



Published in final edited form as:

Front Sustain Food Syst. 2020 February ; 3: . doi:10.3389/fsufs.2019.00124.

## 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

Daniel Weller<sup>1,2,\*</sup>, Alexandra Belias<sup>1</sup>, Hyatt Green<sup>3</sup>, Sherry Roof<sup>1</sup>, Martin Wiedmann<sup>1</sup>

<sup>1</sup>Department of Food Science, Cornell University, Ithaca, NY, United States

<sup>2</sup>Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY, United States

<sup>3</sup>Department of Environmental and Forest Biology, SUNY College of Environmental Science and Forestry, Syracuse, NY, United States

### Abstract

There is a need for science-based tools to (i) help manage microbial produce safety hazards associated with preharvest surface water use, and (ii) facilitate comanagement of agroecosystems for competing stakeholder aims. To develop these tools an improved understanding of foodborne pathogen ecology in freshwater systems is needed. The purpose of this study was to identify (i) sources of potential food safety hazards, and (ii) combinations of factors associated with an increased likelihood of pathogen contamination of agricultural water. Sixty-eight streams were sampled between April and October 2018 (196 samples). At each sampling event separate 10-L grab samples (GS) were collected and tested for *Listeria*, *Salmonella*, and the *stx* and *eaeA* genes. A 1-L GS was also collected and used for *Escherichia coli* enumeration and detection of four host-associated fecal source-tracking markers (FST). Regression analysis was used to identify individual factors that were significantly associated with pathogen detection. We found that *eaeA-stx* codetection [Odds Ratio (OR) = 4.2; 95% Confidence Interval (CI) = 1.3, 13.4] and *Salmonella*

This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

\*Correspondence: Daniel Weller, wellerd2@gmail.com; Daniel\_Weller@urmc.rochester.edu.

#### AUTHOR CONTRIBUTIONS

DW and MW conceived of the project idea, designed the study, and wrote the grant to fund the research. DW oversaw the day-to-day aspects of the project. DW and AB led data collection efforts. DW, AB, SR, and HG performed all laboratory work. DW developed and implemented the data analysis plan. All authors contributed to manuscript development.

#### Specialty section:

This article was submitted to Agro-Food Safety, a section of the journal Frontiers in Sustainable Food Systems

#### DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

#### SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fsufs.2019.00124/full#supplementary-material>

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

isolation (OR = 1.8; CI = 0.9, 3.5) were strongly associated with detection of ruminant and human FST markers, respectively, while *Listeria* spp. (excluding *Listeria monocytogenes*) was negatively associated with  $\log_{10}$  *E. coli* levels (OR = 0.50; CI = 0.26, 0.96). *L. monocytogenes* isolation was not associated with the detection of any fecal indicators. This observation supports the current understanding that, unlike enteric pathogens, *Listeria* is not fecally-associated and instead originates from other environmental sources. Separately, conditional inference trees were used to identify scenarios associated with an elevated or reduced risk of pathogen contamination. Interestingly, while the likelihood of isolating *L. monocytogenes* appears to be driven by complex interactions between environmental factors, the likelihood of *Salmonella* isolation and *eaeA-stx* codetection were driven by physicochemical water quality (e.g., dissolved oxygen) and temperature, respectively. Overall, these models identify environmental conditions associated with an enhanced risk of pathogen presence in agricultural water (e.g., rain events were associated with *L. monocytogenes* isolation from samples collected downstream of dairy farms;  $P = 0.002$ ). The information presented here will enable growers to manage their operations to mitigate the produce safety risks associated with preharvest surface water use.

### Keywords

agricultural water quality; *Listeria*; *Salmonella*; geographic information systems; contaminant sources; GFD; microbial source-tracking

## INTRODUCTION

Over the last two decades, the occurrence of multiple foodborne disease outbreaks linked to contamination of preharvest produce by wildlife (Cody et al., 1999; Jay et al., 2007; Kangas et al., 2008; Laidler et al., 2013; Kwan et al., 2014) or surface water (e.g., during irrigation; Gelting et al., 2011, 2015; Mody et al., 2011; Lundqvist et al., 2013; FDA, 2019) have highlighted the role of wildlife and surface water as on-farm sources of foodborne pathogens. As part of the traceback investigation during a 2006 *Escherichia coli* outbreak linked to bagged spinach, the outbreak strain was isolated from both feral pig feces and preharvest water from the implicated farm (Jay et al., 2007). Following this outbreak, growers reported increased pressure to adopt practices to prevent wildlife intrusion into produce fields, including through the removal of on-farm, non-crop vegetation (e.g., forest and wetland cover, hedgerows; Beretti and Stuart, 2008; Karp et al., 2015; Baur et al., 2016). Since non-crop vegetation provides key ecosystem services (e.g., erosion prevention, water filtration; Sweeney et al., 2004) its removal can directly affect environmental health and a farm's economic resiliency. In fact, studies have shown riparian buffers can prevent up to 90% of nutrients in run-off from entering streams (Schultz et al., 2004) and are effective at reducing fecal inputs into streams (Collins and Rutherford, 2004; Ferguson et al., 2007; Wilkes et al., 2013). Non-crop vegetation removal may, therefore, result in impaired water quality. Despite the potential for negative outcomes following non-crop vegetation removal, there is limited data on the impact of upstream landscape structure on the likelihood of detecting foodborne pathogens in preharvest surface water sources. Landscape structure includes patterns of land use within a watershed (e.g., percent of non-crop vegetation in riparian areas) as well as the presence, location, and distance to potential sources of fecal

contamination (e.g., livestock operations, wastewater discharge sites). Thus, additional research on the association between upstream landscape structure and foodborne pathogen contamination of preharvest surface water sources is needed to (i) develop effective strategies for managing agricultural watersheds for multiple stakeholder aims (e.g., preventing wildlife intrusion, water quality), and (ii) reduce the unintended consequences of on-farm food safety practices (e.g., removal of non-crop vegetation). One aim of the present study therefore is to characterize the association between upstream landscape structure and foodborne pathogen detection.

Recognizing the produce safety concerns surrounding preharvest water use, the US Food and Drug Administration (FDA) proposed preharvest microbial water quality standards as part of the Food Safety Modernization Act (FSMA). The proposed rule states that the *E. coli* concentration in surface water directly applied to preharvest produce must not exceed a geometric mean of 126 CFU/100-mL or a statistical threshold value (90th percentile) of 410 CFU *E. coli*/100-mL (Food Drug Administration, 2015). The geometric mean and statistical threshold value is calculated using 20 water samples collected over a 2–4 year period (Food Drug Administration, 2015). However, interpretation of *E. coli*-based water quality tests is complicated by spatiotemporal variation in the microbial quality of surface water (Goyal et al., 1977; Hipsey et al., 2008; Payment and Locas, 2011; Pandey et al., 2012; Benjamin et al., 2013; Cooley et al., 2014; Weller et al., 2015b, 2020). For example, 71% (15/21), 63% (33/52), and 6% (2/32) of surface water samples collected from the same site in Upstate New York in 2013 (unpublished), 2014 (Weller et al., 2015a), and 2017 (Weller et al., 2020), respectively, were *Listeria monocytogenes*-positive. *E. coli* levels at this site also varied by  $> 2 \log_{10}$  MPN/100-mL over the course of the 2017 growing season (Weller et al., 2020). Thus, to improve growers' ability to identify and address on-farm food safety hazards, targeted approaches that account for this variation are needed. Due to the availability of spatial data (e.g., from government databases, Google), analyses that utilize such data can facilitate identification of factors associated with foodborne pathogen detection in agricultural water (Benjamin et al., 2013; Strawn et al., 2013a; Weller et al., 2015a, 2016); these factors can then be used to develop the aforementioned targeted approaches. Past studies have shown that microbial water quality is affected by the ecological context unique to each water source (e.g., upstream landscape) as well as conditions (e.g., weather, physicochemical water quality) at time of water use (e.g., McEgan et al., 2013a; Bradshaw et al., 2016; Weller et al., 2020). It is therefore essential to understand how contamination risks vary in response to weather, physicochemical water quality, and upstream landscape factors as well as interactions between these factors. Thus, the second aim of this study was to use machine-learning approaches robust to correlation between explanatory factors (i.e., conditional inference trees) to identify combinations of environmental factors that were associated with an increased likelihood of foodborne pathogen detection.

Due to the variability in microbial water quality, past studies concluded that the proposed FSMA standard is not effective at identifying food safety risks associated with preharvest surface water use. Specifically, these studies determined that whether a water source meets the proposed standard is largely a function of when the water samples were collected and was not associated with the presence of food safety hazards at the time of water use (Havelaar et al., 2017; Truitt et al., 2018; Weller et al., 2020). Since the relationship between

*E. coli* levels and foodborne pathogen presence in agricultural water varied widely within and between past studies (Edberg et al., 2000; Harwood et al., 2005; Wilkes et al., 2009; Payment and Locas, 2011; Benjamin et al., 2013; Pachepsky et al., 2015; Antaki et al., 2016; Weller et al., 2020), concerns have also been raised about the standard's reliance on generic *E. coli* as an indicator of potential food safety hazards. In fact, a review that compiled the findings of 40 studies with data on *E. coli* levels and pathogen presence in surface water found a significant association between *E. coli* levels and pathogen presence in only 18% of the datasets (Pachepsky et al., 2015). One potential explanation for this variation in the *E. coli*-pathogen relationship is that *E. coli* is an indicator of fecal contamination and not an index organism for foodborne pathogens. This is problematic since foodborne pathogens are not always fecally-associated. For instance, *L. monocytogenes* is a free-living soil microbe, and *E. coli* and *Salmonella* can naturalize in non-host environments [e.g., water (Hendricks, 1967; Goto and Yan, 2011; McEgan et al., 2013a), submerged aquatic vegetation (Byappanahalli et al., 2003; Whitman et al., 2003; Ksoll et al., 2007; SAVs), soil (Ishii et al., 2010; Nautiyal et al., 2010; Goto and Yan, 2011; NandaKafle et al., 2018)]. Another explanation for variation in the *E. coli*-pathogen relationship between studies (Edberg et al., 2000; Harwood et al., 2005; Wilkes et al., 2009; Payment and Locas, 2011; Benjamin et al., 2013; Pachepsky et al., 2015; Antaki et al., 2016) is that *E. coli* is not host-associated, and as such, *E. coli* levels are indicative of all fecal inflows into a waterway. Thus, even when pathogen contamination is of fecal origin, the strength of the *E. coli*-pathogen relationship may be biased by the presence of other fecal inflows that are not contaminated by the target pathogen. Host-associated markers of fecal contamination may offer one way of identifying fecal sources of foodborne pathogens in agricultural water (Green et al., 2019). Moreover, being able to identify and associate pathogen presence with host-associated fecal inputs, will improve our understanding of the food safety hazards associated with human, wildlife, and livestock fecal inputs, and allow for the development of targeted interventions to manage food safety hazards in agricultural water. Thus, the third aim of the present study was to identify potential pathogen sources by characterizing the relationship between foodborne pathogen detection and (i) host-associated fecal indicators, including fecal source tracking (FST) markers for avian (GFD; Green et al., 2012), canine (DG3; Green et al., 2014b), human (HF183; Green et al., 2014a) and ruminant (Rum2Bac; Mieszkin et al., 2010) fecal contamination, and (ii) upstream sources of fecal contamination (e.g., livestock operations, wastewater discharge sites).

## MATERIALS AND METHODS

### Study Design

Sixty-eight streams in Upstate New York were sampled two to three times each between April and October 2018 (196 samples total; Figure 1); this timeframe was selected to coincide with the produce growing season in New York. At each sampling, one 10-L grab sample was collected for detection of each set of microbial targets: (i) *Listeria* [*Listeria* spp. (excluding *monocytogenes*) and *L. monocytogenes*], (ii) *Salmonella* and (iii) the *stx* and *eaeA* genes (molecular markers for the potential presence of pathogenic *E. coli*; Hamilton et al., 2010; Melton-Celsa, 2014) as previously described (Weller et al., 2020; 30-L total). A separate 1-L grab sample was also collected to characterize *E. coli* levels. Gloves (Nasco,

Fort Atkinson, WI) were changed for each sample collected. All samples were transported on ice and stored at 4°C until samples could be processed. All 10-L grab samples were processed within 18 h of sample collection, while all 1-L grab samples were processed within 6 h of sample collection.

### Grab Sample Processing

The 10-L grab samples were filtered using modified Moore swabs (mMS) as previously described (Sbodio et al., 2013; Weller et al., 2020). After filtration, each mMS was transferred to a separate, sterile Whirl-Pak bag and processed as described below. *E. coli* quantification was performed as previously reported (Weller et al., 2020). Briefly, a 100-mL aliquot of the 1-L grab sample was used for *E. coli* enumeration, which was performed using the Colilert Quanti-Tray 2000 kit (IDEXX, Westbrook, ME) per manufacturer instructions. A second 100-mL aliquot was separately filtered through a 0.45 µm polyethersulfone filter (Whatman, Chicago IL). These filters were then used for detection of FST markers specific to avian, canine, human and ruminant sources as previously described (Table 1). Filters were allowed to thaw prior to adding 29.2 µl of prepared *Caenorhabditis elegans* lysate (Kirtane et al., 2019). The *C. elegans* strain used here contains a *gfp* gene, which can be targeted using the CG4 assay, allowing for (i) estimation of the total amount of DNA recovered from each sample, and (ii) confirmation that qPCR inhibition was absent (Kirtane et al., 2019). Following the addition of the *C. elegans* lysate, the filter and lysate were homogenized using a FastPrep-24-5G (Irvine, CA, MPBio). DNA extraction was then performed using the DNeasy Blood and Tissue Kit (Germantown, MD, Qiagen). Each qPCR reaction consisted of 1X TaqMan Environmental Master Mix (ThermoFisher), DNase and Rnase free water, and assay-specific oligonucleotides (see Table 1). Duplicate reactions were run on a QuantStudio3 or QuantStudio5 (ThermoFisher) under standard cycling conditions: 50°C for 2 min, 95°C for 10 min, and 40 cycles of 95°C for 15 s and 60°C for 1 min. Extraction blanks and three no-template control wells (NTCs) were included in each qPCR run; extraction blanks and NTCs were negative for FST markers in all runs. It is important to note that in the current study, the avian FST assay was modified for hydrolysis probe-based chemistry. The modified GFD assay used the original GFD forward primer [5'-TCG GCT GAG CAC TCT AGG G; Green et al., 2012], a modified GFD reverse primer (5'-GCG TCT CTT TGT ACA TCC CA T TG), and a newly-designed ZEN® probe (5'-ACG TCA AGT CAT CAT GGC CCT TAC GC; Coralville, IA, Integrated DNA Technologies). Specificity and sensitivity of the modified GFD assay approximated that of the original SYBR Green-based assay. Specifically, the modified assay was able to correctly identify 86% (13/14) of bird fecal samples (Table S1). Approximately 5% (2/37) of bovine fecal samples were incorrectly identified as being positive for GFD and avian contamination; however, this occurred at very low concentrations per nanogram DNA in the two false-positive samples (Table S1).

### Listeria Enrichment and Isolation

*Listeria* enrichment and isolation were performed as previously described (Weller et al., 2015a, 2020). Briefly, 225 mL of buffered *Listeria* enrichment broth (Becton Dickinson, Franklin Lakes, NJ) was added to each Whirl-pak containing a modified Moore swab. Following incubation at 30°C for 4h, *Listeria* selective enrichment supplement (Oxoid,



Cambridge, UK) was added to each enrichment. Following incubation at 30°C for a total of 24 h and 48 h, 50 µl of each enrichment were streaked onto *L. monocytogenes* plating medium (LMPM; Biosynth International, Itasca, IL) and Modified Oxford agar (MOX; Becton Dickinson), which were incubated for 48 h at 35 and 30°C, respectively. Up to four presumptive *Listeria* colonies were sub-streaked from MOX to LMPM; the LMPM plates were then incubated at 35°C for 48 h. Up to two presumptive *Listeria* (excluding *L. monocytogenes*) colonies and up to two presumptive *L. monocytogenes* colonies were sub-streaked from LMPM onto brain-heart infusion plates (BHI; Becton Dickinson), which were incubated at 37°C for 24h. For each sample, PCR amplification and sequencing of the partial *sigB* gene (Nightingale et al., 2005; Den Bakker et al., 2010; Bundrant et al., 2011) were used to (i) determine the species of one presumptive *Listeria* (excluding *L. monocytogenes*) isolate, and (ii) confirm one presumptive *L. monocytogenes* isolate as *L. monocytogenes*. The protocol for the *sigB* PCR performed can be found at <https://github.com/wellerd2/Laboratory-Protocols>. Positive (FSL R3-0001, Roberts and Wiedmann, 2006) and negative controls (uninoculated media) were processed in parallel with all samples.

After processing ~85% of samples, we observed that the prevalence of *Listeria* was substantially lower for this study compared to past NY studies (Strawn et al., 2013a,b; Chapin et al., 2014; Weller et al., 2015a,b). The only methodological difference between our study and past studies was the larger volume of water collected (10-L instead of 250-mL), which necessitated filtration through a mMS instead of a 0.45 µm filter. As such, for the last 29 samples collected we filtered 9 L of the 10-L grab sample using the mMs approach and the remaining liter using a 0.45 µm filter. We then used McNemar's  $\chi^2$ -square and Cohen's Kappa to assess the relative ability of paired mMS and 0.45 µm filters to detect each *Listeria* species as well as *Listeria* spp. and *Listeria* spp. excluding *L. monocytogenes*. For modeling purposes, a sample was considered positive for a given *Listeria* species if either the mMS or filter were positive for that species.

### **Salmonella Enrichment and Isolation**

Two-hundred twenty-five milliliters of buffered peptone water supplemented with novobiocin to a concentration of 20 mg/L was added to each Whirl-pak containing a modified Moore swab. Following incubation at 35°C for 24 h, *Salmonella* negative samples and presumptive *Salmonella*-positive samples were identified using real-time BAX *Salmonella* assays (Hygiena, Wilmington, DE). BAX negative samples were considered negative for *Salmonella*, while BAX positive samples underwent culture-confirmation for *Salmonella* (Strawn et al., 2013a). Briefly, 1 mL of the enrichment was added to 9 mL of tetrathionate broth (TT; Oxoid) supplemented with 200 µL of I<sub>2</sub>-KI and 100 µL of Brilliant Green. Separately, 0.1 mL of the enrichment was added to 9.9 mL of Rappaport Vassiliadis broth (RV; Acros Organic, Geel, Belgium). After incubating the TT and RV broth in a 42°C shaking water bath for 24h, 50 µL of each broth were streaked onto separate *Salmonella* CHROMagar (DRG International, Springfield, NJ) and xylose lysine deoxycholate agar (XLD; Neogen, Lansing, MI) plates. The plates were incubated at 37 and 35°C, respectively, for 24 h. Up to 12 presumptive *Salmonella* colonies per sample were confirmed as *Salmonella* by PCR amplification of *invA* (Kim et al., 2007) using the protocol for selecting colonies for culture-confirmation described by Weller et al. (2020). The protocol for

performing primary and secondary enrichment as well as the BAX Assay can be found in the Supplementary Materials under, *Protocol for Salmonella Detection using the Real-time BAX Salmonella Assay*. The protocol for the *invA* PCR performed here can be found at <https://github.com/wellerd2/Laboratory-Protocols>. Positive (FSL F6-0826) and negative (uninoculated enrichment media) controls were processed in parallel with all samples.

### ***eaeA* and *stx* Detection**

A PCR-screen for the *stx* (both *stx1* and *stx2*) and *eaeA* genes was performed using a real-time BAX Shiga-toxin producing *E. coli* (STEC) assay (Hygiena); these genes are considered biomarkers for the potential presence of enteropathogenic *E. coli* (*eaeA*), STEC (*stx*), and enterohemorrhagic *E. coli* (*eaeA* and *stx*). Sample enrichment and processing were performed per manufacturer's instructions and as previously described (Weller et al., 2020). Briefly, 250 mL of tryptic soy broth supplemented with casamino acids to a final concentration of 10 g/L and with novobiocin to a final concentration of 8 mg/L was added to each Whirl-pak containing a modified Moore swab. The enrichment was then incubated at 41°C for 24h. Following enrichment, the BAX assay was performed per manufacturer's instructions. The protocol for performing primary enrichment as well as the BAX Assay can be found in the Supplementary Materials under, *Protocol for eaeA-stx Codetection using the Real-time BAX STEC Assay*

### **Spatial Data Acquisition and Waterway Enrollment**

Hydrological data ([www.usgs.gov/core-science-systems/ngp/national-hydrography](http://www.usgs.gov/core-science-systems/ngp/national-hydrography)), USDA Cropscape data on where produce was grown between 2009 and 2017 ([nassgeodata.gmu.edu/CropScape/](http://nassgeodata.gmu.edu/CropScape/)), transportation and infrastructure data ([tigerweb.geo.census.gov](http://tigerweb.geo.census.gov)) and data on the location of public lands ([cugir.library.cornell.edu](http://cugir.library.cornell.edu); [gis.ny.gov](http://gis.ny.gov)) were obtained to facilitate enrollment of waterways in this study. Watershed delineation and all other spatial analyses were performed using ArcGIS version 10.2 and R version 3.5.3. Watersheds were enrolled by identifying publicly accessible locations (e.g., parks, bridges, Cornell farms) along streams (i) <3.5h from the research laboratory, (ii) with a watershed area of 10 km<sup>2</sup>, and (iii) that were <400 m from a field where produce was grown in 4 of the years between 2009 and 2017. Flowlines for these watersheds were then converted from linear to point features. Sixty points from non-overlapping watersheds were randomly selected and enrolled as sampling sites in this study. During the course of the study, 11 of the sampling sites had to be replaced (e.g., due to construction, insufficient water). We were able to identify downstream sampling sites that met our enrollment criteria for three of these 11 sites. As a result, we collected three samples from each of these three streams, however, not all samples were collected at the same site; each of these streams is therefore represented by two overlapping watersheds in Figure 1. For the remaining 8 sites, we were unable to identify downstream that met our enrollment criteria. Thus, eight replacement streams were selected using the protocol described above. As a result, 60 streams were sampled three times and eight streams were sampled twice ( $N = 196$  samples total). However, because we changed sampling sites for three streams, a total of 71 watersheds are represented in the dataset (3 pairs of overlapping watersheds and 65 non-overlapping watersheds).

To characterize land cover within watersheds, we used inverse-distance weighting (IDW) as described in King et al. (2005). IDW is based on the idea that land cover in areas closer to the sampling site will have a greater impact on water quality than areas farther upstream. The IDW proportion of the total watershed, stream corridor (all area <60 m of the stream channel), and floodplain under each land cover class was calculated (see Table S2 for the list of land covers). Inverse distance weights were calculated using the following distance intervals: 0–100, 100–250, 250–500, 500–1,000, 1,000–2,000, 2,000–5,000, 5,000–10,000, 10,000–20,000, and >20,000 m upstream of the sampling site; all intervals were constrained by either watershed, stream corridor or floodplain boundaries. In addition to characterizing land cover, we also determined if specific landscape features were present upstream of the sampling site (see Table S2 for features considered). If a feature type was present we calculated the upstream flow path distance from the sampling site to the nearest feature. Briefly, flow lines, flow accumulation and flow direction rasters ([www.usgs.gov/core-science-systems/ngp/national-hydrography](http://www.usgs.gov/core-science-systems/ngp/national-hydrography)) were used to create flow networks that accounted for overland and in-channel flow using the Hydrology toolset in ArcGIS. The flow networks were imported into R, and the riverdist package (Matt Tyers, 2017) was then used to calculate the flow path distance to the nearest upstream feature for each sampling site and feature type. For features that were also potential sources of fecal contamination (e.g., livestock operations, wastewater discharge sites), the upstream density was also determined. It is important to note that we generated distance and density data for specific types of livestock operations, including dairy farms, poultry farms, and stables, as well as for all livestock operations (i.e., where density includes any livestock operation within the watershed regardless of livestock type). Since septic system data was aggregated at the census tract level (as opposed to being spatially explicit point data like the wastewater discharge sites), upstream septic system density was estimated using the equation below:

$$Density = \frac{\sum \left( \left( \frac{Area\ of\ Overlap\ Between\ Tract_i\ and\ Watershed}{Area\ of\ Tract_i} \right) * No.\ of\ Septic\ Systems\ in\ Tract_i \right)}{Watershed\ Area}$$

## Weather Data

Weather data were obtained from the NEWA weather station closest to each sample site (Figure 1; <http://newa.cornell.edu/>). The closest station was identified by drawing Thiessen polygons around all stations in Upstate NY. The average distance between the NEWA stations and the sample sites was 9 km (range = < 1–26 km). If a weather station had a malfunction then data from the next nearest station was used for the time period the malfunction persisted. Average solar radiation and total rainfall for the 0–1, 1–2, 2–3, 3–4, 4–5, 5–10, 10–20, and 20–30 days before sample collection (BSC) was calculated. Due to high Spearman's correlation between average air temperature 0–1, 1–2, 2–3, 3–4, and 4–5 days BSC, average air temperature for 0–5, 5–10, 10–20, and 20–30 d BSC was calculated. Air and water temperature were also measured in the field at sample collection.

## Statistical Analyses

All analyses were performed in R (version 3.4.2; R Core Team, Vienna, Austria). Correlation between factors was quantified and visualized using the corrplot package (Wei,



2013) as previously described (Weller et al., 2015a). The study reported here tested each sample for multiple microbial targets including key foodborne pathogens (*Salmonella* and *L. monocytogenes*), pathogen markers (*eaeA* and *stx* genes), and index organisms for key pathogens (non-pathogenic *Listeria*). Separate analyses were performed for each of these microbial targets. Since culture-based methods were used for detection of *Salmonella*, *L. monocytogenes* and non-pathogenic *Listeria*, and a PCR-screen was used to detect the *eaeA* and *stx* genes, care needs to be taken when comparing results between each microbial targets.

**Regression Analyses**—Generalized linear mixed models (GLMM) were developed to investigate potential relationships between likelihood of detecting each target, and (i) indicators of fecal contamination (e.g.,  $\log_{10}$  *E. coli* concentration, detection of FST markers), (ii) weather and physicochemical water quality factors, and (iii) spatial factors (see Table S2 for a list of all covariates considered). While past studies that investigated potential relationships between pathogens and environmental factors often used Spearman's correlation to characterize such relationships (e.g., Bradshaw et al., 2016), GLMMs were used here so stream could be included as a random effect. Stream was included as a random effect and week of the year (i.e., no. of weeks since the week that included Jan. 1st) was included as a fixed effect in all GLMMs to account for pseudoreplication in our dataset. The dependent variable in the GLMMs was detection or non-detection of the microbial target. Since this was a hypothesis-generating study, two thresholds were used for interpreting the GLMMs. Specifically,  $P < 0.05$  indicated that likelihood of microbial target detection and the factor were significantly associated, while a  $0.05 < P < 0.10$  indicated the presence of a potential relationship that warrants investigation in future studies. When interpreting the results of GLMMs where  $0.05 < P < 0.10$ , the magnitude and 95% confidence interval for the change in odds (or odds ratio for categorical explanatory factors) should be considered.

**Characterizing Spatiotemporal Variation in the Likelihood of Detecting Each Microbial Target**—GLMMs were also developed to compare the relative impact of spatial and temporal factors on the likelihood of detecting each microbial target. Using the `r.squaredGLMM` function in the `MuMIn` package, the marginal (variance accounted for by fixed effects) and conditional (variance accounted for by both fixed and random effects)  $R^2$  were estimated for each model. Temporal fixed effects considered included week of the year, day of the week (e.g., Monday, Tuesday, Wednesday), month (e.g., April, May), and hour of the day. To account for pseudo-replication, stream ID was included as a random effect in all models containing temporal fixed effects. Spatial fixed effects considered were latitude and longitude (to detect linear spatial trends at a scale larger than the watershed-level, e.g., due to N-S land use patterns in New York; Figure 1). Spatial random effects considered included county (to account for non-linear spatial trends at a scale larger than the watershed-level) and stream ID. Table S5 lists all models considered. By comparing the percent variance accounted for in these four models, the relative contribution of space and time to the observed variation in detection of each microbial target could be determined.

**Conditional Inference Trees (CTrees)**—To identify combinations of factors (specific scenarios) associated with an increased or decreased likelihood of detecting each microbial

target, CTreeS were developed using the mlr and partykit packages. Five-fold cross validation repeated 20 times was used to tune hyperparameters (minbucket and maxdepth) by optimizing the kappa score (Kuhn, 2018). A primary split, competitor split, and two surrogate splits were identified for each CTree node as described by Bradshaw et al. (2016). For CTreeS where the outcome was binary and imbalanced (frequency of positive samples was 30% < or 70% >), upsampling was performed as part of tuning (Kuhn, 2018). Due to the use of upsampling, these models may be subject to overfitting (i.e., nodes with higher numbers may be less reliable). To minimize the potential for overfitting we: (i) tuned the maxdepth and minbucket parameters, (ii) limited the upper bound of maxdepth to 10, (iii) limited the lower bound of minbucket to 20, (iv) used a mincriterion of 0.95, and (v) used the Bonferroni multiple comparison correction.

## RESULTS

In the current study, 68 streams were sampled between April and October 2018, and 196 sets of samples were collected (6,078 L total). The area of the sampled watersheds was between 9.6 and 850.0 km<sup>2</sup> (Mean = 120.8 km<sup>2</sup>; Median = 50.1 km<sup>2</sup>; Table S3). While the proportion of upstream area under any given land cover varied substantially between sites, on average, forest-wetland, cropland, and pasture predominated at all scales of analysis (e.g., whole watershed, stream corridor, floodplain; Table S3). Of the landscape features (e.g., road crossings, culverts) considered here, road crossings were the most common since all streams had at least one upstream bridge. The min. flow path distance from the sampling site to the nearest upstream road crossing ranged between 0.0 and 4.2 km (Mean = 0.5 km). Of the potential sources of fecal contamination considered here, livestock operations and specifically, dairy farms were the most common; dairy farms were present in 63 of the 71 watersheds sampled (note the 71 watersheds correspond to the 68 sampled streams since sampling sites on three of the 68 streams had to be moved). Summary statistics for all factors are reported in Tables S3, S4. Changes in weather and water quality factors over time are shown in Figure S3, while land cover for the sampled watersheds is shown in Figure 1. The only factors that were strongly correlated with time were air and water temperature, which increased from April to August and decreased from August to October (Figures S2, S3).

### Pathogen Prevalence

The presence of potential food safety hazards was determined using culture-based methods to detect *Listeria* (*L. monocytogenes*, and *Listeria* spp. excluding *monocytogenes*) and *Salmonella*, and PCR-based methods to codetect two genes, *stx* and *eaeA*, associated with pathogenic *E. coli* presence; we refer to these collectively within the paper as microbial targets. The most frequently detected targets were the *eaeA* (96%; 190/196 samples) and *stx* (68%; 133/196 samples), which were detected in samples collected from 100% (68/68) and 96% (65/68) of the sampled streams (Table 2), respectively. Temporal factors accounted for between 2% (time of day) and 94% (month) of variance in the likelihood of *eaeA-stx* codetection. Spatial factors accounted for between <1% (latitude) and 10% (longitude) of variance in the likelihood of *eaeA-stx* codetection (Table S5). *Salmonella* was isolated from 40% of samples (79/196). Temporal factors accounted for between 2% (week of the year)

and 73% (month) of variance in the likelihood of *Salmonella* isolation, Spatial factors accounted for between 1% (latitude) and 6% (longitude) of variance in the likelihood of *Salmonella* isolation (Table S5). The base model used in the univariate regression analysis accounted for 4% of variance in the likelihood of *Salmonella* isolation (Table S5). *Listeria* spp. (excluding *monocytogenes*) was isolated from 28% (55/196) of samples and 71% (48/68) streams, while *L. monocytogenes* was isolated from 10% (20/196) of samples and 28% (19/68) of streams (Table 2). Temporal factors accounted for between <1% (week of the year) and 4% (month) of variance in the likelihood of *Listeria* spp. (excluding *monocytogenes*) isolation, and for between <1% (time of day) and 8% (month) of variance in the likelihood of *L. monocytogenes* isolation. Spatial factors accounted for between <1% (longitude) and 26% (stream) of variance in the likelihood of *Listeria* spp. isolation, and for between 0% (stream) and 6% (county) of variance in the likelihood of *L. monocytogenes* isolation (Table S5). After collecting 85% of all samples we noted that the prevalence of *Listeria* was substantially lower in this study compared to past NY studies (Strawn et al., 2013a,b; Chapin et al., 2014; Weller et al., 2015a,b). We, therefore, used the last 29 samples collected to compare the ability of paired mMS and 0.45  $\mu$ m filters to detect *Listeria* in 9 L and 1 L of water, respectively. There was significant disagreement in the number of *Listeria* spp. positive samples identified using 0.45  $\mu$ m filters (18/29) compared to mMs (3/29;  $P < 0.001$ ; Table 3). Based on the Kappa test, the filters were substantially better than the mMS at recovering both *Listeria* spp. (excluding *monocytogenes*) and *L. monocytogenes* (Table 3). For example, the frequency of *L. monocytogenes* detection was 7 times greater using the filters [24%; (7/29)] compared to the mMS [3%; (1/29); Table 3].

### Association Between Indicators of Fecal Contamination and Foodborne Pathogen Detection

Each set of grab samples was tested for five indicators of potential fecal contamination: generic *E. coli* (a non-specific indicator of fecal contamination), and host-associated markers for canine (DG3), avian (GFD), human (HF183), and ruminant (Rum2Bac) fecal contamination. Figure S1 shows how levels of all five fecal indicators changed over the course of the study. *E. coli* was detected in all samples, and *E. coli* levels ranged between 0.3 and 3.4  $\log_{10}$  MPN/100-mL (Median = 2.3). Canine, avian, human, and ruminant FST markers were detected in <1% (1/196), 4% (8/196), 25% (49/196), and 17% (34/196) of samples, respectively. The average number of copies/100-mL of the avian, human and ruminant FST markers in samples positive for the respective marker were 1,251 (Min. = 64; Max. = 7,040), 1,643 (Min. = 49; Max. = 320,449), and 1,974 (Min. = 145; Max. = 117,490), respectively (Table S4). The association between likelihood of microbial target detection, and (i) detection and  $\log_{10}$  concentration of human and ruminant FST markers, and (ii)  $\log_{10}$  *E. coli* levels were assessed using GLMMs (Table 4). Both *Salmonella* isolation and *eaeA-stx* codetection were positively associated with  $\log_{10}$  *E. coli* levels (Table 4). For each  $\log_{10}$  increase in the MPN of *E. coli*/100-mL the odds of isolating *Salmonella* increased 1.8-fold (Odds Ratio [OR] = 1.8; 95% Confidence Interval [CI] = 1.1, 3.1; Table 4). *Salmonella* isolation and *eaeA-stx* codetection were also strongly and positively associated with the detection and  $\log_{10}$  concentration of human and ruminant FST markers, respectively (Table 4). Specifically, the odds of isolating *Salmonella* approx. doubled (OR = 1.8; 95% CI = 0.91, 3.5) when human FST markers were detected compared to when human

FST markers were not detected in the sample. Similarly, the odds of *eaeA-stx* codetection increased by a factor of 4 (OR = 4.2; 95% CI = 1.3, 13.4) when ruminant FST markers were detected in the samples compared to when ruminant FST markers were not detected. While we failed to find evidence of an association between *L. monocytogenes* isolation and any of the fecal indicators considered, the likelihood of isolating *Listeria* spp. (excluding *monocytogenes*) was negatively associated with *E. coli* levels (OR = 0.50; 95% CI = 0.26, 0.96;  $P = 0.036$ ). Due to the low number of DG3 and GFD positive samples, the association between likelihood of microbial target detection, and detection and  $\log_{10}$  concentration of DG3 and GFD could not be statistically assessed. However, when GFD was present the odds of *Salmonella* isolation and of *eaeA-stx* codetection were both  $\sim 1.5$  times greater compared to when GFD was not detected [*Salmonella* Odds Ratio [OR] =  $1.51 = (4/75)/(4/113)$ ; *eaeA-stx* OR =  $1.54 = (6/127)/(2/65)$ ]. The sole sample positive for DG3 was also positive for *Listeria* spp. (excluding *monocytogenes*).

### Association Between Environmental Factors and Foodborne Pathogen Detection

GLMMs were used to identify potential relationships between individual weather and water quality factors, and likelihood of microbial target detection; the results of these GLMMs are reported in Table 4 and summarized here. While rainfall 0–1 d BSC and 2–3 d BSC were positively associated with *Salmonella* and *Listeria* spp. (excluding *monocytogenes*) isolation, respectively, rainfall 10–20 d BSC was negatively associated with *Listeria* spp. (excluding *monocytogenes*) isolation. For each one cm increase in rainfall 0–1 d before sample collection, the odds of *Salmonella* isolation increased 2.2-fold (OR = 2.2; 95% CI = 1.31, 3.62), while each one cm increase in rainfall 2–3 d before sample collection the odds of *Listeria* spp. (excluding *monocytogenes*) isolation increased 3.3-fold (OR = 3.3; 95% CI = 1.44, 7.51). Likelihood of *Listeria* spp. (excluding *monocytogenes*) and *Salmonella* isolation were both negatively associated with solar radiation and temperature; *eaeA-stx* codetection was also negatively associated with solar radiation. Additionally, likelihood of *Salmonella* isolation was negatively associated with dissolved oxygen levels and pH, while the likelihood of *Listeria* spp. (excluding *monocytogenes*) was positively associated with both factors. For instance, the odds of *Salmonella* isolation decreased 1.4-fold for each one mg/L increase in dissolved oxygen levels (OR = 0.74; 95% CI = 0.72, 0.89), while the odds of *Listeria* spp. (excluding *monocytogenes*) isolation increased 1.3-fold for each one mg/L increase in dissolved oxygen levels (OR = 1.34; 95% CI = 1.06, 1.69). We also found evidence of a potential strong, positive relationship between *eaeA-stx* codetection and flow rate (OR = 2.90; 95% CI = 0.86, 9.77), and of a negative association between *L. monocytogenes* isolation and  $\log_{10}$  turbidity levels (OR = 0.32; 95% CI = 0.09, 1.08).

Multiple spatial factors were also associated with likelihood of microbial target detection (see Table 5). The likelihood of *eaeA-stx* codetection was positively associated with forest-wetland cover and negatively associated with developed land. For instance, for each 1% increase in the amount of forest-wetland cover in the stream corridor the odds of *eaeA-stx* codetection increased 1.02-fold (OR = 1.02; 95% CI = 1.00, 1.05), while for each 1% increase in developed non-open space in the stream corridor the odds of *eaeA-stx* codetection decreased 1.08-fold (OR = 0.93; 95% CI = 0.86, 1.00). While we did not find significant associations ( $P < 0.05$ ) between land use-factors and *Listeria* spp. (excluding

*monocytogenes*), *L. monocytogenes* or *Salmonella* isolation, we did find evidence of potential associations ( $0.05 < P < 0.10$ ) that warrant future investigation. For example, we found evidence of a negative association between likelihood of *L. monocytogenes* isolation and developed non-open space (OR = 0.96; 95% CI = 0.91, 1.01). Similarly, we found evidence of a positive association between likelihood of *Listeria* spp. (excluding *monocytogenes*) isolation and pasture (OR = 1.03; 95% CI = 1.00, 1.07). In addition to land-use factors, we also found significant associations between upstream features, including hydrological factors (e.g., in-stream waterbodies, ditches, stormwater outfalls) and potential sources of fecal contamination (e.g., livestock operations; Table 5). For example, both the odds of *Salmonella* isolation (OR = 2.04;  $P = 0.042$ ) and the odds of *eaeA-stx* codetection (OR = 2.17;  $P = 0.044$ ) approx. doubled when ditches were present upstream of the sampling site. The odds of *Listeria* spp. (excluding *L. monocytogenes*) isolation was negatively associated with multiple livestock-related factors, including the presence of upstream dairy farms (OR = 0.26;  $P = 0.028$ ) and stables (OR = 0.32;  $P = 0.018$ ) upstream.

### Combinations of Factors Associated With an Increased or Decreased Likelihood of Foodborne Pathogen Detection

Conditional inference trees (CTrees) were used to identify and visualize combinations of factors (i.e., specific scenarios) associated with an increased probability of detecting foodborne pathogens in New York agricultural water. CTree results are reported in Figures 2, 3; the CTree where *Listeria* spp. (excluding *L. monocytogenes*) isolation was the outcome did not include any splits and is therefore not reported as a figure. The final CTrees for *Salmonella* isolation and *eaeA-stx* codetection both consisted of a single split (Figure 2). The primary, competitor and first surrogate splits in the *Salmonella* CTree were based on physicochemical water quality at the time of sample collection (Figure 2A). Based on the CTree, the probability of isolating *Salmonella* was highest when dissolved oxygen was below 8.5 mg/L, pH was below 8, or turbidity was above 0.5 log<sub>10</sub> NTUs. The primary split in the *eaeA-stx* CTree was based on week of the year, while the competitor and surrogate splits were based on air temperature (Figure 2B). Specifically, the probability of *eaeA-stx* codetection was highest between Jun. and Oct. (compared to between Apr. and May), and when air temperature 20–30 d BSC was >12°C, air temperature 10–20 d BSC was >24°C, and air temperature 5–10 BSC was >16.8°C. Since air temperature was strongly correlated with week of the year (Figure S2), the association between *eaeA-stx* codetection and week of the year identified using CTree analysis may be a product of seasonal trends in weather (see Figures S1, S2 and Table S5). The *L. monocytogenes* CTree was more complex than either the *Salmonella* or the *eaeA-stx* CTrees. For instance, the *L. monocytogenes* CTree consisted of four interior and five terminal nodes while the *Salmonella* and *eaeA-stx* CTrees consisted of a single interior node and two terminal nodes. Based on the primary splits, the likelihood of isolating *L. monocytogenes* was greatest when avg. solar radiation 2–3 d before sampling was >0.4 mJ/m<sup>2</sup>, the sample was collected after August 19th, and the upstream density of cattle operations was 1 per 10 km<sup>2</sup>. The likelihood of isolating *L. monocytogenes* was lowest when either: (i) avg. solar radiation 2–3 d before sampling was >0.4 mJ/m<sup>2</sup>, the sample was collected after August 19th, and the upstream density of cattle operations was <1 per 10 km<sup>2</sup>, or (ii) avg. solar radiation 2–3 d before sampling was >0.4



mJ/m<sup>2</sup>, the sample was collected before August 19th, and there were no upstream sources of human fecal contamination.

## DISCUSSION

The objectives of this study were to identify (i) potential foodborne pathogen contamination sources using FST markers and geospatial data, and (ii) combinations of spatial, water quality, and weather factors associated with an increased likelihood of detecting foodborne pathogens (*Salmonella* and *L. monocytogenes*), pathogen markers (*eaeA* and *stx* genes), and index organisms for key pathogens (non-pathogenic *Listeria*) in agricultural water samples; we refer to these collectively as microbial targets in the study reported here. As such, this was a hypothesis-generating study, and regression was used to identify factors (i) that were significantly associated with microbial target detection ( $P < 0.05$ ), and (ii) that were not significantly associated with microbial target detection but where a trend that warrants investigation in future studies was present ( $0.05 < P < 0.10$ ). This study is novel due to the diversity of data types used (e.g., weather data, land use data from federal databases, data scraped from Google and government permits, field-collected water quality data), and the computational approaches used to generate these data. For instance, this study calculated flow path distances that account for topography and represent the physical distance a contaminant travels from its source to the sampling site; the Euclidean distance measures used in previous studies (e.g., Strawn et al., 2013a; Weller et al., 2016) do not capture this complexity. However, it is also important to recognize the limitations associated with spatial data. While most of the spatial data is inherently comprehensive (e.g., the wastewater discharge site data includes all sites in NY since the data were generated using NY State permit data) and spatially explicit (e.g., wastewater discharge sites exist as a single point), this is not true for the livestock, campground, or trailer park data. These three data types are based on addresses, and as such, these features, which can cover large land areas (e.g., one campground in the study area is 100 acres), are reduced to single points. This may result in underestimation of feature density, and overestimation of min. flow path distances. For example, dairy farms with addresses outside the watershed but with pastures inside the watershed would be considered absent. As such, the failure to identify significant ( $P < 0.05$ ) associations between livestock operation, trailer park, and campground factors with microbial target detection does not prove a lack of association. Despite this limitation, this study was able to identify potential associations between microbial target detection and these landscape features; as such, these associations should be explored in future studies once more accurate datasets are available. Overall, the integration of diverse data types, as well as the methods used for obtaining said data, is novel and provides a blueprint for how such data and approaches can be used in future studies. This study also illustrates how robust, non-parametric statistical approaches (e.g., conditional inference trees) can be used to investigate how interactions between correlated environmental factors drive foodborne pathogen contamination of preharvest environments.



## Observed *Listeria* Species Prevalence Was Significantly Lower When Modified Moore Swabs, as Opposed to 0.45 µm Filters, Were Used to Process Grab Samples

In this study, we tested all water samples for the presence of *Listeria*, *Salmonella*, and the *eaeA* and *stx* genes (molecular markers associated with the potential presence of pathogenic *E. coli*). While it is difficult to compare pathogen prevalence between this and past studies (Strawn et al., 2013a,b; Chapin et al., 2014; Weller et al., 2015a,b; Falardeau et al., 2017) due to the larger volume of water collected here (10-L) compared to these studies (between 250 and 532-mL), one would expect a higher prevalence in the current study due to the larger amount of water collected. While the *Salmonella* prevalence observed here was substantially higher than the *Salmonella* prevalence observed in these past studies, it is surprising that the *Listeria* prevalence observed here was substantially lower than the *Listeria* prevalence in these past studies (Strawn et al., 2013a,b; Chapin et al., 2014; Weller et al., 2015a,b). While we isolated *L. monocytogenes* from 10% of samples, studies that sampled the same waterways as the study reported here isolated *L. monocytogenes* from 71% (15/21; unpublished) and 63% [33/52; (Weller et al., 2015a)] of samples collected in 2013 and 2014, respectively. The only methodological difference between our study and these past studies was the larger volume of water collected here, which necessitated filtration through a mMS instead of a 0.45 µm filter. Thus, we used the last 29 samples collected to compare *Listeria* recovery using paired mMS and 0.45 µm filters. While our analyses are limited by the small sample size, our data indicate that *Listeria* recovery using mMS was significantly lower than *Listeria* recovery using 0.45 µm filters, even though 9 times as much water was filtered through the mMS as opposed to 0.45 µm filters. In fact, only 14% (2/7) samples that were identified as *L. monocytogenes*-positive using filters were also identified as *L. monocytogenes*-positive using mMS. Since mMS work by capturing large particles (e.g., sediment) to which bacteria are attached, one explanation for the lower than expected *Listeria* prevalence in our study may be differences in attachment mechanisms between *Listeria* and the enteric bacteria species used to validate the mMS approach (Bisha et al., 2011, 2014; McEgan et al., 2013b; Sbodio et al., 2013; Zhu et al., 2019). Indeed, differences in bacterial surface structures, hydrophobicity, surface charge, cell size, and cell sphericity have been shown to affect bacterial attachment to sediment and other surfaces (Faille et al., 2002; Ukuku and Fett, 2002; Vorst et al., 2004; Walker et al., 2005; Silva et al., 2008; Wan Norhana et al., 2009; Liao et al., 2015). One study found a strong correlation between cell surface hydrophobicity and the strength of *Listeria*, *Salmonella*, and *E. coli* attachment to cantaloupe rinds, with *Salmonella* and *E. coli* attaching more strongly to the rind than *Listeria*. Given our findings, follow-up studies are needed to (i) determine why *Listeria* recovery is lower using mMS compared to 0.45 µm filters, and (ii) how the mMS approach can be adapted to facilitate *Listeria* recovery. Such studies are essential if mMS are to be incorporated into water testing programs as previously suggested (Bisha et al., 2014). Due to the use of mMS in the present study, the results of analyses where the likelihood of *Listeria* isolation was the outcome need to be interpreted in the context of the sampled population. Thus, factors identified as significant here should be considered associated with *Listeria* isolation from mMS as opposed to *Listeria* isolation from water samples.

## ***Salmonella* Isolation and *eaeA-stx* Codetection Were Associated With Human and Ruminant Fecal Contamination, Respectively**

In the present study, we found evidence of a strong, positive association between ruminant FST markers (Rum2Bac) and *eaeA-stx* codetection, and between human FST markers (HF183) and *Salmonella* isolation. In general, these findings are consistent with past studies that found strong associations between ruminant fecal contamination and detection of pathogenic *E. coli* markers (Walters et al., 2007; Petit et al., 2017), and between human fecal contamination and *Salmonella* detection (Marti et al., 2013; Liang et al., 2015; Stea et al., 2015). For example, Stea et al. (2015) found that the odds of detecting molecular markers for *Salmonella* in Nova Scotia surface water samples was 2.2 times greater when human FST markers were present as opposed to when human FST markers were not detected. Bradshaw et al. (2016) reported that a model containing log<sub>10</sub> ruminant FST marker (Rum2Bac) concentration and water temperature was able to identify 100% of *stx*-positive water samples collected in Georgia, USA. Bradshaw et al. (2016) also reported that the odds of *stx* detection increased by a factor of 2 for each log<sub>10</sub> increase in ruminant FST marker concentration, which is similar to the odds ratio calculated here (OR = 1.6; 95% CI = 1.1, 2.4). The identification of an association between host-associated FST markers, and *Salmonella* and *eaeA-stx* codetection is also consistent with the associations between spatial factors and microbial target detection identified here and in past studies (Sassoubre et al., 2011; Walters et al., 2011; Wilkes et al., 2011; Bradshaw et al., 2016). For instance, Wilkes et al. (2011) identified a positive association between pasture being present 0–5 km upstream of a sampling site, and an increased likelihood of isolating *E. coli* O157:H7 from water samples collected in Ontario, Canada. Interestingly, in the present study, we did not find associations between upstream agricultural land and *eaeA-stx* codetection but did find associations between *eaeA-stx* codetection and forest-wetland cover. The lack of an association between upstream agricultural land use and *eaeA-stx* codetection as well as the identification of an association between forest-wetland cover and *eaeA-stx* codetection may indicate that the ruminant fecal contamination detected in the present study is of cervid as opposed to bovine origin. This is supported by the fact that (i) multiple *E. coli* outbreaks have been attributed to deer intrusion into recreational and farm environments (Cody et al., 1999; Feldman et al., 2002; Laidler et al., 2013), and (ii) past studies have isolated pathogenic *E. coli* from deer feces (Rice et al., 1995; Dunn et al., 2004; Díaz et al., 2011). Despite the potential association between deer fecal contamination and *eaeA-stx* detection in agricultural water samples, the authors are not recommending the removal or alteration of upstream habitat to reduce deer populations. In fact, based on other associations identified here, conversion of forest-wetland cover to developed or agricultural land uses would increase the likelihood of *Salmonella* or *L. monocytogenes* being present in the waterway. This highlights the potential for unintended consequences (increased risk of *Salmonella* or *L. monocytogenes* contamination) when food safety management practices (removal of forest-wetland cover) are implemented with a single aim in mind (such as reducing contamination by pathogenic *E. coli*).

*Salmonella* isolation was positively associated with both human fecal contamination and proxies for increased human presence in upstream areas [proximity to campgrounds, developed open space (e.g., parks)], which is consistent with the existing literature (e.g.,

Christensen et al., 1978; Varness et al., 1978; Dasher et al., 1981; Hendry and Leggatt, 1982; Flack et al., 1988; Sassoubre et al., 2011; Walters et al., 2011; Vereen et al., 2013). Previous research has shown that increases in recreational activities (e.g., camping, hiking, swimming) can affect the microbial quality of downstream surface water sources (Christensen et al., 1978; Varness et al., 1978; Dasher et al., 1981; Hendry and Leggatt, 1982; Flack et al., 1988), which may explain the association between proximity to campgrounds and *Salmonella* isolation reported here. Interestingly, we also found a positive association between proximity to poultry operations and *Salmonella* isolation; although too few samples tested positive for the avian FST marker to perform statistical analyses, the odds of detecting *Salmonella* when the avian FST marker was present was 1.5 times greater than when the marker was not detected. Overall, the association between avian fecal sources and *Salmonella* isolation is consistent with previous studies (e.g., Vereen et al., 2013; Hernandez et al., 2016; Steele et al., 2018) as well as the fact that *Salmonella* is a well-known contaminant of poultry production systems (Louis et al., 1988; Rodrigue et al., 1990; Behraves et al., 2014). In a survey of *Salmonella* prevalence in the Satilla River Basin, Vereen et al. (2013) found that sources of poultry and human fecal contamination were positively associated with *Salmonella* isolation from water samples; frequency of *Salmonella* detection was two-times greater in watersheds with poultry operations compared to watersheds without poultry operations. Overall, our findings suggest that one approach to managing the food safety risks associated with *Salmonella* contamination of preharvest surface water in NY is to address upstream sources of human and poultry fecal contamination. Such efforts could include identifying and addressing failing infrastructure upstream of irrigation pumps or treating irrigation water sources downstream of campgrounds, poultry farms or other fecal sources.

Our failure to find evidence of an association between *L. monocytogenes* and fecal indicator bacteria (FIBs, e.g., *E. coli*) is not unexpected as past studies also failed to find positive associations between *L. monocytogenes* and FIBs in water (e.g., Schaffter et al., 2004; Wilkes et al., 2009; Economou et al., 2013; Weller et al., 2020). For instance, in their survey of the microbial quality of the South Nation River Basin, Wilkes et al. (2009) found a negative association between the likelihood of *L. monocytogenes* isolation and *E. coli* levels. While expected, our failure to identify an association between *L. monocytogenes* isolation and fecal indicators is interesting given the identification of positive associations between *L. monocytogenes* isolation and sources of human (e.g., campgrounds, wastewater discharge sites) and livestock (e.g., dairy farms) fecal contamination in this and other studies (Watkins and Sleath, 1981; Dijkstra, 1982; Paillard et al., 2005; Lyautey et al., 2007; Odjadjare et al., 2010; Strawn et al., 2013a; Weller et al., 2016). The failure to identify an association between *L. monocytogenes* and fecal indicators but the identification of an association between *L. monocytogenes* and sources of fecal contamination may be due to the fact *L. monocytogenes* is a microbe adapted to living in non-host environments (Vivant et al., 2013) but the FIBs used here are not (Lee et al., 2008; Bae and Wuertz, 2009). Thus, even if *L. monocytogenes* contamination is initially of fecal origin, *L. monocytogenes* may not correlate to presence or levels of fecal indicators since *L. monocytogenes* can persist in non-host environments longer than the indicators. Overall, these findings are illustrative of the

need for alternatives to traditional indicator-based water quality tests for identifying when and where *L. monocytogenes* contamination of agricultural water sources is likely to occur.

### **Rain Events, Ditches, and Other Factors Associated With Increased Stream Inflows Were Positively Associated With Pathogen Detection and Appeared to Mediate the Relationship Between Land Use and *L. monocytogenes* Detection**

Multiple factors included in the analyses conducted here were hydrological in nature (e.g., flow rate), or influenced stream hydrology and inflows (e.g., presence of upstream ditches, rainfall). Interestingly, we found evidence of associations between these “hydrological” factors and detection of all targets in the present study. For example, the odds of *eaeA-stx* codetection was approx. two times greater when a ditch intersected the stream channel near the sample site compared to when no ditch was present. Similarly, we found evidence that increased rainfall was associated with an enhanced risk of *Salmonella*, *Listeria* spp. (excluding *monocytogenes*), and *L. monocytogenes* detection. Overall, these findings are not unexpected since past studies have found associations between “hydrological” factors and pathogen isolation (e.g., Goyal et al., 1977; Baudart et al., 2000; Kistemann et al., 2002; Cooley et al., 2007; Haley et al., 2009; Wilkes et al., 2009, 2011; Walters et al., 2011; Luo et al., 2015; Harris et al., 2018). A study that examined associations between environmental factors and pathogen isolation from Ontario river water only detected *Salmonella* when surface water discharge rates were elevated, which led the authors to conclude that events that promote off-farm and in-stream transfer of microbes (e.g., rain events) must occur for detection of *Salmonella* in the sampled rivers (Wilkes et al., 2011). Overall, the association between hydrological factors and microbial target detection is logical since hydrological factors are associated with processes that facilitate pathogen movement from terrestrial to freshwater environments. Falbo et al. (2013) investigated the role of ditches in bacterial loading of central New York waterways and found that, on average, 22% of a watershed’s area drains to roadside ditches, which capture and transport runoff and associated-bacteria to waterways from road surfaces and adjacent land areas. Falbo et al. (2013) also found that *E. coli* was capable of surviving in ditch sediments during dry periods and of re-entering the water column following resuspension of these sediments during rain events, suggesting ditches are a source of bacterial contaminants as well as a conduit for bacterial movement. Thus, efforts focused on reducing pathogen movement and survival in ditches may provide one set of strategies for preventing foodborne pathogens from entering agricultural water sources. Rain events can also physically transport pathogens from terrestrial sources to freshwater systems through overland run-off and facilitate the release of pathogens from soil, feces and other matrices so pathogens are available for transport from terrestrial sources to freshwater systems (Thelin and Gifford, 1983; Muirhead et al., 2004, 2005; Guber et al., 2006, 2007, 2015; Boyer, 2008). The key role of rain (and associated increases in run-off and instream flow) in bacterial release and transport is evidenced by the repeated inclusion of rainfall as a model parameter (Guber et al., 2009, 2011, 2013) or an experimental condition in studies investigating bacterial release from terrestrial sources and transport to freshwater systems (Thelin and Gifford, 1983; Muirhead et al., 2004, 2005; Guber et al., 2006, 2007, 2015; Boyer, 2008). In the *L. monocytogenes* CTree reported here, rain and hydrological factors appear to mediate the relationship between land use and *L. monocytogenes* isolation. The *L. monocytogenes* CTree implies that, for the sampled

streams, a high upstream dairy density was associated with an increased likelihood of *L. monocytogenes* isolation only when there was a recent rain event. While these findings may seem obvious, they suggest several strategies for managing food safety hazards in agricultural water. For instance, if potential sources of fecal contamination (e.g., dairy farms) or features that facilitate pathogen movement from terrestrial sources to agricultural water sources (e.g., ditches, culverts) are present upstream, growers may want to treat water from these sources prior to use or use an alternative water source following rain events. As suggested by past studies (e.g., Diaz et al., 2010, 2012), impeding overland flow from potential sources of pathogen contamination to agricultural waterways (e.g., by constructing wetlands and vegetated buffers, changing ditch hydrology to facilitate settling out bacteria) may help mitigate the food safety risks associated with preharvest water use. However, additional research is needed to (i) determine how such strategies could be incorporated into a larger comanagement framework, and (ii) experimentally quantify the risk reduction (if any) associated with implementations of such measures.

### **Weather and Physicochemical Water Quality Appear to Be Key Factors Associated With *eaeA-stx* and *Salmonella* Detection, Respectively**

The majority of splits in the *Salmonella* CTree were based on physicochemical water quality; water quality factors were also associated with the greatest change in odds of *Salmonella* isolation according to regression analysis. Similarly, air temperature comprised the competitor and surrogate splits in the *eaeA-stx* CTree, while flow rate (a proxy for unmeasured upstream rainfall) was more strongly associated with the likelihood of *eaeA-stx* codetection than any other factor except detection of ruminant FST markers according to regression analysis. These findings suggest that physicochemical water quality and weather are key factors associated with *Salmonella* isolation and *eaeA-stx* detection, respectively. While past studies also found significant associations between temperature and pathogenic *E. coli* detection, the direction of these associations varied between studies (Wilkes et al., 2011; Francy et al., 2013; Luo et al., 2015; Bradshaw et al., 2016; Truchado et al., 2018; Weller et al., 2020). For example, Wilkes et al. (2011) found that temperatures above 14 C were associated with a reduced likelihood of isolating *E. coli* O157:H7 from water samples collected in eastern Ontario, while the present study found a strong positive association between *eaeA-stx* detection and temperature. Moreover, a 2017 study that characterized factors associated with temporal variability in the microbial quality of New York streams and Arizona canals, reported a polynomial relationship between *eaeA-stx* detection and air temperature in both states (Weller et al., 2020). Indeed, the existence of a complex (e.g., polynomial) relationship between temperature and *eaeA-stx* codetection could explain why we found evidence of a significant association between temperature and *eaeA-stx* codetection using CTree but not regression analysis. While the nature of the temperature-pathogenic *E. coli* relationship varies in direction and shape between studies, the relationships between *Salmonella* isolation and dissolved oxygen (Liang et al., 2015; Bradshaw et al., 2016; Weller et al., 2020), pH (Liang et al., 2015; Bradshaw et al., 2016; Weller et al., 2020), and turbidity (Liang et al., 2015; Bradshaw et al., 2016; Partyka et al., 2018; Weller et al., 2020) appear to be consistent across studies. As part of their study on the association between fecal indicators and human pathogens in tropical surface waters, Liang et al. (2015) reported a positive correlation between *Salmonella* levels and turbidity.



Similarly, the 2017 study discussed above identified a strong, positive monotonic relationship between *Salmonella* detection and turbidity in New York and Arizona, and a strong, negative monotonic relationship between *Salmonella* detection and dissolved oxygen levels in Arizona (Weller et al., 2020). Overall the findings of this and other studies (Liang et al., 2015; Bradshaw et al., 2016; Partyka et al., 2018; Weller et al., 2020) suggest that physicochemical water quality factors could be incorporated as supplementary indicators into traditional microbial indicator-based water quality monitoring programs, as suggested by past studies (Weller et al., 2020). For example, growers concerned about *Salmonella* may want to measure pH or turbidity levels in agricultural water sources immediately prior to use; if pH is low or turbidity is high, the grower may want to use an alternative source or treat the water. Integration of physicochemical water quality parameters into water testing would provide information growers can use to guide decision-making in real-time; however additional research is needed to clarify how exactly physicochemical water quality data can be integrated with traditional microbial testing to inform hazard management efforts.

## CONCLUSION

Due to the number of streams sampled ( $N=68$ ), this study is one of the most geographically widespread surveys of the food safety hazards present in New York agricultural water to-date, and therefore provides key insights into pathogen dynamics in New York freshwater environments. Using a diversity of data types and advanced geographic analyses this study identified combinations of environmental conditions (e.g., rain events were associated with *L. monocytogenes* isolation from samples collected downstream of dairy farms) associated with an increased likelihood of pathogen presence in agricultural water. As such, this study provides data that will enable growers to better manage the food safety risks associated with preharvest surface water use. Specifically, the associations identified here can help New York growers qualitatively assess the food safety hazards associated with preharvest water use for individual water sources and in real-time. For example, we found evidence that likelihood of *Salmonella* isolation was positively associated with the sources of human fecal contamination, and with physicochemical water quality. These findings suggest several hazard management strategies, including identifying and addressing failing infrastructure upstream of irrigation pumps or incorporating physicochemical water quality factors as supplementary indicators into traditional microbial indicator-based water quality monitoring programs. Our study also highlights the potential for unintended consequences when food safety management practices are implemented with a single aim in mind. For example, although we found evidence that *eaeA-stx* codetection was associated with forest-wetland cover, our findings also suggest conversion of forest-wetland cover to developed or agricultural land cover could increase the potential for *Salmonella* or *L. monocytogenes* contamination. Given the strong association between hydrological factors and detection of all microbial targets in the present study, strategies for reducing overland flow from potential sources of fecal contamination (e.g., dairy farms) to agricultural water sources may help mitigate bacterial contamination of agricultural water sources as suggested in previous studies (Díaz et al., 2010, 2012). However, additional research is needed to determine (i) how such strategies could be incorporated into a larger comanagement framework, and (ii) if implementation of these strategies could have unexpected, detrimental impacts on food



safety and/or environmental health as suggested above. Overall, this is a hypothesis-generating study that can guide future research. For example, the associations identified here (e.g., between *Salmonella* and physicochemical water quality) can be used to develop models to predict when and where a specific water source has an enhanced risk of being contaminated by a given pathogen. Our findings can also be used to guide efforts to quantify trade-offs associated with implementing specific food safety practices on-farm (e.g., removing forest-wetland cover to reduce risks associated with pathogenic *E. coli* contamination). Moreover, the methods used to obtain (e.g., scraping Google, government permits) and analyze (e.g., flow path distance as opposed to Euclidean) the geospatial data in this study, are novel and can serve as a template for the integration of these data types into future studies, including in other produce-growing regions.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

We are grateful for the technical assistance of Maureen Gunderson, Tanzy Love, Sriya Sunil, Ahmed Gaballa, Kayla Ferris, and Julia Muuse. We also thank Ariana Fenty and Maxwell Wilder for their assistance with the FST qPCR analysis.

### FUNDING

Data collection and cleaning were funded by a grant from the Center for Produce Safety under award number 2017CPS09. Data analysis and manuscript preparation were supported by the National Institute of Environmental Health Sciences of the National Institutes of Health (NIH) under award number T32ES007271. The content is solely the responsibility of the authors and does not represent the official views of the NIH.

## REFERENCES

- Antaki EM, Vellidis G, Harris C, Aminabadi P, Levy K, and Jay-Russell MT (2016). Low concentration of *Salmonella enterica* and generic *Escherichia coli* in farm ponds and irrigation distribution systems used for mixed produce production in southern Georgia. *Foodborne Pathog. Dis* 13, 551–558. doi: 10.1089/fpd20162117 [PubMed: 27400147]
- Bae S, and Wuertz S (2009). Discrimination of viable and dead fecal Bacteroidales bacteria by quantitative PCR with propidium monoazide. *Appl. Environ. Microbiol* 75, 2940–2944. doi: 10.1128/AEM.01333-08 [PubMed: 19270114]
- Baudart J, Grabulos J, Barusseau J-P, and Lebaron P (2000). *Salmonella* spp. and fecal coliform loads in coastal waters from a point vs. nonpoint source of pollution. *J. Environ. Qual* 29:241. doi: 10.2134/jeq2000.00472425002900010031x
- Behravesh CB, Brinson D, Hopkins BA, and Gomez TM (2014). Backyard poultry flocks and salmonellosis: a recurring, yet preventable public health challenge. *Clin. Infect. Dis* 58, 1432–1438. doi: 10.1093/cid/ciu067 [PubMed: 24501387]
- Benjamin L, Atwill ER, 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 on the central California coast. *Int. J. Food. Microbiol* 165, 65–76. doi: 10.1016/j.ijfoodmicro.2013.04.003 [PubMed: 23697918]
- Bernhard AE, and Field KG (2000). A PCR assay To discriminate human and ruminant feces on the basis of host differences in Bacteroides-Prevotella genes encoding 16S rRNA. *Appl. Environ. Microbiol* 66, 4571–4574. doi: 10.1128/AEM.66.10.4571-4574.2000 [PubMed: 11010920]

- Bisha B, Adkins JA, Jokerst JC, Chandler JC, Perez-Mendez A, Coleman SM, et al. (2014). Colorimetric paper-based detection of *Escherichia coli*, *Salmonella* spp., and *Listeria monocytogenes* from large volumes of agricultural water. *J Vis Exp*. 88:e51414. doi: 10.3791/51414
- Bisha B, Perez-Mendez A, Danyluk MD, and Goodridge LD (2011). Evaluation of modified moore swabs and continuous flow centrifugation for concentration of *Salmonella* and *Escherichia coli* O157:H7 from large volumes of water. *J. Food Prot* 74, 1934–1937. doi: 10.4315/0362-028X.JFP-11-133 [PubMed: 22054197]
- Boyer DG (2008). Fecal coliform dispersal by rain splash on slopes. *Agric. For. Meteorol* 148, 1395–1400. doi: 10.1016/j.agrformet.2008.04.001
- Bradshaw JK, Snyder BJ, Oladeinde A, Spidle D, Berrang ME, Meinersmann RJ, 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 [PubMed: 27295624]
- Bundrant BN, Hutchins T, den Bakker HC, Fortes E, and Wiedmann M (2011). Listeriosis outbreak in dairy cattle caused by an unusual *Listeria monocytogenes* serotype 4b strain. *J. Vet. Diagn. Invest* 23, 155–158. doi: 10.1177/104063871102300130 [PubMed: 21217050]
- Byappanahalli MN, Shively DA, Nevers MB, Sadowsky MJ, and Whitman RL (2003). Growth and survival of *Escherichia coli* and enterococci populations in the macro-alga *Cladophora* (Chlorophyta). *FEMS Microbiol. Ecol* 46, 203–211. doi: 10.1016/S0168-6496(03)00214-9 [PubMed: 19719574]
- Chapin TK, Nightingale KK, Worobo RW, Wiedmann M, and Strawn LK (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 [PubMed: 25364926]
- Christensen H, Pacha R, Varness K, and Lapen R (1978). “Human use in a dispersed recreation area and its effect on water quality,” in *Proceedings, Recreational Impact on Wildlands* (Seattle, WA: U.S. Department of Agriculture), 107–119.
- Cody SH, Glynn MK, Farrar JA, Cairns KL, Griffin PM, Kobayashi J, et al. (1999). An outbreak of *Escherichia coli* O157:H7 infection from unpasteurized commercial apple juice. *Ann. Intern. Med* 130, 202–209. doi: 10.7326/0003-4819-130-3-199902020-00005 [PubMed: 10049198]
- Collins R, and Rutherford K (2004). Modelling bacterial water quality in streams draining pastoral land. *Water Res.* 38, 700–712. doi: 10.1016/j.watres.2003.10.045 [PubMed: 14723940]
- Cooley M, Carychao D, Crawford-Miksza L, Jay MT, Myers C, Rose C, et al. (2007). Incidence and tracking of *Escherichia coli* O157:H7 in a major produce production region in California. *PLoS ONE* 2:e1159. doi: 10.1371/journal.pone.0001159 [PubMed: 18174909]
- Cooley MB, Quinones B, Oryang D, Mandrell RE, and 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* 4:30. doi: 10.3389/fcimb.2014.00030 [PubMed: 24624367]
- Dasher D, Urban L, Dvoracek M, and Fish E (1981). Effects of recreation on water quality in Guadalupe Mountains National Park. *Trans. ASAE* 24, 1181–1187. doi: 10.13031/2013.34417
- Den Bakker HC, Bundrant BN, Fortes ED, Orsi RH, and Wiedmann M (2010). A population genetics-based and phylogenetic approach to understanding the evolution of virulence in the genus *Listeria*. *Appl. Environ. Microbiol* 76, 6085–6100. doi: 10.1128/AEM.00447-10 [PubMed: 20656873]
- Díaz FJ, O’Geen AT, and Dahlgren RA (2010). Efficacy of constructed wetlands for removal of bacterial contamination from agricultural return flows. *Agric. Water Manage* 97, 1813–1821. doi: 10.1016/j.agwat.2010.06.015
- Díaz FJ, O’Geen AT, and Dahlgren RA (2012). Agricultural pollutant removal by constructed wetlands: implications for water management and design. *Agric. Water Manage* 104, 171–183. doi: 10.1016/j.agwat.2011.12.012
- Díaz S, Vidal D, Herrera-Leon S, and Sanchez S (2011). O157:H7 in free-ranging red deer in south-central Spain. *Foodborne Pathog. Dis* 8, 1313–1315. doi: 10.1089/fpd.2011.0923 [PubMed: 21819212]

- Dijkstra RG (1982). The occurrence of *Listeria monocytogenes* in surface water of canals and lakes, in ditches of one big polder and in the effluents and canals of a sewage treatment plant. *Zentralbl. Bakteriologie. Mikrobiologie. Hygiene B* 176, 202–205. [PubMed: 6812325]
- Dunn JR, Keen JE, Moreland D, and Alex T (2004). Prevalence of *Escherichia coli* O157:H7 in white-tailed deer from Louisiana. *J. Wildl. Dis* 40, 361–365. doi: 10.7589/0090-3558-40.2.361 [PubMed: 15362843]
- Economou V, Gousia P, Kansouzidou A, Sakkas H, Karanis P, and Papadopoulou C (2013). Prevalence, antimicrobial resistance and relation to indicator and pathogenic microorganisms of *Salmonella enterica* isolated from surface waters within an agricultural landscape. *Int. J. Hyg. Environ. Health* 216, 435–444. doi: 10.1016/j.ijheh.2012.07.004 [PubMed: 22901425]
- Ederberg SC, Rice EW, Karlin RJ, and Allen MJ (2000). *Escherichia coli*: the best biological drinking water indicator for public health protection. *J. Appl. Microbiol* 88, 106S–116S. doi: 10.1111/j.1365-2672.2000.tb05338.x
- Faillie C, Jullien C, Fontaine F, Bellon-Fontaine M-N, Slomianny C, and Benezech T (2002). Adhesion of *Bacillus* spores and *Escherichia coli* cells to inert surfaces: role of surface hydrophobicity. *Can. J. Microbiol* 48, 728–738. doi: 10.1139/w02-063 [PubMed: 12381029]
- Falardeau J, Johnson RP, Pagotto F, and 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:e0185437. doi: 10.1371/journal.pone.0185437 [PubMed: 28953937]
- Falbo K, Schneider RL, Buckley DH, Walter MT, Bergholz PW, and Buchanan BP (2013). Roadside ditches as conduits of fecal indicator organisms and sediment: implications for water quality management. *J. Environ. Manage* 128, 1050–1059. doi: 10.1016/j.jenvman.2013.05.021 [PubMed: 23933218]
- FDA (2019). Investigation Summary: Factors Potentially Contributing to the Contamination of Romaine Lettuce Implicated in the Fall 2018 Multi-State Outbreak of *E. coli* O157:H7 (2019). Available online at: <https://www.fda.gov/food/outbreaks-foodborne-illness/investigation-summary-factors-potentially-contributing-contamination-romaine-lettuce-implicated-fall#Factors>
- Feldman KA, Mohle-Boetani JC, Ward J, Furst K, Abbott SL, Ferrero DV, et al. (2002). A cluster of *Escherichia coli* O157: nonmotile infections associated with recreational exposure to lake water. *Public Health Rep.* 117, 380–385. doi: 10.1016/S0033-3549(04)50175-9 [PubMed: 12477920]
- Ferguson CM, Davies CM, Kaucner C, Krogh M, Rodehutsors J, Deere DA, et al. (2007). Field scale quantification of microbial transport from bovine faeces under simulated rainfall events. *J. Water Health* 5, 83–95. doi: 10.2166/wh.2006.050 [PubMed: 17402281]
- Flack JE, Medine AJ, and Hansen-Bristow KJ (1988). Stream water quality in a mountain recreation area. *Mt. Res. Dev* 8:11. doi: 10.2307/3673402
- Food and Drug Administration (2015). Standards for the Growing, Harvesting, Packing, and Holding of Produce for Human Consumption. Food Safety Modernization Act.
- Francy DS, Stelzer EA, Duris JW, Brady AMG, Harrison JH, Johnson HE, et al. (2013). Predictive models for *Escherichia coli* concentrations at inland lake beaches and relationship of model variables to pathogen detection. *Appl. Environ. Microbiol* 79, 1676–1688. doi: 10.1128/AEM.02995-12 [PubMed: 23291550]
- Gelting RJ, Baloch MA, Zarate-Bermudez M, Hajmeer MN, Yee J, Brown T, et al. (2015). A systems analysis of irrigation water quality in an environmental assessment of an *E. coli* O157:H7 outbreak in the United States linked to iceberg lettuce. *Agric. Water Manage* 150, 111–118. doi: 10.1016/j.agwat.2014.12.002
- Gelting RJ, Baloch MA, Zarate-Bermudez MA, and Selman C (2011). Irrigation water issues potentially related to the 2006 multistate *E. coli* O157:H7 outbreak associated with spinach. *Agric. Water Manage* 98, 1395–1402. doi: 10.1016/j.agwat.2011.04.004
- Goto DK, and Yan T (2011). Genotypic diversity of *Escherichia coli* in the water and soil of tropical watersheds in Hawaii. *Appl. Environ. Microbiol* 77, 3988–3997. doi: 10.1128/AEM.02140-10 [PubMed: 21515724]

- Goyal SM, Gerba CP, Melnick JL, Bonsdorff C-H, von, Torvela N, Heikinheimo A, et al. (1977). Occurrence and distribution of bacterial indicators and pathogens in canal communities along the Texas coast. *Appl. Environ. Microbiol* 34, 139–149. [PubMed: 20838]
- Green HC, Dick LK, Gilpin B, Samadpour M, and Field KG (2012). Genetic markers for rapid PCR-based identification of gull, Canada goose, duck, and chicken fecal contamination in water. *Appl. Environ. Microbiol* 78, 503–510. doi: 10.1128/AEM.05734-11 [PubMed: 22081573]
- Green HC, Haugland RA, Varma M, Millen HT, Borchardt MA, Field KG, et al. (2014a). Improved HF183 quantitative real-time PCR assay for characterization of human fecal pollution in ambient surface water samples. *Appl. Environ. Microbiol* 80, 3086–3094. doi: 10.1128/AEM.04137-13 [PubMed: 24610857]
- Green HC, White KM, Kelty CA, and Shanks OC (2014b). Development of rapid canine fecal source identification PCR-based assays. *Environ. Sci. Technol* 48, 11453–11461. doi: 10.1021/es502637b [PubMed: 25203917]
- Green HC, Wiedmann M, and Weller DL (2019). Microbial source-tracking reveals origins of fecal contamination in a recovering watershed. *Water* 11:2162. doi: 10.3390/w11102162
- Guber AK, Fry J, Ives RL, and Rose JB (2015). *Escherichia coli* survival in, and release from, white-tailed deer feces. *Appl. Environ. Microbiol* 81, 1168–1176. doi: 10.1128/AEM.03295-14 [PubMed: 25480751]
- Guber AK, Karns JS, Pachepsky YA, Sadeghi AM, Van Kessel JS, and Dao TH (2007). Comparison of release and transport of manure-borne *Escherichia coli* and enterococci under grass buffer conditions. *Lett. Appl. Microbiol* 44, 161–167. doi: 10.1111/j.1472-765X.2006.02065.x [PubMed: 17257255]
- Guber AK, Pachepsky YA, Dao TH, Shelton DR, and Sadeghi AM (2013). Evaluating manure release parameters for nonpoint contaminant transport model KINEROS2/STWIR. *Ecol. Modell* 263, 126–138. doi: 10.1016/j.ecolmodel.2013.05.008
- Guber AK, Pachepsky YA, Yakirevich AM, Shelton DR, Sadeghi AM, Goodrich DC, et al. (2011). Uncertainty in modelling of faecal coliform overland transport associated with manure application in Maryland. *Hydrol. Process* 25, 2393–2404. doi: 10.1002/hyp.8003
- Guber AK, Shelton DR, Pachepsky YA, Sadeghi AM, and Sikora LJ (2006). Rainfall-induced release of fecal coliforms and other manure constituents: comparison and modeling. *Appl. Environ. Microbiol* 72, 7531–7539. doi: 10.1128/AEM.01121-06 [PubMed: 17028232]
- Guber AK, Yakirevich AM, Sadeghi AM, Pachepsky YA, and Shelton DR (2009). Uncertainty evaluation of coliform bacteria removal from vegetated filter strip under overland flow condition. *J. Environ. Qual* 38:1636. doi: 10.2134/jeq2008.0328 [PubMed: 19549940]
- Haley BJ, Cole DJ, and Lipp EK (2009). Distribution, diversity, and seasonality of waterborne salmonellae in a rural watershed. *Appl. Environ. Microbiol* 75, 1248–1255. doi: 10.1128/AEM.01648-08 [PubMed: 19124594]
- Hamilton MJ, Hadi AZ, Griffith JF, Ishii S, and Sadowsky MJ (2010). Large scale analysis of virulence genes in *Escherichia coli* strains isolated from Avalon Bay, CA. *Water Res* 44, 5463–5473. doi: 10.1016/j.watres.2010.06.058 [PubMed: 20643468]
- Harris CS, Tertuliano M, Rajeev S, Vellidis G, and Levy K (2018). Impact of storm runoff on Salmonella and *Escherichia coli* prevalence in irrigation ponds of fresh produce farms in southern Georgia. *J. Appl. Microbiol* 124, 910–921. doi: 10.1111/jam.13689 [PubMed: 29316043]
- Harwood VJ, Levine AD, Scott TM, Chivukula V, Lukasik J, Farrah SR, et al. (2005). Validity of the indicator organism paradigm for pathogen reduction in reclaimed water and public health protection. *Appl. Environ. Microbiol* 71, 3163–3170. doi: 10.1128/AEM.71.6.3163-3170.2005 [PubMed: 15933017]
- Havelaar AH, Vazquez KM, Topalcengiz Z, Muñoz-Carpena, and R., and Danyluk MD, (2017). Evaluating the U.S. Food Safety Modernization Act produce safety rule standard for microbial quality of agricultural water for growing produce. *J. Food Prot* 80, 1832–1841. doi: 10.4315/0362-028X.JFP-17-122
- Hendricks C (1967). Multiplication and growth of selected enteric bacteria in clear mountain stream water. *Water Res.* 1, 567–576. doi: 10.1016/0043-1354(67)90039-5

- Hendry GS, and Leggatt EA (1982). Some effects of shoreline cottage development on lake bacteriological water quality. *Water Res.* 16, 1217–1222. doi: 10.1016/0043-1354(82)90140-3
- Hernandez SM, Welch CN, Peters VE, Lipp EK, Curry S, Yabsley MJ, et al. (2016). Urbanized white ibises (*Eudocimus albus*) as carriers of *Salmonella enterica* of significance to public health and wildlife. *PLoS ONE* 11:e0164402. doi: 10.1371/journal.pone.0164402 [PubMed: 27768705]
- Hipsey MR, Antenucci JP, and Brookes JD (2008). A generic, process-based model of microbial pollution in aquatic systems. *Water Resour. Res.* 44. doi: 10.1029/2007WR006395
- Ishii S, Yan T, Vu H, Hansen DL, Hicks RE, and Sadowsky MJ (2010). Factors controlling long-term survival and growth of naturalized *Escherichia coli* populations in temperate field soils. *Microbes Environ.* 25, 8–14. doi: 10.1264/jsme2.ME09172 [PubMed: 21576846]
- Jay MT, Cooley M, Carychao D, Wiscomb GW, Sweitzer RA, Crawford-Miksza L, et al. (2007). *Escherichia coli* O157:H7 in feral swine near spinach fields and cattle, central California coast. *Emerg. Infect. Dis* 13, 1908–1911. doi: 10.3201/eid1312.070763 [PubMed: 18258044]
- Kangas S, Takkinen J, Hakkinen M, Nakari UM, Johansson T, Henttonen H, et al. (2008). *Yersinia pseudotuberculosis* O:1 traced to raw carrots, Finland. *Emerg. Infect. Dis* 14, 1959–1961. doi: 10.3201/eid1412.080284 [PubMed: 19046537]
- Kim JS, Lee GG, Park JS, Jung YH, Kwak HS, Kim SB, 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, 1656–1662. doi: 10.4315/0362-028X-70.7.1656 [PubMed: 17685339]
- King RS, Baker ME, Whigham DF, Weller DE, Jordan TE, Kazyak PF, et al. (2005). Spatial considerations for linking watershed land cover to ecological indicators in streams. *Ecol. Appl* 15, 137–153. doi: 10.1890/04-0481
- Kirtane A, Wilder ML, and Green HC (2019). Development and validation of rapid eDNA detection methods for bog turtle (*Glyptemys muhlenbergii*). *PLoS ONE* 14:e0222883 doi: 10.1371/journal.pone.0222883 [PubMed: 31725720]
- Kistemann T, Classen T, Koch C, Dangendorf F, Fischeder R, Gebel J, et al. (2002). Microbial load of drinking water reservoir tributaries during extreme rainfall and runoff. *Appl. Environ. Microbiol* 68, 2188–2197. doi: 10.1128/AEM.68.5.2188-2197.2002 [PubMed: 11976088]
- Ksoll WB, Ishii S, Sadowsky MJ, and Hicks RE (2007). Presence and sources of fecal coliform bacteria in epilithic periphyton communities of Lake Superior. *Appl. Environ. Microbiol* 73, 3771–3778. doi: 10.1128/AEM.02654-06 [PubMed: 17468280]
- Kuhn M (2018). caret: Classification and Regression Training 2018. Available online at: <https://cran.r-project.org/package=caret>
- Kwan PSL, Xavier C, Santovenia M, Pruckler J, Stroika S, Joyce K, et al. (2014). Multilocus sequence typing confirms wild birds as the source of a *Campylobacter* outbreak associated with the consumption of raw peas. *Appl. Environ. Microbiol* 80, 4540–4546. doi: 10.1128/AEM.00537-14 [PubMed: 24837383]
- Laidler MR, Tourdjman M, Buser GL, Hostetler T, Repp KK, Leman R, et al. (2013). *Escherichia coli* O157:H7 infections associated with consumption of locally grown strawberries contaminated by deer. *Clin. Infect. Dis* 57, 1129–1134. doi: 10.1093/cid/cit468 [PubMed: 23876397]
- Lee Y-J, Molina M, Santo Domingo JW, Willis JD, Cyterski M, Endale DM, et al. (2008). Temporal assessment of the impact of exposure to cow feces in two watersheds by multiple host-specific PCR assays. *Appl. Environ. Microbiol* 74, 6839–6847. doi: 10.1128/AEM.00601-08 [PubMed: 18806002]
- Liang L, Goh SG, Vergara GG, Fang HM, Rezaeinejad S, Chang SY, et al. (2015). Alternative fecal indicators and their empirical relationships with enteric viruses, *Salmonella enterica*, and *Pseudomonas aeruginosa* in surface waters of a tropical urban catchment. *Appl. Environ. Microbiol* 81, 850–860. doi: 10.1128/AEM.02670-14 [PubMed: 25416765]
- Liao C, Liang X, Soupier ML, and Jarboe LR (2015). Cellular, particle and environmental parameters influencing attachment in surface waters: a review. *J. Appl. Microbiol* 119, 315–330. doi: 10.1111/jam.12860 [PubMed: 26033178]



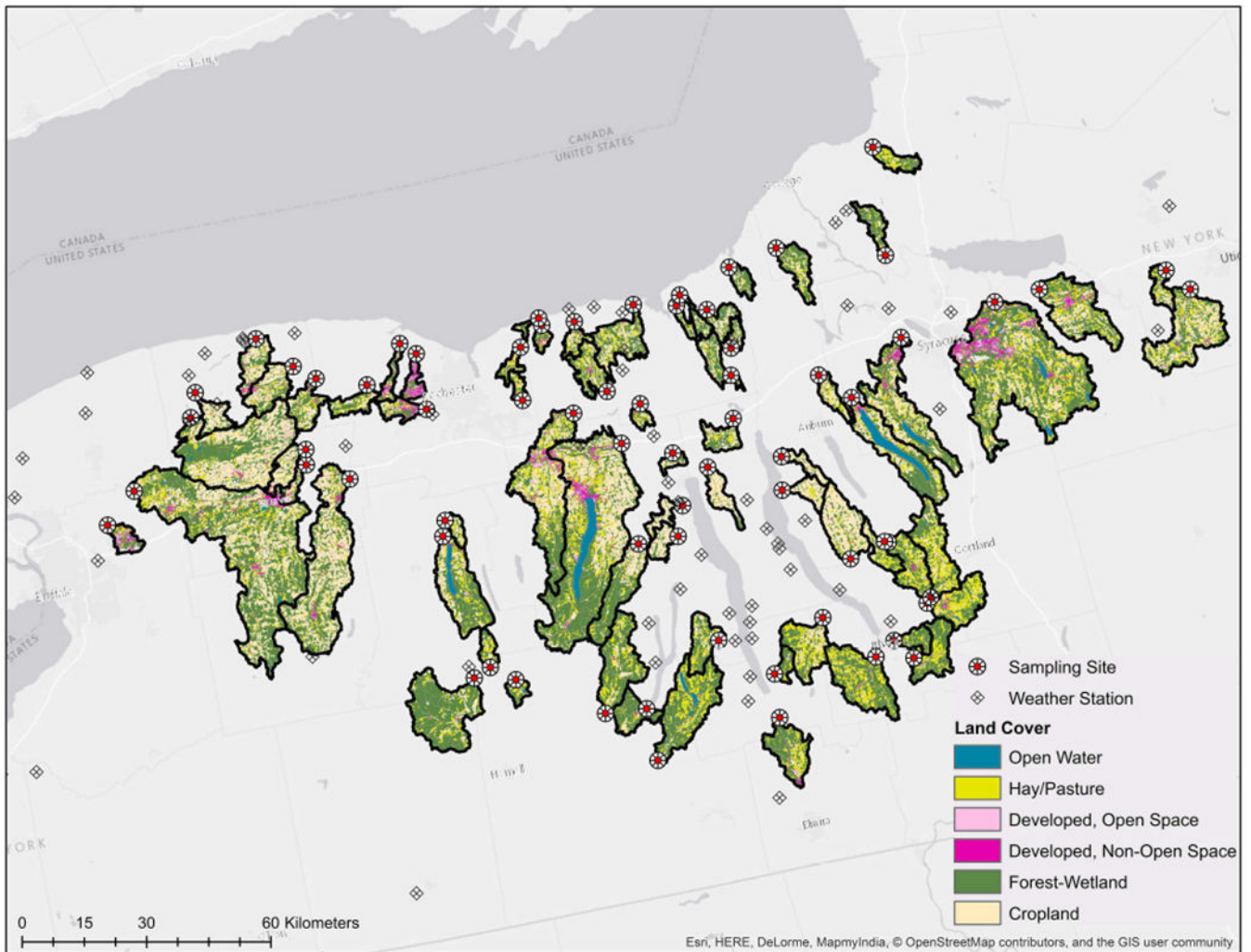
- Louis ME St., Morse, Potter ME, DeMelfi TM, Guzewich JJ, Tauxe RV, et al. (1988). The emergence of grade a eggs as a major source of *Salmonella enteritidis* infections. *JAMA* 259:2103. doi: 10.1001/jama.1988.03720140023028. [PubMed: 3279240]
- Lundqvist PJ, Gelting RJ, and Baloch M (2013). A systems analysis of irrigation water quality in environmental assessments related to foodborne outbreaks. *Aquat. Proc* 1, 130–137. doi: 10.1016/j.aapro.2013.07.011
- Luo Z, Gu G, Ginn A, Giurcanu MC, Adams P, Vellidis G, et al. (2015). Distribution and characterization of *Salmonella enterica* isolates from irrigation ponds in the Southeastern, U.S.A. *Appl. Environ. Microbiol* 81, 4376–4387. doi: 10.1128/AEM.04086-14 [PubMed: 25911476]
- Lyatey E, Lapen DR, Wilkes G, McCleary K, Pagotto F, Tyler K, et al. (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 [PubMed: 17630309]
- Marti R, Gannon VPJ, Jokinen C, Lanthier M, Lapen DR, Neumann NF, et al. (2013). Quantitative multi-year elucidation of fecal sources of waterborne pathogen contamination in the South Nation River basin using Bacteroidales microbial source tracking markers. *Water Res.* 47, 2315–2324. doi: 10.1016/j.watres.2013.02.009 [PubMed: 23497974]
- Matt Tyers (2017). riverdist: River Network Distance Computation and Applications. Available online at: <https://cran.r-project.org/package=riverdist>
- McEgan R, Mootian G, Goodridge LD, Schaffner DW, and Danyluk MD (2013a). Predicting *Salmonella* populations from biological, chemical, and physical indicators in Florida surface waters. *Appl. Environ. Microbiol* 79, 4094–4105. doi: 10.1128/AEM.00777-13 [PubMed: 23624476]
- McEgan R, Rodrigues CA, Sbodio A, Suslow TVV, Goodridge LD, and Danyluk MD (2013b). Detection of *Salmonella* spp. from large volumes of water by modified Moore swabs and tangential flow filtration. *Lett. Appl. Microbiol* 56, 88–94. doi: 10.1111/lam.12016 [PubMed: 23106632]
- Melton-Celsa AR (2014). Shiga Toxin (Stx) classification, structure, and function. *Microbiol. Spectr.* 2:EHEC-0024-2013. doi: 10.1128/microbiolspec.EHEC-0024-2013
- Mieszkin S, Yala J-F, Joubrel R, and Gourmelon M (2010). Phylogenetic analysis of Bacteroidales 16S rRNA gene sequences from human and animal effluents and assessment of ruminant faecal pollution by real-time PCR. *J. Appl. Microbiol* 108, 974–984. doi: 10.1111/j.1365-2672.2009.04499.x [PubMed: 19735325]
- Mody RK, Greene S, Gaul L, Sever A, Pichette S, Zambrana I, et al. (2011). National outbreak of *Salmonella* serotype Saintpaul infections: importance of Texas restaurant investigations in implicating Jalapeno peppers. *PLoS ONE* 6:e16579. doi: 10.1371/journal.pone.0016579 [PubMed: 21373185]
- Muirhead RW, Collins RP, and Bremer PJ (2005). Erosion and subsequent transport state of *Escherichia coli* from cowpats. *Appl. Environ. Microbiol* 71, 2875–2879. doi: 10.1128/AEM.71.6.2875-2879.2005 [PubMed: 15932980]
- Muirhead RW, Davies-Colley RJ, Donnison AM, and Nagels JW (2004). Faecal bacteria yields in artificial flood events: quantifying in-stream stores. *Water Res.* 38, 1215–1224. doi: 10.1016/j.watres.2003.12.010 [PubMed: 14975655]
- NandaKafle G, Christie AA, Vilain S, and Brözel VS (2018). Growth and extended survival of *Escherichia coli* O157:H7 in soil organic matter. *Front. Microbiol* 9:762. doi: 10.3389/fmicb.2018.00762 [PubMed: 29740410]
- Nautiyal CS, Rehman A, and Chauhan PS (2010). Environmental *Escherichia coli* occur as natural plant growth-promoting soil bacterium. *Arch. Microbiol* 192, 185–193. doi: 10.1007/s00203-010-0544-1 [PubMed: 20084366]
- Nightingale KK, Windham K, and Wiedmann M (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 [PubMed: 16077098]
- Odjadjare EEO, Obi LC, and Okoh AI (2010). Municipal wastewater effluents as a source of listerial pathogens in the aquatic milieu of the Eastern Cape Province of South Africa: a concern of public



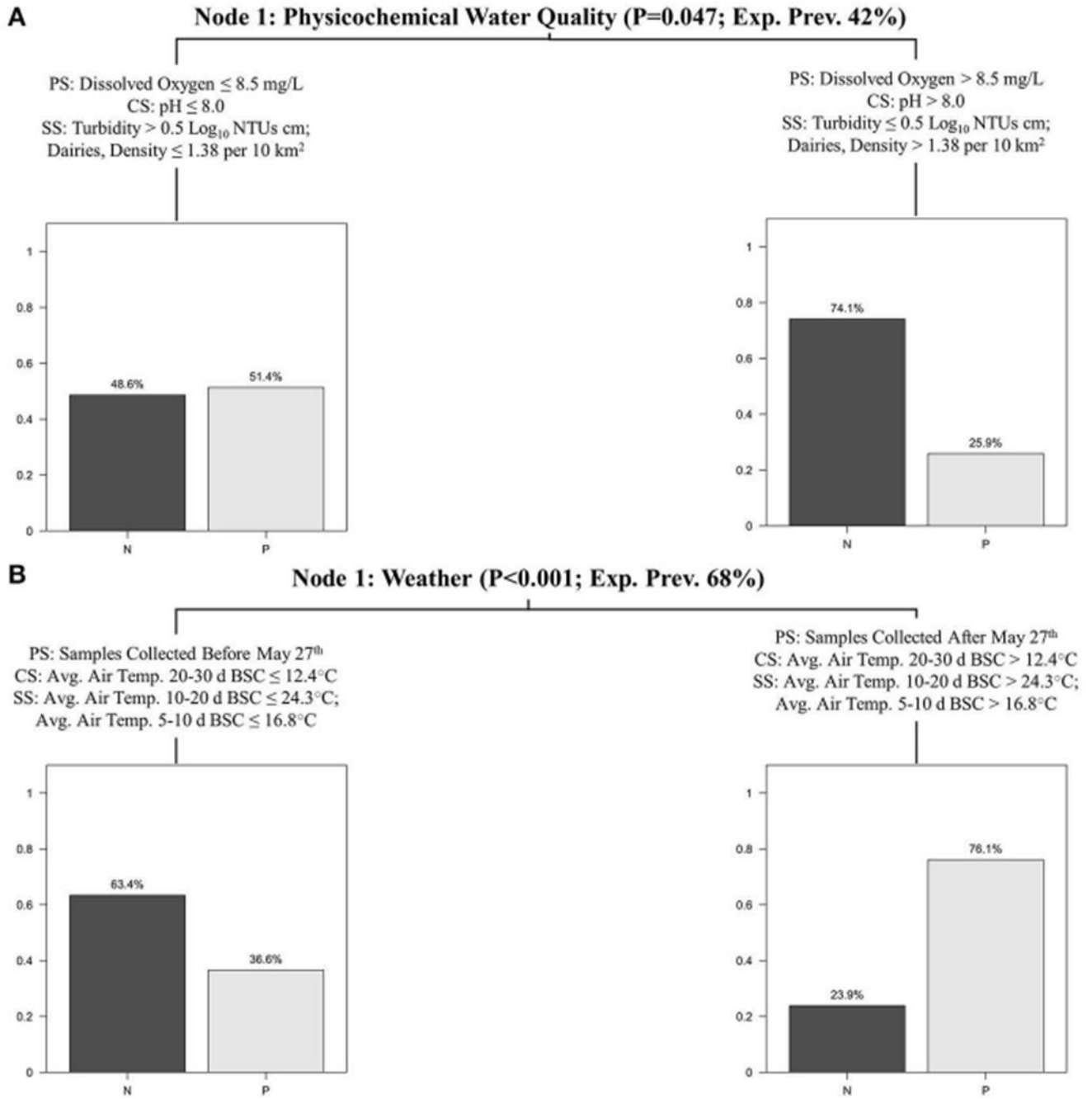
- health importance. *Int. J. Environ. Res. Public Health* 7, 2376–2394. doi: 10.3390/ijerph7052376 [PubMed: 20623030]
- Pachepsky Y, Shelton D, Dorner S, and Whelan G (2015). Can *E. coli* or thermotolerant coliform concentrations predict pathogen presence or prevalence in irrigation waters? *Crit. Rev. Microbiol* 42, 384–393. doi: 10.3109/1040841X.2014.954524
- Paillard D, Dubois V, Thiebaut R, Nathier F, Hoogland E, Caumette P, et al. (2005). Occurrence of *Listeria* spp. in effluents of French urban wastewater treatment plants. *Appl. Environ. Microbiol* 71, 7562–7566. doi: 10.1128/AEM.71.11.7562-7566.2005 [PubMed: 16269804]
- Pandey PK, Soupir ML, Haddad M, and Rothwell JJ (2012). Assessing the impacts of watershed indexes and precipitation on spatial in-stream *E. coli* concentrations. *Ecol. Indic* 23, 641–652. doi: 10.1016/j.ecolind.2012.05.023
- Partyka ML, Bond RF, Chase JA, and Atwill ER (2018). Spatiotemporal variability in microbial quality of western US agricultural water supplies: a multistate study. *J. Environ. Qual* 47, 939–948. doi: 10.2134/jeq2017.12.0501 [PubMed: 30272786]
- Payment P, and Locas A (2011). Pathogens in water: value and limits of correlation with microbial indicators. *Ground Water* 49, 4–11. doi: 10.1111/j.1745-6584.2010.00710.x [PubMed: 20477877]
- Petit F, Clermont O, Delannoy S, Servais P, Gourmelon M, Fach P, et al. (2017). Change in the structure of *Escherichia coli* population and the pattern of virulence genes along a rural aquatic continuum. *Front. Microbiol* 8:609. doi: 10.3389/fmicb.2017.00609 [PubMed: 28458656]
- Rice DH, Hancock DD, and Besser TE (1995). Verotoxigenic *E. coli* O157 colonisation of wild deer and range cattle. *Vet. Rec* 137:524. doi: 10.1136/vr.137.20.524
- Roberts AJ, and Wiedmann M (2006). Allelic exchange and site-directed mutagenesis probe the contribution of ActA amino-acid variability to phosphorylation and virulence-associated phenotypes among *Listeria monocytogenes* strains. *FEMS Microbiol. Lett* 254, 300–307. doi: 10.1111/j.1574-6968.2005.00041.x [PubMed: 16445760]
- Rodrigue DC, Tauxe RV, and Rowe B (1990). International increase in *Salmonella enteritidis*: a new pandemic? *Epidemiol. Infect* 105, 21–27. doi: 10.1017/S0950268800047609 [PubMed: 2200698]
- Sassoubre LM, Walters SP, Russell TL, and Boehm AB (2011). Sources and fate of Salmonella and fecal indicator bacteria in an urban creek. *J. Environ. Monit* 13:2206. doi: 10.1039/c1em10213c [PubMed: 21687857]
- Sbodio A, Maeda S, Lopez-Velasco G, and Suslow TV (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
- Schaffter N, Zumstein J, and Parriaux A (2004). Factors influencing the bacteriological water quality in mountainous surface and groundwaters. *Acta Hydrochim. Hydrobiol* 32, 225–234. doi: 10.1002/aheh.200300532
- Silva S, Teixeira P, Oliveira R, and Azeredo J (2008). Adhesion to and viability of *Listeria monocytogenes* on food contact surfaces. *J. Food Prot* 71, 1379–1385. doi: 10.4315/0362-028X-71.7.1379 [PubMed: 18680936]
- Stea EC, Truelstrup Hansen L, Jamieson RC, and Yost CK (2015). Fecal contamination in the surface waters of a rural- and an urban-source watershed. *J. Environ. Qual* 44:1556. doi: 10.2134/jeq2014.11.0459 [PubMed: 26436273]
- Steele JA, Blackwood AD, Griffith JF, Noble RT, and Schiff KC (2018). Quantification of pathogens and markers of fecal contamination during storm events along popular surfing beaches in San Diego, California. *Water Res* 136, 137–149. doi: 10.1016/j.watres.2018.01.056 [PubMed: 29501758]
- Strawn LK, Fortes ED, Bihn E, Nightingale KK, Grohn YT, Worobo RW, 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 [PubMed: 23144137]
- Strawn LK, Grohn YT, Warchocki S, Worobo RW, Bihn E, and 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 [PubMed: 24077713]

- Thelin R, and Gifford GF (1983). Fecal coliform release patterns from fecal material of cattle. *J. Environ. Qual* 12:57. doi: 10.2134/jeq1983.00472425001200010008x
- Truchado P, Hernandez N, Gil MI, Ivanek R, and Allende A (2018). Correlation between *E. coli* levels and the presence of foodborne pathogens in surface irrigation water: establishment of a sampling program. *Water Res.* 128, 226–233. doi: 10.1016/j.watres.2017.10.041 [PubMed: 29107907]
- Truitt LN, Vazquez KM, Pfunter RC, Rideout SL, Havelaar AH, and Strawn LK (2018). Microbial quality of agricultural water used in produce preharvest production on the eastern shore of Virginia. *J. Food Prot* 81, 1661–1672. doi: 10.4315/0362-028X.JFP-18-185 [PubMed: 30212229]
- Ukuku DO, and Fett WF (2002). Relationship of cell surface charge and hydrophobicity to strength of attachment of bacteria to Cantaloupe rind. *J. Food Prot* 65, 1093–1099. doi: 10.4315/0362-028X-65.7.1093 [PubMed: 12117240]
- Varness KJ, Pacha RE, and Lapen RF (1978). Effects of dispersed recreational activities on the microbiological quality of forest surface water. *Appl. Environ. Microbiol* 36, 95–104. [PubMed: 100055]
- Vereen E, Lowrance RR, Jenkins MB, Adams P, Rajeev S, and Lipp EK (2013). Landscape and seasonal factors influence *Salmonella* and *Campylobacter* prevalence in a rural mixed use watershed. *Water Res.* 47, 6075–6085. doi: 10.1016/j.watres.2013.07.028 [PubMed: 23969398]
- Vivant A-L, Garmyn D, and Piveteau P (2013). *Listeria monocytogenes*, a down-to-earth pathogen. *Front. Cell Infect. Microbiol* 3:87. doi: 10.3389/fcimb.2013.00087 [PubMed: 24350062]
- Vorst KL, Todd EC, and Rysert ET (2004). Improved quantitative recovery of *Listeria monocytogenes* from stainless steel surfaces using a one-ply composite tissue. *J. Food Prot* 67, 2212–2217. doi: 10.4315/0362-028X-67.10.2212 [PubMed: 15508632]
- Walker SL, Redman JA, and Elimelech M (2005). Influence of growth phase on bacterial deposition: interaction mechanisms in packed-bed column and radial stagnation point flow systems. *Environ. Sci. Technol* 39, 6405–6411. doi: 10.1021/es050077t [PubMed: 16190193]
- Walters SP, Gannon VP, and Field KG (2007). Detection of Bacteroidales fecal indicators and the zoonotic pathogens *E. coli* O157:H7, *Salmonella*, and *Campylobacter* in river water. *Environ. Sci. Technol* 41, 1856–1862. doi: 10.1021/es0620989 [PubMed: 17410775]
- Walters SP, Thebo AL, and Boehm AB (2011). Impact of urbanization and agriculture on the occurrence of bacterial pathogens and *stx* genes in coastal waterbodies of central California. *Water Res.* 45, 1752–1762. doi: 10.1016/j.watres.2010.11.032 [PubMed: 21168181]
- Wan Norhana MN, Goulter RM, Poole SE, Deeth HC, and Dykes GA (2009). Relationship between the physicochemical properties of nonchitinolytic *Listeria* and *Salmonella* and their attachment to shrimp carapace. *J. Food Prot* 72, 1181–1189. doi: 10.4315/0362-028X-72.6.1181 [PubMed: 19610328]
- Watkins J, and Sleath KP (1981). Isolation and enumeration of *Listeria monocytogenes* from sewage, sewage sludge and river water. *J. Appl. Bacteriol* 50, 1–9. doi: 10.1111/j.1365-2672.1981.tb00865.x [PubMed: 6262293]
- Wei T (2013). corrplot: Visualization of a Correlation Matrix. R Package Version 0.73. Available online at: <http://cran.r-project.org/package=corrplot>
- Weller D, Shiwakoti S, Bergholz P, Grohn Y, Wiedmann M, and Strawn LK (2016). Validation of a previously developed geospatial model that predicts *Listeria monocytogenes* prevalence for New York State produce fields. *Appl. Environ. Microbiol* 82, 797–807. doi: 10.1128/AEM.03088-15 [PubMed: 26590280]
- Weller D, Wiedmann M, and Strawn L (2015a). 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 [PubMed: 26038903]
- Weller D, Wiedmann M, and Strawn L (2015b). Spatial and temporal factors associated with an increased prevalence of *L. monocytogenes* in spinach fields in New York State. *Appl. Environ. Microbiol* 81, 6059–6069. doi: 10.1128/AEM.01286-15 [PubMed: 26116668]
- Weller DL, Brassill N, Rock CM, Ivanek R, Mudrak E, Ganda E, et al. (2020). Complex interactions between weather, and microbial and physicochemical water quality impact the likelihood of detecting foodborne pathogens in agricultural water used for produce production. *bioRxiv*. doi: 10.1101/2020.01.02.892851

- Whitman RL, Shively DA, Pawlik H, Nevers MB, and Byappanahalli MN (2003). Occurrence of *Escherichia coli* and enterococci in Cladophora (Chlorophyta) in nearshore water and beach sand of Lake Michigan. *Appl. Environ. Microbiol* 69, 4714–4719. doi: 10.1128/AEM.69.8.4714-4719.2003 [PubMed: 12902262]
- Wilkes G, Brassard J, Edge TA, Gannon V, Jokinen CC, Jones TH, et al. (2013). Coherence among different microbial source tracking markers in a small agricultural stream with or without livestock exclusion practices. *Appl. Environ. Microbiol* 79, 6207–6219. doi: 10.1128/AEM.01626-13 [PubMed: 23913430]
- 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 [PubMed: 19339033]
- Wilkes G, Edge TA, Gannon VPJ, Jokinen C, Lyautey E, Neumann NF, 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 [PubMed: 21889781]
- Zhu L, Torres M, Betancourt WQ, Sharma M, Micallef SA, Gerba C, et al. (2019). Incidence of fecal indicator and pathogenic bacteria in reclaimed and return flow waters in Arizona, U.S.A. *Environ. Res* 170, 122–127. doi: 10.1016/j.envres.2018.11.048 [PubMed: 30579985]
- Baur P, Driscoll L, Gennet S, and Karp D (2016). Inconsistent food safety pressures complicate environmental conservation for California produce growers. *Calif. Agric* 70, 142–151. doi: 10.3733/ca.2016a0006
- Karp DS, Baur P, Atwill ER, De Master K, Gennet S, Iles A, et al. (2015). The unintended ecological and social impacts of food safety regulations in California’s central coast region. *Bioscience.* 65, 1173–1183. doi: 10.1093/biosci/biv152
- Beretti M, and Stuart D (2008). Food safety and environmental quality impose conflicting demands on Central Coast growers. *Calif. Agric* 62, 68–73. doi: 10.3733/ca.v062n02p68
- Sweeney BW, Bott TL, Jackson JK, Kaplan LA, Newbold JD, Standley LJ, et al. (2004). Riparian deforestation, stream narrowing, and loss of stream ecosystem services. *Proc. Natl. Acad. Sci* 101, 14132–14137. doi: 10.1073/pnas.0405895101 [PubMed: 15381768]
- Schultz RC, Isenhardt TM, Simpkins W, and Colletti JP (2004). Riparian forest buffers in agroecosystems - lessons learned from the Bear Creek Watershed, central Iowa, USA. *Agrofor. Syst* 61, 35–50. doi: 10.1023/B:AGFO.0000028988.67721.4d

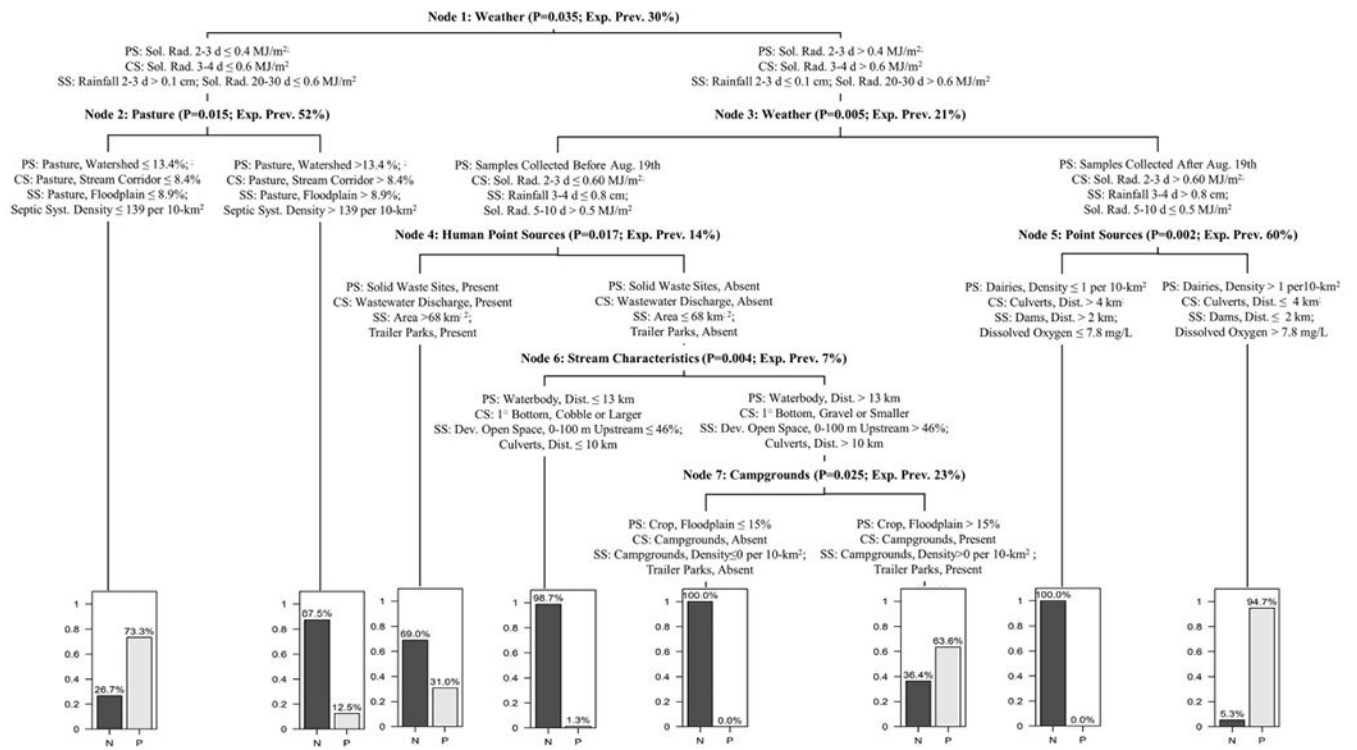


**FIGURE 1 |** Map showing the study region, including land cover in the sampled watersheds as well as the location of the sampling sites and the NEWA weather stations.



**FIGURE 2 |** *Salmonella* (A), and *eaeA-stx* (B) CTrees showing the primary split (PS), competitor split (CS), and two surrogate splits (SS) for each node. The splits that comprise each node are generalized in the node description (e.g., most splits in node 1A are physicochemical water quality factors). The *p-value* is the *p-value* associated with the primary split for the given node. The expected prevalence in Node 1 was the prevalence of the given microbial target in the study reported here. Bar plots show the exp. prevalence of samples that were negative (N) or positive (P) for the given target(s) in each terminal node.





**FIGURE 3 |**

*L. monocytogenes* CTree showing the primary split (PS), competitor split (CS), and two surrogate splits (SS) for each node. The splits that comprise each node are generalized in the node description (e.g., all splits in node 1 are weather factors). Each node description also includes the *P*-value associated with the node’s primary splits, and the expected *L. monocytogenes* prevalence (exp. prev.) based on the node’s parent splits [e.g., exp. prev. in Node 2 for samples collected when avg. solar radiation (sol. rad) 2–3 d BSC was  $< 0.4$  MJ/m<sup>2</sup> was 52%]. The expected prevalence in Node 1 is greater than that reported in Table 1 since upsampling was used to address the imbalanced nature of the *L. monocytogenes* data when performing CTree analysis. Bar plots show the exp. prevalence of *L. monocytogenes* negative (N) and positive (P) samples in each terminal node. For example, the expected prevalence of *L. monocytogenes* in samples collected on a day when sol. rad 3-4 d BSC was  $0.4$  MJ/m<sup>2</sup> and pasture accounted for  $> 13.4\%$  of the upstream area was 12.5%.

TABLE 1 |

FST qPCR assays used and their performance parameters.

Assay name	Target	References	F and R primer conc. (μM)	Probe conc. (μM)	R <sup>2</sup>	Eff. (%) <sup>a</sup>	LLOQ <sup>b</sup>
DG3	Canine	(Green et al., 2014b)	1.4	0.1	0.994	97.2	10
HF183	Human	(Bernhard and Field, 2000; Green et al., 2014a)	1	0.08	0.994	98.8	10
Modified GFD	Avian	(Green et al., 2012)	1	0.1	0.993	97.4	10
Rum2Bac	Ruminant	(Mieszkin et al., 2010)	0.3	0.1	0.991	99.9	10
CG4	<i>C. elegans</i> internal control	(Kirtane et al., 2019)	1.4	0.1	0.943	99.9	10

<sup>a</sup> Calculated using %Eff =  $(10^{(-1/\text{slope})} - 1)$ , where slope is the assay-specific slope of 4–6 pooled standard curves constructed using synthetic gBlocks® (IDT).

<sup>b</sup> Assay lower limits of quantification (gBlock® copies/reaction).

TABLE 2 |

Number of times each target was detected in a given watershed.

No. of Times Target Detected	Culture-Confirmed				PCR-Screen Positive <sup>d</sup>	
	<i>Listeria</i> <sup>a</sup>		<i>Salmonella</i> <sup>c</sup>		<i>eaeA</i>	<i>stx</i>
	<i>Listeria</i> spp. excluding <i>monocytogenes</i> <sup>b</sup>	<i>monocytogenes</i>				
	<i>innocua</i>	<i>marthii</i>	<i>seeligeri</i>	<i>welshimeri</i>		
<b>IN STREAMS SAMPLED TWICE (N = 8)</b>						
Zero	8	8	6	6	5	1
Once	0	0	1	2	3	1
Twice	0	0	1	0	0	3
						8
<b>IN STREAMS SAMPLED THRICE (N = 60)</b>						
Zero	49	55	43	47	44	13
Once	10	5	16	11	15	30
Twice	1	0	1	1	1	12
Thrice	0	0	0	1	0	5
						54
						19

<sup>a</sup>Overall *L. innocua*, *L. marthii*, *L. monocytogenes*, *L. seeligeri*, and *L. welshimeri* were isolated from 6% (12/196), 3% (5/196), 10% (20/196), 11% (21/196), and 9% (18/196) of samples, respectively. One sample was positive for both *L. monocytogenes* and *L. marthii*, and one sample was positive for both *L. innocua* and *L. seeligeri*. *L. innocua*, *L. marthii*, *L. monocytogenes*, *L. seeligeri*, and *L. welshimeri* were isolated from 16% (11/68), 7% (5/68), 28% (19/68), 28% (19/68), and 22% (15/68) of streams, respectively.

<sup>b</sup>*Listeria* spp. excluding *monocytogenes* was isolated from 28% (55/196) of samples and 71% (48/68) of streams.

<sup>c</sup>*Salmonella* was isolated from 42% of samples (79/196) and 75% (51/68) of streams. Four samples were PCR-screen positive for *Salmonella* but culture-negative for *Salmonella*; these samples were not considered positive for *Salmonella* in the analyses reported here.

<sup>d</sup>The *eaeA* and *stx* genes were detected in 96% (190/196) and 68% (133/196) of samples, respectively. Additionally, the *eaeA* and *stx* genes were detected in 100% (68/68) and 96% (65/68) of streams, respectively. All six-positive samples were also *eaeA*-positive.

Ability of modified Moore swabs (mMS) compared to 0.45 µm filters to recover *Listeria* from 29 paired 9-L and 1-L grab samples, respectively.

TABLE 3 |

Species <sup>a</sup>	Prevalence (No. Pos.)		Proportion of		P-value for McNemar's $\chi^2$ <sup>b</sup>	Kappa <sup>c</sup>
	mMS	Filters	Pos. Agreement	Neg. Agreement		
<i>L. seeligeri</i>	7% (2)	17% (5)	29%	90%	0.375	0.278
<i>L. monocytogenes</i>	3% (1)	24% (7)	0%	84%	0.070	0.581
<i>Listeria</i> spp. excluding <i>monocytogenes</i>	7% (2)	38% (11)	15%	76%	0.012	-
<i>Listeria</i> spp. including <i>monocytogenes</i>	10% (3)	62% (18)	10%	49%	<0.001	-

<sup>a</sup>Due to the small number of samples that were *L. innocua*, *L. marthii*, and *L. welshimeri* positive (i.e., two samples each) analyses could not be performed to compare recovery using mMS and 0.45 µm filters.

<sup>b</sup>McNemar's  $\chi^2$  square tests for symmetry in a two-by-two contingency table; a  $P < 0.05$  indicates significant disagreement in the number of mMS and the number of filters that were positive for the target organism.

<sup>c</sup>Cohen's Kappa score is a measure of the level of agreement between two tests beyond what would have been expected by random chance. A  $p < 0.05$  indicates that the extent of agreement is greater than expected by random chance. Due to the identification of significant disagreement in *Listeria* spp. recovery by mMS compared to 0.45 µm filter by McNemar's  $\chi^2$  test, the kappa test was not performed to determine if there was significant agreement in *Listeria* spp. recovery by mMS compared to 0.45 µm filter.

TABLE 4 |

Associations between detection of each microbial or molecular target, and weather and water quality factors according to generalized linear mixed models<sup>a</sup>.

Outcome	Factor	Change in odds	95% CI <sup>b</sup>	P-value
<i>eaeA-sfA</i> Codetection	Avg. Solar Radiation 0-1d BSC <sup>c</sup> (MJ/m <sup>2</sup> )	0.37	0.12, 4.15	0.084
	<i>E. coli</i> Levels (log <sub>10</sub> MPN/100-mL)	1.77	0.98, 3.20	0.057
	Flow Rate (m/s)	2.90	0.86, 9.77	0.087
	Rum2Bac, Concentration (log <sub>10</sub> Copies/100-mL)	1.64		0.012
	Rum2Bac, Detection (Failure to Detect = Reference)	4.25	1.35, 13.37	0.013
	<i>Listeria</i> species excluding <i>L. monocytogenes</i>	Avg. Solar Radiation 2-3 d BSC (MJ/m <sup>2</sup> )	0.16	0.02, 1.29
Avg. Air Temperature 20-30 d BSC (C)		0.93	0.86, 1.00	0.061
Dissolved Oxygen Levels (mg/L)		1.34	1.06, 1.69	0.015
<i>E. coli</i> Levels (log <sub>10</sub> MPN/100-mL)		0.50	0.26, 0.96	0.036
PH		3.65	1.14, 11.64	0.029
Total Rainfall 2-3 d BSC (cm)		3.28	1.44, 7.51	0.005
Total Rainfall 10-20 d BSC (cm)		0.78	0.59, 1.03	0.081
Turbidity (log NTUs)		0.36	0.14, 0.94	0.037
Water Temperature (C)		0.92	0.84, 0.99	0.029
Turbidity (log <sub>10</sub> NTUs)		0.32	0.09, 1.08	0.066
<i>L. monocytogenes</i> Isolation		Avg. Solar Radiation 0-1 d BSC (MJ/m <sup>2</sup> )	0.27	0.09, 0.77
	Avg. Solar Radiation 2-3 d BSC (MJ/m <sup>2</sup> )	0.36	0.12, 1.08	0.068
	Avg. Air Temperature 10-20 d BSC (C)	0.94	0.87, 1.01	0.095
	Dissolved Oxygen Levels (mg/L)	0.74	0.72, 0.89	0.001
	<i>E. coli</i> Levels (log <sub>10</sub> MPN/100-mL)	1.83	1.08, 3.09	0.025
	HF183, Concentration (log <sub>10</sub> copies/100-mL)	1.24	1.01, 1.52	0.039
<i>Salmonella</i> Isolation	HF183, Detection (Failure to Detect = Reference)	1.78	0.91, 3.50	0.095



Outcome	Factor	Change in odds	95% CI <sup>b</sup>	P-value
	pH	0.24	0.10, 0.58	0.002
	Rainfall 0–1 d BSC (cm)	2.18	1.31, 3.62	0.003

<sup>a</sup>Since this was a hypothesis-generating study, two thresholds were used for interpreting the results of the GLMMs. Specifically,  $P < 0.05$  indicated that likelihood of microbial target detection and the factor were significantly associated, while a 0.10  $P < 0.10$  indicated the presence of a potential relationship that warrants investigation in future studies.

<sup>b</sup>CI, Confidence interval.

<sup>c</sup>BSC, before sample collection.

TABLE 5 |

Associations between detection of each microbial or molecular target, and spatial factors according to generalized linear mixed models.<sup>a</sup>

Outcome	Factor	Change in Odds	95% CI <sup>b</sup>	P-value	
<i>eteA-stx</i> Codetection	Bottom Substrate, Cobble or Larger (Absent = Reference)	0.50	0.24, 1.05	0.068	
	Developed Non-Open Space, IDW <sup>c</sup> % of Flood Plain	0.92	0.85, 0.99	0.012	
	Developed Non-Open Space, IDW % of Stream Corridor	0.93	0.86, 1.00	0.036	
	Developed Non-Open Space, IDW % of Total Watershed	0.95	0.90, 1.00	0.034	
	Developed Open Space, IDW % of Flood Plain	0.95	0.91, 0.99	0.012	
	Developed Open Space, IDW % of Stream Corridor	0.95	0.91, 0.99	0.010	
	Developed Open Space, IDW % of Total Watershed	0.93	0.88, 0.99	0.018	
	Ditch, Present Immediately Upstream of Site (Absent = Reference)	2.17	1.02, 4.62	0.044	
	Forest Wetland, IDW % of Flood Plain	1.02	1.00, 1.04	0.022	
	Forest Wetland, IDW % of Stream Corridor	1.02	1.00, 1.05	0.038	
	Open Water, IDW % of Flood Plain	0.98	0.95, 1.00	0.043	
	Open Water, IDW % of Stream Corridor	0.90	0.81, 0.99	0.033	
	Open Water, IDW % of Total Watershed	0.83	0.69, 0.99	0.033	
	Wastewater Discharge Sites, Upstream Density (per 10 km <sup>2</sup> )	0.23	0.04, 1.30	0.096	
	In-stream Waterbodies, Flow Path Distance to Nearest (km)	1.10	1.01, 1.19	0.032	
	In-stream Waterbodies, Present Upstream (Absent = Reference)	2.04	0.96, 4.35	0.065	
	<i>Listeria</i> species excluding <i>L. monocytogenes</i>	Bottom Substrate, Coarse Gravel (Absent = Reference)	2.29	0.97, 5.41	0.059
		Bottom Substrate, Fine Gravel (Absent = Reference)	2.12	0.87, 5.16	0.097
		Bottom Substrate, Organic Matter (Absent = Reference)	0.42	0.16, 1.10	0.077
Bottom Substrate, Sand (Absent = Reference)		0.45	0.19, 1.04	0.062	
Campgrounds, Present Upstream (Absent = Reference)		0.37	0.13, 1.11	0.076	
Dairy Operations, Flow Path Distance to Nearest (km)		0.84	0.03, 0.73	0.027	
Dairy Operations, Present Upstream (Absent = Reference)		0.26	0.75, 0.98	0.028	
Ditch, Present Immediately Upstream of Site (Absent = Reference)		0.41	0.16, 1.05	0.064	
In-stream Waterbodies, Flow Path Distance to Nearest (km)		0.86	0.75, 0.98	0.028	
Livestock Operation, Present Upstream (Absent = Reference)		0.27	0.08, 0.93	0.039	

Outcome	Factor	Change in Odds	95% CI <sup>b</sup>	P-value	
<i>L. monocytogenes</i> Isolation	Pasture, IDW % of Flood Plain	1.03	1.00, 1.07	0.070	
	Pig Farms, Present Upstream (Absent = Reference)	0.16	0.04, 0.69	0.014	
	Pig Farms, Upstream Density (per 10 km <sup>2</sup> )	0.00 <sup>d</sup>	0.00, 5.17	0.077	
	Stables, Present Upstream (Absent = Reference)	0.32	1.06, 1.67	0.018	
	Campgrounds, Upstream Density (per 10 km <sup>2</sup> )	565.4 <sup>e</sup>	0.91, 3.50*10 <sup>5</sup>	0.053	
	Developed Non-Open Space, 0–100m Upstream of Site (%)	0.96	0.91, 1.01	0.086	
	<i>Salmonella</i> Isolation	Bottom Substrate, Cobble or Larger (Absent = Reference)	0.55	0.30, 1.01	0.054
		Campground, Flow Path Distance to Nearest (km)	0.94	0.88, 1.01	0.068
		Dairy Operations, Upstream Density (per 10 km <sup>2</sup> )	0.47	0.28, 0.81	0.007
		Pasture, 0–100m Upstream of Site (%)	0.99	0.97, 1.00	0.053
Developed Open Space, IDW % of Stream Corridor		1.03	1.00, 1.06	0.096	
Ditch, Stormwater Outfalls Present Upstream (Absent = Reference)		2.04	1.03, 4.05	0.042	
Poultry Operations, Flow Path Distance to Nearest (km)	0.93	0.86, 1.00	0.060		

<sup>a</sup>Since this was a hypothesis-generating study, two thresholds were used for interpreting the results of the GLMMs. Specifically,  $P < 0.05$  indicated that likelihood of microbial target detection and the factor were significantly associated, while a 0.05  $P < 0.10$  indicated the presence of a potential relationship that warrants investigation in future studies.

<sup>b</sup>CI, Confidence interval.

<sup>c</sup>IDW, Inverse distance weighted.

<sup>d</sup>The odds ratio was  $>0.000$  but  $<0.001$ .

<sup>e</sup>This confidence interval is wide due to sparse data bias, which is a product of the low prevalence of *L. monocytogenes* (10% ; 20/196) and of watersheds with campgrounds upstream of the sampling site (22% ; 16/71); 54% of *L. monocytogenes*-positive samples were collected from watersheds with campgrounds upstream.