019.

17. Chase JA, Atwill ER, Partyka ML, Bond RF, Oryang D. 2017. Inactivation of *Escherichia coli* O157:H7on romaine lettuce when inoculated in a fecal slurry mix. J Food Prot 80(5): 792-798.

18. Chase JA, Partyka ML, Bond RF, Atwill ER. 2019. Environmental inactivation and irrigation-mediated regrowth of *Escherichia coli* O157:H7 on romaine lettuce when inoculated in a fecal slurry matrix. PeerJ 7:e6591.

19. Choi S, Bang J, Kim H, Beuchat LR, Ryu JH. 2011. Survival and colonization of *Escherichia coli* O157:H7 on spinach leaves as affected by inoculum level and carrier, temperature and relative humidity. J Appl Microbiol 111:1465-1472.

20. Cooley MB, Chao D, Mandrell RE. 2006. *Escherichia coli* O157:H7 survival and growth on lettuce is altered by the presence of epiphytic bacteria. J Food Prot 69(10):2329-2335.

21. Curtiss R III, Kelly SM. 1987. *Salmonella* Typhimurium deletion mutants lacking adenylate cyclase and cyclic AMP receptor protein are avirulent and immunogenic. Infect Immun 55(12):3035-3043.

22. Delaquis P, Bach S, Dinu LD. 2007. Behavior of *Escherichia coli* O157:H7 in leafy vegetables. J Food Prot 70(8):1966-1974.

23. de Moraes MH, Chapin TK, Ginn A, Wright AC, Parker K, Hoffman C, Pascual DW, Danyluk MD, Teplitski M. 2016. Development of an avirulent *Salmonella* surrogate for modeling pathogen behavior in pre- and postharvest environments. Appl Environ Microbiol 82(14):4100-4111.

24. Erickson MC, Webb CC, Diaz-Perez JC, Phatak SC, Silvoy JJ, Davey L, Payton AS, Liao J, Ma L, Doyle MP. 2010. Surface and internalized *Escherichia coli* O157:H7 on field-grown spinach and lettuce treated with spray-contaminated irrigation water. J Food Prot 73(6):1023-1029.

25. Erickson MC, Liao JY, Payton AS, Cook PW, Den Bakker HC, Bautista J, Diaz-Perez JC. 2019. Pre-harvest internalization and surface survival of *Salmonella* and *Escherichia coli* O157:H7 sprayed onto different lettuce cultivars under field grown chamber conditions. Int J Food Microbiol 291:197-204.

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

27. Gutierrez-Rodriguez E, Gunderson A, Sbodio AO, Suslow TV. 2011. Variable agronomic practices, cultivar, strain source and initial contamination dose differentially affect survival of *Escherichia coli* on spinach. J Appl Microbiol 112:109-118.

28. Gutierrez-Rodriguez E, Gunderson 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.

29. Hanning IB, Nutt JD, Ricke SC. 2009. Salmonellosis outbreaks in the United States due to fresh produce: sources and potential intervention methods. Foodborne Pathog Dis 6(6):635-648.

30. Harel O, Perkins N, Schisterman EF. 2014. The use of multiple imputations for data subject to limits of detection. Sri Lankan J Appl Stat 5(4):227-246.

31. 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 andfresh-cut produce. Compr Rev Food Sci Food Saf 2(s1):78–141.

32. Heaton JC, Jones K. 2008. Microbial contamination of fruit and vegetables and the behaviour of enteropathogens in the phyllosphere: a review. J Appl Microbiol 104:613-626.

33. Hutchison ML, Avery SM, Monaghan JM. 2008. The air-borne distribution of zoonotic agents from livestock waste spreading and microbiological risk to fresh produce from contaminated irrigation sources. J Appl Microbiol 105:848-857.

34. Ibekwe AM, Watt PM, Shouse PJ, Grieve SM. 2004. Fate of *Escherichia coli* O157:H7 in irrigation water on soils and plants as validated by culture method and real-time PCR. Can J Microbiol 50(12):1007-1014.

35. Ilic S, Rajic A, Britton CJ, Grasso E, Wilkins W, Totton S, Wilhelm B, Waddell L, LeJeune J. 2012. A scoping study characterizing prevalence, risk factors and intervention research, published between 1990 and 2010, for microbial hazards in leafy green vegetables. Food Control 23:7-19.

36. Ingram DT, Patel J, Sharma M. 2011. Effects of repeated irrigation with water containing varying levels of total organic carbon on the persistence of *Escherichia coli* O157:H7 on baby spinach. J Food Prot 74(5):709-717.

37. Islam M, Doyle MP, Phatak SC, Millner P, Jiang X. 2004. Persistence of Enterohemorrhagic *Escherichia coli* O157:H7 in soil and on leaf lettuce and parsley grown in fields treated with contaminated manure composts or irrigation water. J Food Prot 67(7):1365-1370.

38. Ivanek R, Grohn YT, Wells MT, Lembo AJ Jr, Sauders BD, Wiedmann M. 2009. Modeling of spatially referenced environmental and meteorological factors influencing the probability of *Listeria* species isolation from natural environments. Appl Environ Microbiol 75(18):5893-5909.

39. Jarvis B, Corry JEL, Hedges AJ. 2007. Estimates of measurement uncertainty from proficiency testing schemes, internal laboratory quality monitoring and during routine enforcement examination of foods. J Appl Microbiol 103:462-467.

40. Jeamsripong S, Chase JA, Jay-Russell MT, Buchanan RL, Atwill ER. 2019. Experimental in-field transfer and survival of *Escherichia coli* from animal feces to romaine lettuce in Salinas Valley, California. Microorganisms 7(408).

41. Kass, RE, Wasserman, L. 1995. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. J Am Stat Assoc 90 (431): 928–934.

42. Lee D, Tertuliano M, Harris C, Vellidis G, Levy K, Coolong T. 2019. *Salmonella* survival in soil and transfer onto produce via splash events. J Food Prot 82(12):2023-2037.

43. Lopez-Galvez F, Gil MI, Allende A. 2018. Impact of relative humidity, inoculum carrier and size, and native microbiota on *Salmonella* ser. Typhimurium survival on baby lettuce. Food Microbiol 70:155-161.

44. Lopez-Velasco G, Sbodio A, Tomas-Callejas A, Wei P, Tan KH, Suslow TV. 2012. Assessment of root uptake and systematic vine-transport of *Salmonella enterica* sv. Typhimurium by melon (*Cucumis melo*) during field production. Int J Food Microbiol 158:65-72.

45. Lopez-Velasco G, Tomas-Callejas A, Sbodio AO, Pham X, Wei P, Diribsa D, Suslow TV. 2015. Factors affecting cell population density during enrichment and subsequent molecular detection of *Salmonella enterica* and *Escherichia coli* O157:H7 on lettuce contaminated during field production. Food Control 54:165-175.

46. Machado-Moreira B, Richards K, Brennan F, Abram F, Burgess CM. 2019. Microbial contamination of fresh produce: what, where, and how? Compr Rev Food Sci Food Saf 18(6):1727-1750.

47. McKellar RC, Perez-Rodriguez F, Harris LJ, Moyne AL, Blais B, Topp E, Bezanson G, Bach S, Delaquis P. 2014. Evaluation of different approaches for modeling *Escherichia coli* O157:H7 survival on field lettuce. Int J Food Microbiol 184:74-85.

48. Medina-Martinez MS, Allende A, Barbera GG, Gil MI. 2015. Climatic variations influence the dynamic epiphyte bacteria of baby lettuce. Food Res Int 68:54-61.

49. Moyne AL, Sudarshana MR, Blessington T, Koike ST, Cahn MD, Harris LJ. 2011. Fate of *Escherichia coli* O157:H7 in field-inoculated lettuce. Food Micorbiol 28:1417-1425.

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

51. Munther DS, Carter MQ, Aldric CV, Ivanek R, Brandl MT. 2019. Formation of E. coli O157:H7 persister cells in the lettuce phyllosphere and application of differential equation models to predict their prevalence on lettuce plants in the field. Appl Environ Microbiol 01602-19.

52. Oliveira M, Vinas I, Usall J, Anguera M, Abadias M. 2012. Presence and survival of *Escherichia coli* O157:H7on lettuce leaves and in soil treated with contaminated compost and irrigation water. Int J Food Microbiol 156:133-140.

53. Ottoson JR, Nyberg K, Lindqvist R, Albihn A. 2011. Quantitative microbial risk assessment for *Escherichia coli* O157:H7 on lettuce, based on survival data from controlled studies in a climate chamber. J Food Prot 74(12):2000-2007.

54. Pachepsky Y, Shelton DR, McLain JET, Patel J, Mandrell RE. 2011. Irrigation waters as a source of pathogenic microorganisms in produce: a review. Adv Agron 113:73-138.

55. Park S, Szonyi B, Gautam R, Nightingale K, Aniciso J, Ivanek R. 2012. Risk factors for microbial contamination in fruits and vegetables at the preharvest level: a systematic review. J Food Prot 75(11):2055-2081.

56. Poza-Carrion C, Suslow T, Lindow S. 2013. Resident bacteria on leaves enhance survival of immigrant cells of *Salmonella enterica*. Phytopathology 103(4):341-351.

57. Rubin, DB. 1987. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley and Sons.

58. Solomon EB, Pang HJ, Matthews KR. 2003. Persistence of *Escherichia coli* O157:H7 on lettuce plants following spray irrigation with contaminated water. J Food Prot 66(12):2198-2202.

59. Steele M, Odumeru J. 2004. Irrigation water as a source of foodborne pathogens on fruit and vegetables. J Food Prot 67(12):2839-2849.

60. Stine SW, Song I, Choi CY, Gerba CP. Effect of relative humidity on preharvest survival of bacterial and viral pathogens on the surface of cantaloupe, lettuce, and bell peppers. J Food Prot 68(7):1352-1358.

61. Therneau T, Atkinson B. 2019. rpart: Recursive Partitioning and Regression Trees. R package version 4.1-15. https://CRAN.R-project.org/package=rpart

62. Tomas-Callejas A, Lopez-Velasco G, Camacho AB, Artes F, Artes-Hernandez F, Suslow TV. 2011. Survival and distribution of *Escherichia coli* on diverse fresh-cut baby leafy greens under preharvest through postharvest conditions. Int J Food Microbiol 151:216-222.

63. U.S. Food and Drug Administration. 2015. Standards for the Growing, Harvesting, Packing, and Holding of Produce for Human Consumption. Final Rule, Docket No. FDA–2011–N–0921, Accessed on December 15, 2019. https://www.federalregister.gov/documents/2015/11/27/2015-28159/standards-for-the-growing-harvesting-packing-and-holding-of-produce-for-human-consumption

64. U.S. Food and Drug Administration. 2019. FDA, CDC and other health partners investigated outbreak of E. coli O157:H7 possibly linked to romaine lettuce, outbreak appears to be over. Accessed on December 15, 2019. https://www.fda.gov/news-events/fda-brief/fda-cdc-and-other-health-partners-investigated-outbreak-e-coli-o157h7-possibly-linked-romaine

65. U.S. Food and Drug Administration. 2019. Recalls, market withdrawals, & safety alerts. Accessed on December 15, 2019. https://www.fda.gov/safety/recalls-market-withdrawals-safety-alerts

66. Uyttendaele M, Jaykus LA, Amoah P, Chiodini A, Cunliffe D, Jacxsens L, Holvoet K, Korsten L, Lau M, McClure P, Medema G, Sampers I, Jasti PR. 2015. Microbial hazards in irrigation water: standards, norms, and testing to manage use of water in fresh produce primary production. Compr Rev Food Sci Food Saf 14:336-356.

67. van Buuren S, Groothuis-Oudshoorn K. 2011. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1-67. https://www.jstatsoft.org/v45/i03/.

68. Van der Linden I, Cottyn B, Uyttendaele M, Vlaemynck G, Maes M, Heyndrickx M. 2014. Evaluation of an attachement assay on lettuce leaves with temperature- and starvation-stressed *Escherichia coli* O157:H7 MB3885. J Food Prot 77(4):549-557.

69. Venables WN, Ripley BD. 2002. Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0.

70. Walk ST, Alm EW, Gordon DM, Ram JL, Toranzos GA, Tiedje JM, Whittam TS. 2009. Cryptic lineages of genus *Escherichia.* Appl Environ Microbiol 75(20):6534-6544.

71. Wall GL, Clements DP, Fisk CL, Stoeckel DM, Woods KL, Bihn EA. 2019. Meeting report: keys outcomes from a collaborative summit on agricultural water standards for fresh produce. Compr Rev Food Sci Food Saf 18:723-737.

72. Weller DL, Kovac J, Roof S, Kent DJ, Tokman JI, Kowalcyk B, Oryang D, Ivanek R, Aceituno A, Sroka C, Wiedmann M. 2017. Survival of *Escherichia coli* on lettuce under field conditions encountered in the northeastern United States. J Food Prot 80(7):1214-1221.

73. Wood JD, Bezanson GS, Gordon RJ, Jamieson R. 2010. Population dynamics of *Escherichia coli* inoculated by irrigation into the phyllosphere of spinach grown under commercial production conditions. Int J Food Microbiol 143:198-204.

CHAPTER 7

CONCLUSIONS

L. monocytogenes, *Salmonella* spp., and EHEC present many unique challenges along the produce supply chain. As such, mitigating risks associated with each of these pathogens is complex and requires multiple control strategies. The research presented here (i) identified and characterized microbial food safety hazards along the produce supply chain and (ii) tested strategies for controlling these hazards. In particular, the first three studies reviewed and assessed *Listeria* spp. and *L. monocytogenes* presence, diversity, and associated risks in the pre- and post-harvest environments and proposed a root cause analysis procedure for reducing or eliminating persistent and persistent transient *Listeria* from post-harvest operations. On the other hand, the final two studies investigated *Salmonella* and *E. coli* in the pre-harvest environment. Specifically, they (i) used spatial-temporal modeling to predict *Salmonella* spp. and EHEC presence in southwestern U.S. canals utilized for agricultural water and (ii) characterized the effects of weather on *Salmonella* and *E. coli* dynamics on leafy greens contaminated via a simulated irrigation event.

The review indicated *Listeria* spp. and *L. monocytogenes* are prevalent in a variety of environments, including natural, urban, primary production, processing, and retail environments, among others. As such, there are a variety of points along the supply chain where control measures for preventing *L. monocytogenes* contamination are necessary to protect against associated public health and business risks. These control measures should include GMPs; regular cleaning and sanitation programs; the use of equipment with sanitary designs (i.e., equipment without niches); and one-directional flow of employees, equipment, and food products. In addition, food products and their processing environments should be monitored for *Listeria* presence, persistence, and persistent transience; these environmental monitoring programs need to be linked to specific goals (e.g., verification of certain food safety programs, such as sanitation programs). Furthermore, the large number of potential sources of *Listeria* throughout the food supply chain make it difficult to

identify the true source of contamination when detected in the environment or products. As such, the use of formal and well executed root cause analysis, "for-cause" investigations, and subtyping tools are essential in investigations of *Listeria* to (i) address the immediate problem and (ii) implement control measures to prevent similar events from occurring in the future.

Sampling of feces and agricultural water sources from a New York State produce farm indicated *Listeria* is prevalent in this environment. As such, feces and water can represent direct (i.e., through contact with produce) and indirect (i.e., introduction into the packing or processing environment and subsequent produce contamination) routes of produce contamination. The population of *Listeria* observed in the relatively small area of the produce farm (<0.5 km^2) was diverse, and there was evidence of transfer between sample types or continuous reintroduction of *Listeria* in this environment. This indicates successful traceback of isolates from finished products or the processing environment back to specific farm sources is likely difficult, unless large "swab-a-thons" are performed that include subtyping of a large number of isolates per sample type. However, even with large "swab-a-thons" successful traceback to the pre-harvest environment may not be possible in some cases.

Since persistent and persistent transient *Listeria* contamination of produce packing and processing environments has been linked to public health and business risks, the root cause of the contamination must be identified and controlled. As such, a modified version of the RCA protocol described in Chapter 2 was tested in an apple packinghouse to identify, implement, test interventions to eliminate known persistent and persistent transient *Listeria* populations. Two of the 4 site-specific interventions tested in the packinghouse were successful, indicating RCA was a useful strategy for identifying the initial cause of *Listeria* contamination in the packinghouse. Specifically, the use of quat powder was effective at preventing *Listeria* isolation from (i) forklift stops and (ii) a floor crack. On the other hand, the use of a cleaner sanitizer in drains and the removal of a dead-end pipe did not eliminate persistent or transient persistent *Listeria* populations. This highlights that fact that while these 2 interventions were not initially successful, use of an iterative process to test several possible root causes or several interventions for a given root cause

is often necessary to identify and fix the true cause of a problem. Regardless, the RCA protocol tested in this study can be used as a guide by other produce operations (i.e., packinghouse or processing facilities) to identify the root cause of the contamination and eliminate persistent or persistent transient *Listeria*.

Salmonella spp. and EHEC are also associated with public health risks, especially in pre-harvest produce environments. As such, strategies for rapidly identifying *Salmonella* and EHEC presence in these environments, and in particular in agricultural water sources is needed. Machine learning-based predictive models, such as conditional forest models, show promise for predicting *Salmonella* presence and EHEC marker detection in southwestern US canals used as sources for agricultural water. The use of machine learning models, in addition to regression analysis, provides a more complete assessment of the relationships between spatial and temporal factors and foodborne pathogen presence in agricultural water due to the complexity in the system. Furthermore, the use of predictive modeling, and real-time predictive models (i.e., using no microbiological data), may be able to replace traditional generic *E. coli* testing to determine when and where water may require treatment before produce application. The conditional forest models developed specifically indicated a positive monotonic relationship between the percent of natural land cover around the sampling sites and the probability of a sampling being positive for *Salmonella* presence or EHEC marker detection. However, the effect appears to be slight and does not warrant changes to the landscape surrounding produce production areas, including unintended consequences that may be associated with these types of interventions. In addition, the condition forest models indicated that the probability of a sample being positive for *Salmonella* presence and EHEC marker detection, respectively, also increases substantially after approx. September; this indicates a potential seasonal contamination source or transmission pathway. Regardless, prior to application of predictive modeling for this purpose in the real-world, additional studies are needed to determine (i) if the models developed remain effective from year to year and over space (i.e., in other produce growing regions), (ii) if canal specific models would be more effective for predicting

foodborne pathogen presence, and (iii) if the important predictors identified hold across space and time.

Furthermore, in the case that *Salmonella* and EHEC presence in agricultural water is not identified prior to application of the water to produce, control measures must be utilized to reduce the risk of associated illnesses and outbreaks. The final study presented here quantified in-field die-off of *Salmonella* and generic *E. coli* on baby spinach and lettuce following a simulated irrigation event with contaminated water. The goal was to determine if the use of wait times between water application and harvest can be used as a feasible approach to control these hazards as proposed in the Food Safety Modernization Act (FSMA) Produce Safety Rule. In addition, the effect of weather on die-off rate was assessed to determine the conditions under which wait times between water application and harvest were more effective. The results indicated die-off of *Salmonella* and *E. coli* on baby spinach and lettuce follows a biphasic segmented log-linear pattern, with more rapid die-off in the first phase compared to the second in most cases. The die-off rate in the first segment was variable and appeared to be associated with relative humidity, produce type, and bacteria. After the breakpoint, the die-off rate was less variable, however, there was still a large variation in the microbial counts at each time point in this segment across experimental plots and trials. Additionally, these results provided evidence that relative humidity can be used to estimate when experimental die-off is compliant with the FSMA. As such, the use of a single die-off rate, as proposed by FSMA, is likely not appropriate. Rather it is likely more useful to use wait times between water application and harvest in conjunction with other control methods (i.e., antimicrobial water treatments) to reduce the likelihood of illnesses, outbreaks, and recalls considering no control measure is 100% effective.

Overall, these studies have identified sources of *L. monocytogenes*, *Salmonella* spp., and EHEC across the supply chain, including feces and agricultural water in the pre-harvest packinghouse environments. This information provides a better understanding of the conditions under which these pathogens can be found to help guide the implementation of control strategies. In addition, these studies have tested strategies for controlling these pathogens across the produce

supply chain, including wait times between water application to produce and harvest, as well as the implementation of a root cause analysis procedure for identifying interventions to eliminate *Listeria* from the post-harvest environment. The findings from these studies can be used by industry to help guide the development and revision of food safety programs, by regulatory agencies to help guide implementation of new rules and regulations, and by academia to guide additional research in the produce safety field.

SUPPLEMENTAL MATERIALS

Table S4.1. Description of sampling sites within the apple packinghouse.

Site ID	Site description
301	Inside of large PVC by rot bin next to flume.
302	Inside of small PVC pipe under end of flume before rollers. S of site #301
303	Inside of small PVC by chemical barrels. N of site #'s 301 and 302
304	Inside of large PVC by start of flume near bin submerge
305	Inside of PVC in NW corner at end of brush bed
306[a]	Inside dead-end (removed to swab) of PVC in NW corner. Same pipe as site #305
307[a]	Same as 306
308[a]	Same as 306 and 307
309	Top of PVC pipe in NW corner (part that connects with the catch pan)
310	Catch pan at N end of the brush bed by NW corner. Crack/ metal juncture
311	Plastic/ metal joints supporting south end of brush bed by west platform stairs
312	Metal cross beams with wires wrapped around them under the brush bed. Two by stairs at the SE part of the platform
313	Large screws under brush-bed (4 total). The four on the E side of brush bed
314	Plastic/ metal juncture that supports brush bed. Under 2nd set of fans. Accessed from the NW corner
315	Gap between box for brush bed motors and catch pan in NW corner under the brush bed.
316	Frame of brush bed on NW side. By second set of fans
317	Large screws under brushes (directly). 4 in total under second set of fans
318	Forklift stop – full bins, by bin loader
319	Forklift stop – empty bin return
320	Floor crack on NW side of bin return
321	Trench drain by NW corner PVC pipe. 3rd grade from corner
322	Square drain by dryer – PVC pipe in center
323	Square drain by dryer – outer edge
324	PVC pipe (inside) by juice reject belt
325	Square drain between packing area and brush beds – PVC outlet
326	Square drain between packing area and brush beds – outer edge
327	Square drain by pelletizer – outside/ top
328	Square drain by flume/ bin dump – PVC pipe
329	Square drain by flume/ bin dump – outer edge
330	Flume connection flange – south-most by half wall
331	Metal frame lip/ gap at rollers after flume. W side, on platform
332	Overflow door by flume (inner section) by leaf pile
333	Condensation on PVC pipe under flume, at U-bend
334	Four large screws under the brush bed. Last four before fan box, near stairs. E side
335	Small trench leading to square drain by flume

[a]Samples were not collected from these sites during sampling 5 (10/30/21), as the dead-end was removed from the pipe.

Table S4.2. Mixed effects logistic regression model results explaining the relationship between *Listeria* isolation in the packinghouse and intervention status (i.e., if the sample was collected before or after increasing the frequency of cleaning and sanitation to twice a week[a]). Sampling site was included in the model as a random effect[b].

Variable	Log-odds (95% CI[d])	P-value
Intercept	-6.7 (-11.2, -3.0)	0.002
After intervention implementation[c]	0.3 (-1.3, 2.1)	0.711

[a]Prior to increasing the cleaning and sanitation frequency, cleaning and sanitation was performed once a week.

[b]For the sampling site random effect, the residual variance and intercept were 28.34 and 5.324, respectively.

[c]The baseline of the intervention status variable is before intervention implementation (i.e., when cleaning and sanitation was performed once a week). After intervention implementation, cleaning and sanitation was performed twice a week.

[d]95% CI= 95% confidence interval.

Table S5.1. Description of factors used in regression modeling and conditional random models.

Factor Type	Factor	Units
Sampling site	Day of year	Julian day, 1-365
	Latitude	Decimal degrees, N[g]
	Longitude	Decimal degrees, W[g]
	Irrigation district[a]	NA
	Bridge present	Yes/ No
	Point of discharge present	Yes/ No
	Intersecting canal present	Yes/ No
	Construction present	Yes/ No
	Road crossing present	Yes/ No
	Bottom type	Concrete/ Earth
	Parallel road present	Yes/ No
Water quality	*E. coli* level	Log_{10} MPN/ 100 mL
	pH	NA
	Water temperature	°C
	Turbidity	NTU
	Dissolved oxygen	mg/L
	Conductivity	µS/cm
	Surface flow rate	m/s
	Trash present at sample collection	Yes/ No
	Silt present at sample collection	Yes/ No
Land use[b]	Developed (open) cover in 0-250 ft buffer	Percent of buffer area
	Developed cover in 0-250 ft buffer	Percent of buffer area
	Barren cover in 0-250 ft buffer	Percent of buffer area
	Natural cover in 0-250 ft buffer	Percent of buffer area
	Pasture in 0-250 ft buffer	Percent of buffer area
	Cropland in 0-250 ft buffer	Percent of buffer area
	Developed (open) cover in 250-500 ft buffer	Percent of buffer area
	Developed cover in 250-500 ft buffer	Percent of buffer area
	Barren cover in 250-500 ft buffer	Percent of buffer area
	Natural cover in 250-500 ft buffer	Percent of buffer area
	Pasture in 250-500 ft buffer	Percent of buffer area
	Cropland in 250-500 ft buffer	Percent of buffer area
	Developed (open) cover in 500-1,000 ft buffer	Percent of buffer area
	Developed cover in 500-1,000 ft buffer	Percent of buffer area
	Barren cover in 500-1,000 ft buffer	Percent of buffer area
	Natural cover in 500-1,000 ft buffer	Percent of buffer area
	Pasture in 500-1,000 ft buffer	Percent of buffer area
	Cropland in 500-1,000 ft buffer	Percent of buffer area
	Developed (open) cover in 1,000-5,000 ft buffer	Percent of buffer area
	Developed cover in 1,000-5,000 ft buffer	Percent of buffer area
	Barren cover in 1,000-5,000 ft buffer	Percent of buffer area
	Natural cover in 1,000-5,000 ft buffer	Percent of buffer area

	Pasture in 1,000-5,000 ft buffer	Percent of buffer area
	Cropland in 1,000-5,000 ft buffer	Percent of buffer area
	Developed (open) cover in 5,000-10,000 ft buffer	Percent of buffer area
	Developed cover in 5,000-10,000 ft buffer	Percent of buffer area
	Barren cover in 5,000-10,000 ft buffer	Percent of buffer area
	Natural cover in 5,000-10,000 ft buffer	Percent of buffer area
	Pasture in 5,000-10,000 ft buffer	Percent of buffer area
	Cropland in 5,000-10,000 ft buffer	Percent of buffer area
	Number of CAFOs in 0-10,000 ft buffer[c]	NA
Weather[d]	Temperature at sample collection	°C
	Avg temperature 0-12h before sample collection	°C
	Avg temperature 12-24h before sample collection	°C
	Avg temperature 1-2d before sample collection	°C
	Avg temperature 2-3d before sample collection	°C
	Max solar radiation at sample collection	mJ/m^2
	Solar radiation 0-12h before sample collection	mJ/m^2
	Max solar radiation 12-24h before sample collection	mJ/m^2
	Max solar radiation 1-2d before sample collection	mJ/m^2
	Max solar radiation 2-3d before sample collection	mJ/m^2
	Precipitation 0-12h before sample collection[e]	Yes/ No
	Precipitation 12-24h before sample collection	Yes/ No
	Precipitation 1-2d before sample collection	Yes/ No
	Precipitation 2-3d before sample collection	Yes/ No
	Wind speed at sample collection	m/s
	Max wind speed 0-12h before sample collection	m/s
	Max wind speed 12-24h before sample collection	m/s
	Max wind speed 1-2d before sample collection	m/s
	Max wind speed 2-3d before sample collection	m/s
	Vapor pressure at sample collection	kPas
	Avg vapor pressure 0-12h before sample collection	kPas
	Avg vapor pressure 12-24h before sample collection	kPas
	Avg vapor pressure 1-2d before sample collection	kPas
	Avg vapor pressure 2-3d before sample collection	kPas

[a]The sites came from 9 irrigation districts. The irrigation districts were assigned codes to keep site locations confidential. Irrigation district was only used in logistic regression models as a random effect and was not included in conditional random forest models; latitude and longitude coordinates were used in conditional random forest models in place of irrigation district.

[b]Buffer areas were calculated around each sampling point. All land use data, except CAFO, came from the 2016 USGS National Land Cover Database. The land use categories are grouped as follows: developed, open (parks, golf courses, large-lot single family homes), developed, barren,

natural (forest and wetlands), pasture, crop.

[c]CAFO are concentrated animal feeding operation. The locations of each CAFO in the study area came from an internal database. There were 11 CAFOs in the study area.

[d]All weather data was collected from the AZMET database (https://cals.arizona.edu/AZMET/). The closest weather station to each point was used to create weather factors.

[e]Precipitation at sample collection was excluded as a potential variable, as there were no samples with precipitation at sample collection.

[g]WGS 1984 coordinate system was used.

Table S5.2. Summary statistics for continuous water quality factors used in regression and conditional forest analysis.

Factor	Min[a]	Q1	Median	Q3	Max[a]	Mean	SD[a]
E. coli level (log_{10} MPN/ 100 mL)	0.0	0.9	1.4	1.9	3.5	1.4	0.7
pH	6.2	7.1	7.6	8.1	10.1	7.6	0.6
Water temperature (°C)	12.5	21.8	24.8	27.3	33.7	24.3	4.5
Turbidity (NTU)	0.3	2.6	5.3	17.5	798.0	32.7	92.7
Dissolved oxygen (mg/L)	1.1	7.0	8.0	8.9	12.4	8.0	1.6
Conductivity (µS/cm)	4.9	1025.0	1119.0	1285.0	4650.0	1441.9	880.8
Surface flow rate (m/s)	0.0	0.5	0.9	1.4	4.0	1.0	0.6

[a]Min=minimum, Max=maximum, SD= standard deviation.

Table S5.3. Summary statistics for continuous land use factors used in regression and conditional forest analysis.

Buffer Area	Cover Type[a]	Min[b]	Q1	Median	Q3	Max[b]	Mean	SD[b]
0-250ft	Developed (open)	0.0	0.0	8.0	19.1	36.9	11.5	11.8
	Developed	0.0	0.0	0.0	8.8	71.3	8.7	17.3
	Barren	0.0	0.0	0.0	0.0	100.0	2.9	14.1
	Natural	0.0	0.0	0.0	24.0	100.0	18.8	32.5
	Pasture	0.0	0.0	0.0	0.0	2.0	0.1	0.3
	Crop	0.0	14.0	70.9	87.0	100.0	56.2	36.9
250-500ft	Developed (open)	0.0	0.0	8.7	12.8	38.9	9.0	8.8
	Developed	0.0	0.0	1.4	8.4	60.5	7.7	14.0
	Barren	0.0	0.0	0.0	0.0	65.2	1.8	9.0
	Natural	0.0	0.0	0.0	30.2	100.0	19.2	29.3
	Pasture	0.0	0.0	0.0	0.0	7.5	0.2	1.2
	Crop	0.0	28.2	71.0	90.2	100.0	60.7	35.0
500-1,000ft	Developed (open)	0.0	2.8	7.0	11.5	34.8	8.1	7.6
	Developed	0.0	0.0	1.5	5.9	54.1	6.9	12.5
	Barren	0.0	0.0	0.0	0.0	38.9	1.3	5.8
	Natural	0.0	0.0	3.1	25.6	98.8	16.4	25.4
	Pasture	0.0	0.0	0.0	0.0	5.8	0.4	1.2
	Crop	0.0	40.8	79.9	91.2	100.0	64.9	31.9
1,000-5,000ft	Developed (open)	0.1	2.3	3.8	6.2	27.7	5.3	5.3
	Developed	0.0	0.7	1.1	2.6	44.5	3.6	7.1
	Barren	0.0	0.0	0.0	0.2	25.2	0.9	3.6
	Natural	0.0	0.6	7.7	28.5	89.3	16.6	20.2
	Pasture	0.0	0.0	0.0	0.4	2.3	0.2	0.4
	Crop	0.0	55.3	76.0	93.3	99.9	70.4	25.5
5,000-10,000ft	Developed (open)	0.5	2.5	3.6	6.3	14.1	4.9	3.5
	Developed	0.1	0.5	1.1	2.7	43.9	3.9	8.2
	Barren	0.0	0.0	0.3	1.5	28.1	2.1	4.9
	Natural	0.0	3.6	17.8	43.0	82.4	25.4	21.7
	Pasture	0.0	0.0	0.2	0.4	2.3	0.3	0.4
	Crop	1.5	35.8	51.1	85.4	97.5	56.3	27.5
0-10,000ft	CAFO[c]	0.0	0.0	0.0	0.0	3.0	0.1	0.4

[a]Values are the percentage of land in the given category.

[b]Min= minimum, Max= maximum, SD= standard deviation.

[c]CAFO= concentrated animal feeding operation.

Table S5.4. Summary statistics for continuous weather factors used in regression and conditional forest analysis.

Factor[a]	Time[b]	Min[a]	Q1	Median	Q3	Max[a]	Mean	SD[a]
Avg temperature (°C)	at SC	4.7	19.9	23.8	27.8	40.3	24.0	6.8
	0-12h	7.9	16.8	22.5	27.2	35.1	21.6	7.1
	12-24h	15.8	23.3	30.1	33.8	37.5	28.3	6.2
	1-2d	12.0	22.1	27.5	30.8	33.8	25.4	6.4
	2-3d	12.0	20.8	26.4	31.0	33.5	25.3	6.4
Max solar radiation (mJ/m^2)	at SC	0.0	0.5	1.2	2.1	3.6	1.3	1.0
	0-12h	0.0	0.7	1.4	2.5	3.6	1.7	1.1
	12-24h	0.3	2.3	3.0	3.4	3.7	2.8	0.9
	1-2d	0.9	3.0	3.2	3.6	3.7	3.2	0.5
	2-3d	2.2	3.0	3.3	3.6	3.8	3.2	0.4
Max wind speed (m/s)	at SC	0.0	1.0	1.6	2.8	6.3	1.9	1.2
	0-12h	1.6	2.5	3.1	3.9	8.1	3.4	1.4
	12-24h	1.3	2.5	3.6	4.8	9.5	4.0	1.8
	1-2d	1.7	3.0	4.1	5.2	8.3	4.2	1.5
	2-3d	1.9	3.3	4.9	5.7	7.6	4.7	1.6
Avg vapor pressure (kPas)	at SC	0.3	0.6	0.8	1.4	2.7	1.1	0.7
	0-12h	0.4	0.6	0.8	1.4	2.6	1.1	0.6
	12-24h	0.3	0.5	0.8	1.5	2.6	1.1	0.6
	1-2d	0.3	0.6	0.8	1.5	2.4	1.1	0.6
	2-3d	0.3	0.6	0.7	1.5	2.2	1.0	0.6

[a]Avg= average, Min= minimum, Max= maximum, SD= standard deviation.

[b]Time refers to the time before sample collection.

Table S5.5. Results of principal component analysis of land cover, weather, sampling site characteristics, and water quality factors identified to be associated ($P \leq 0.1$) with pathogen detection outcomes (*Salmonella* and EHEC marker presence) in univariable regression analysis (with day of year and irrigation district included as random effects).

Outcome	Factor	Principle Component[a]
Salmonella presence	*E. coli* level (log_{10}MPN/100 mL water)	No loading
	Turbidity (NTU)	No loading
	Developed (open) land in 250-500 ft buffer (%)	PC4
	Developed land in 250-500 ft buffer (%)	No loading
	Natural land in 250-500 ft buffer (%)	PC3
	Developed (open) land in 500-1,000 ft buffer (%)	No loading
	Developed land in 500-1,000 ft buffer (%)	No loading
	Natural land in 500-1,000 ft buffer (%)	No loading
	Developed (open) land in 1,000-5,000 ft buffer (%)	PC2
	Developed land in 1,000-5,000 ft buffer (%)	No loading
	Natural land in 1,000-5,000 ft buffer (%)	No loading
	Developed (open) land in 5,000-10,000 ft buffer (%)	PC4
	Developed land in 5,000-10,000 ft buffer (%)	No loading
	Natural land in 5,000-10,000 ft buffer (%)	No loading
	Crop land in 5,000-10,000 ft buffer (%)	No loading
	Precipitation 12-24h before sample collection (Yes/ no)	NA
	Precipitation 1-2d before sample collection (Yes/ no)	NA
	Max wind speed 2-3d before sample collection (m/s)	No loading
	Vapor pressure at sample collection (kPas)	PC1
	Avg vapor pressure 0-12h before sample collection (kPas)	PC1
	Avg vapor pressure 12-24h before sample collection (kPas)	PC1
EHEC marker	Point of discharge present (Yes/ no)	NA
	Road crossing present (Yes/ no)	NA
	E. coli level (log_{10}MPN/100 mL water)	PC3
	Developed (open) land in 0-250 ft buffer (%)	PC4
	Developed (open) land in 250-500 ft buffer (%)	PC2
	Natural land in 250-500 ft buffer (%)	No loading
	Developed (open) land in 500-1,000 ft buffer (%)	PC2
	Natural land in 500-1,000 ft buffer (%)	PC1
	Developed (open) land in 1,000-5,000 ft buffer (%)	No loading
	Natural land in 1,000-5,000 ft buffer (%)	PC1
	Crop land in 1,000-5,000 ft buffer (%)	PC1
	Developed (open) land in 5,000-10,000 ft buffer (%)	No loading
	Precipitation 0-12h before sample collection (Yes/ no)	NA

	Precipitation 12-24h before sample collection (Yes/ no)	NA
	Precipitation 1-2d before sample collection (Yes/ no)	NA

[a]Principal component analysis was performed separately for each outcome and was applicable only to continuous factors significantly associated ($P \leq 0.1$) by univariable analysis [therefore, the principal component (PC) next to categorical variables and outcomes are listed as "NA"]. "No loading" indicates the factor does not have a major loading on one of the retained principal components. For *Salmonella* presence, 4 principal components (PC) explained 97% of the variation in the data. For EHEC marker presence, 4 PC's explained 94% of the variation in the data.

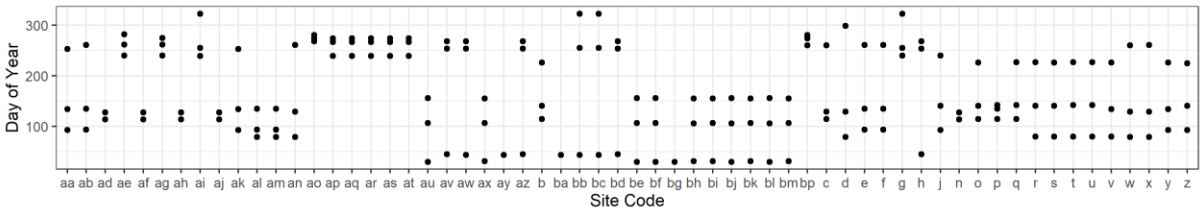


Figure S5.1. Date of sample collection for each canal. Site codes were randomly assigned to each canal to maintain confidentiality of the sampling locations

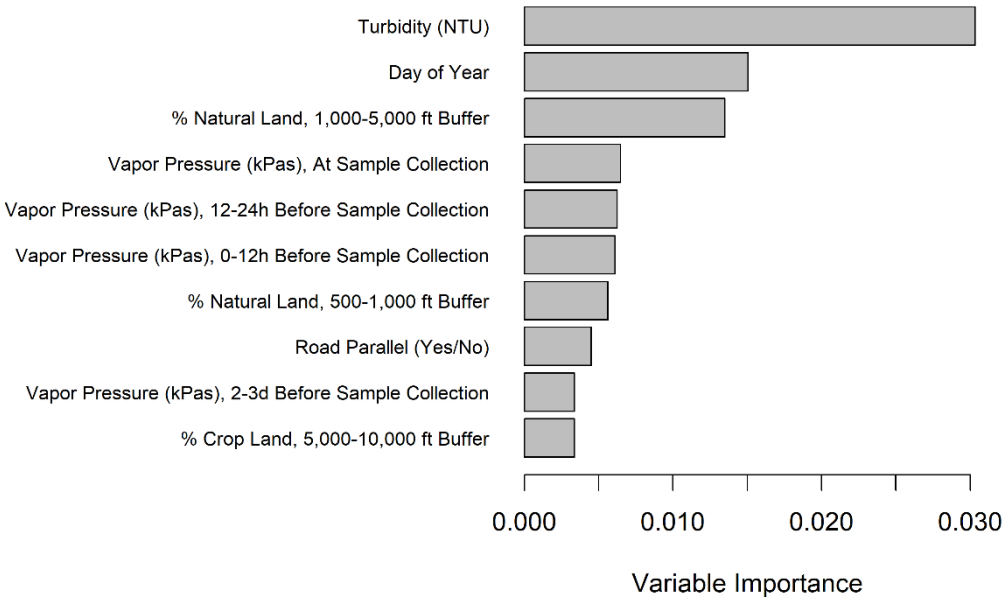


Figure S5.2. Variable importance values for the *Salmonella* presence conditional forest model with generic *E. coli* excluded as a predictor. Only the top 10 predictors are included and are listed from most to least important.

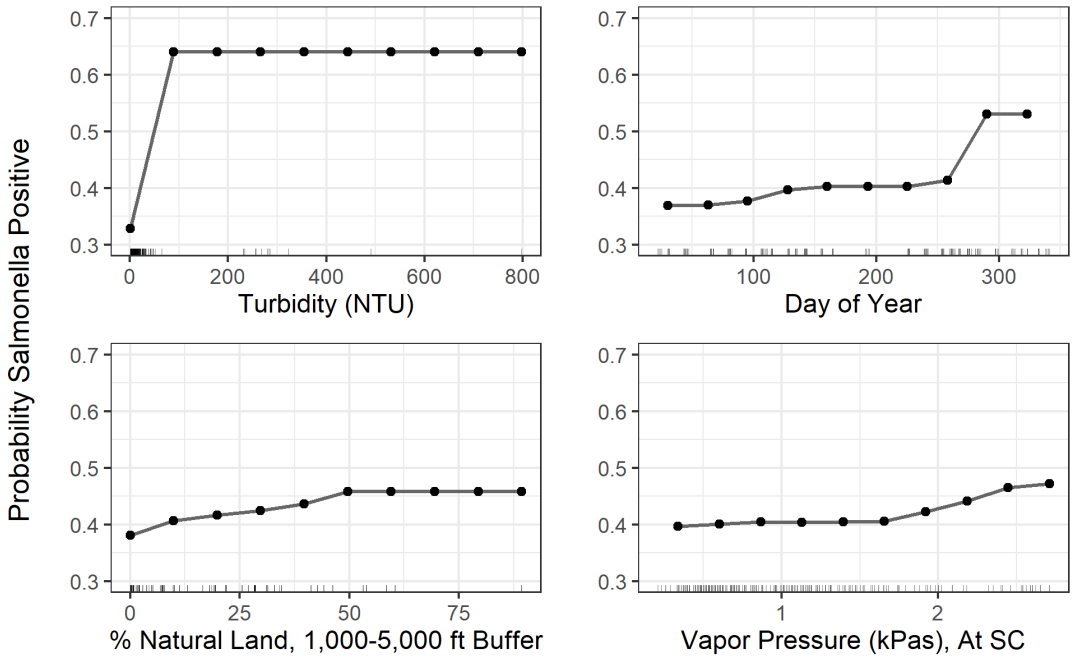


Figure S5.3. Partial dependence plots for the 4 top-ranked predictors according to variable importance in the *Salmonella* conditional forest model with generic *E. coli* excluded as a predictor; the plots indicate how the predicted probability of a water sample being positive for *Salmonella* presence changes as the x-axis variable (predictor) changes. The tick marks along the x-axis indicate values of the predictor variable in samples used to fit the conditional forest model.

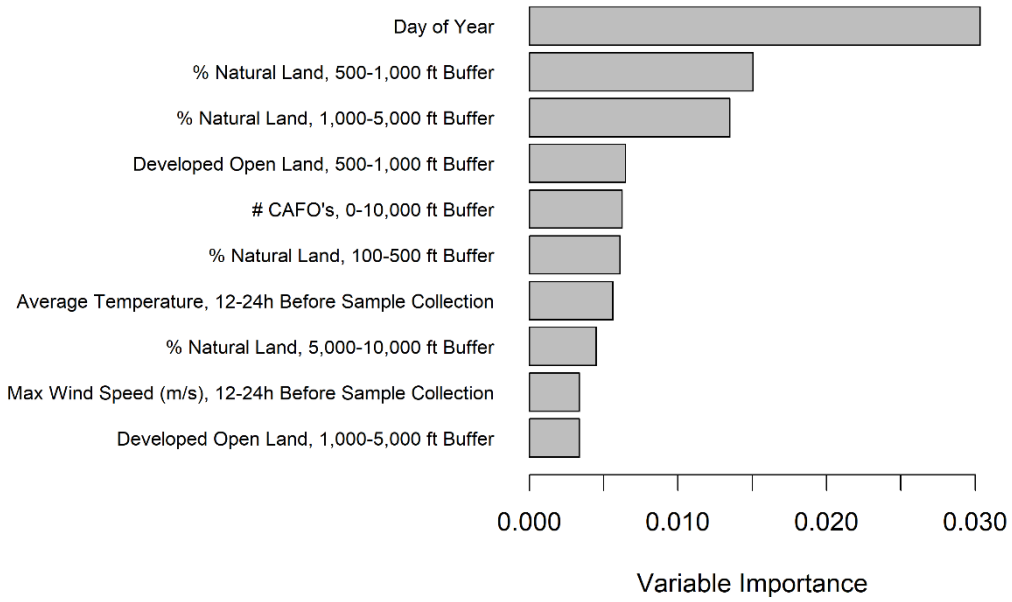


Figure S5.4. Variable importance values for the EHEC marker detection conditional random model with generic *E. coli* excluded as a predictor. Only the top 10 predictors are included and are listed from most to least important.

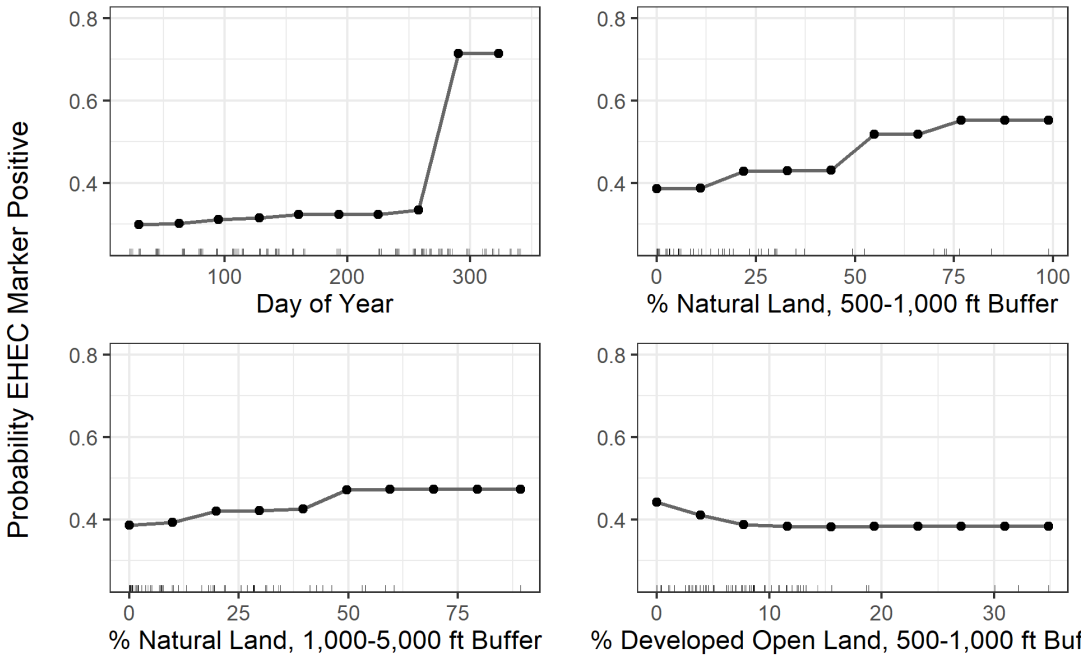


Figure S5.5. Partial dependence plots for the 4 top-ranked predictors according to variable importance in the EHEC marker conditional forest model with generic *E. coli* excluded as a predictor; the plots indicate how the predicted probability of a water sample being positive for EHEC marker detection changes as the x-axis variable (predictor) changes. The tick marks along the x-axis indicate values of the predictor variable in samples used to fit the conditional forest model.