Bacterial Sources of Pollution in Urban Stormwater Impacted Bodies of Water

Megan Beaudry

2019-11-23

# Introduction

Freshwater is a finite resource. Countries all over the world are struggling to meet the growing demands on freshwater resources. Several countries, (e.g., Australia, Israel, and the United States) are investigating ways to address freshwater shortages, including the reuse of wastewater and stormwater as alternative water resources. Harvesting stormwater provides cities with a strategy to address the growing demands on water resources due to climate change and projected population growth. However, stormwater reuse poses a variety of challenges due to the potential of this source to be of low water quality, and to be contaminated with human and animal feces, and thus enteric bacterial pathogens including *Campylobacter* spp., *Salmonella* spp., and pathogenic *E. coli*. In addition, storm events are correlated with an increased prevalence of disease, likely due to the mobilization of pathogens in the environment, leading to increased exposure and transmission risks. The contamination of water with human and/or animal excreta possesses significant risks to human health - albeit the risks associated with pathogens found in sewage are greater than those associated with animal wastes. There are several factors to consider in assessing microbial contamination of stormwater, including the process by which stormwater becomes contaminated, the duration to which stormwater is exposed to contaminants, and the overall levels of contamination (Fong, 2009). Therefore, many governmental agencies or departments, as the United States Environmental Protection Agency (USEPA) or Health Canada recommend the use of microbial fecal indicator bacteria (FIB) (e.g., *Enterococcus* spp., thermotolerant coliforms, and *E. coli*) to help determine water quality. FIB are enteric bacteria found in the digestive tract of humans and animals and excreted in the feces, and therefore their presence in stormwater is indicative of fecal contamination. By association, the presence of FIB implies the potential presence of fecal- orally transmitted pathogens (e.g., *Campylobacter* spp., *Salmonella* spp., *Arcobacter butzleri*).

The field of microbial source tracking aims to identify the specific host sources of fecal pollution impacting water quality (e.g., human, cattle, bird, etc.). Microbial source tracking tools are focused on genotypic and phenotypic differences in microbial populations found in the gut of different animals (Curtis, 2016).Many molecular microbial source tracking techniques have emerged over the last two decades, many of which are based on Polymerase Chain Reaction (PCR) methods targeting DNA sequences unique to certain bacteria found in various animals. The Bacteroides-Prevotella group is a common target for microbial source tracking assays. Other targets include *Escherichia*, the genera Catellicoccus, and members of the order Bacteroidales (Wuertz, 2011). The PCR-based methods provide a rapid means for identifying potential sources of fecal pollution in the environment. However, there is some uncertainty associated with the use of these types of tools, which are similar to that of FIB, as little is known about the persistence of these microbial signatures in the environment (Wuertz, 2011) (Harwood, 2014). Bacteroidales are gram- negative, obligate anaerobes, non-spore forming, and rod-shaped bacteria, and within this order is the genus Bacteroides (Harwood, 2014). Bacteroides are bacteria that are commonly found in the intestine of warm-blooded animals and constitute a large portion of their gut microbiota (e.g., human, dog, etc.) (Bower, 2005) (Shanks, 2009). Several *Bacteroides* spp. have strong host or group specificities as indicated by several studies (Layton A., 2006) (Shanks, 2009) . Therefore, the 16S rRNA genes of host-specific *Bacteroides* spp. can be used as markers for ascertaining human or animal fecal sources of pollution in water (Layton A., 2006).

Although there is an abundance of literature on enteric bacterial pathogens in water systems, there is a lack of data on their presence in urban stormwater. Poor water quality within urban stormwater-impacted bodies of water represents a potentially important public health problem, as many of these water bodies are subjected to chronic issues with human and animal sources of fecal contamination, inferring the potential for zoonotic and anthropogenic enteric bacterial pathogens to be present. Pathogens as *Arcobacter butzleri* (Douidah, 2011) (Van Driessche & Houf, 2005) (Levican, 2013) (Hafliger, 2013) (Craun, 2005), and other enteric bacterial pathogens (e.g., *Campylobacter spp.*, *Salmonella* spp., Shiga-toxin producing *E. coli*) (Moore, 2001) (Krometis, 2010) (Rangel, 2005) (Adams, 2016), have all been implicated in waterborne outbreaks throughout the world. Rain events mobilize and transport fecal pathogen in the environment thereby increasing the effects of non-point and point sources of contamination, which in turn can augment the risk to public health (Staley, 2018).

Stormwater-impacted bodies of water can serve as reservoirs for transmission for enteric bacterial pathogens by the fecal-oral route through ingestion of contaminated water during recreational activities (e.g., swallowing water while swimming); by way of contaminated irrigation water on food that is then eaten; or by accidental ingestion that occurs during irrigation. An outbreak of HUS associated with a recreational water body in Connecticut, USA, occurred where STEC was detected in a storm drain that emptied onto the beach (McCarthy, 2001). In a study of stormwater discharges and gastrointestinal illness following wet weather in California, USA, (Soller, 2017) that wet weather exposure during surfing lead to higher than average illness rates due to human enteric viruses. In addition, they observed Campylobacter spp. above the method detection limit in over half of their samples from stormwater discharges. Meng et al., 2018 identified *Campylobacter spp.* in stormwater constructed wetlands intended for reuse activities, and found the concentrations to be similar between wet and dry weather events. Furthermore, they found that log reduction targets for reuse activities were not being met (Meng, 2018).

The *Arcobacter* genus was created in 1991 in an effort to accommodate aero-tolerant *Campylobacter* spp. (Vandenberg, 2004). *Arcobacter* spp. are gram-negative, curved bacteria (Van Driessche & Houf, 2005). Currently, there are 22 species (Van Driessche & Houf, 2005). Potential sources of *Arcobacter* spp. include humans, birds, and livestock. This genus has been characterized as a potential food or waterborne pathogen; and has been implicated in causing human disease, with such symptoms as bacteremia, diarrhea, and gastroenteritis from three species: *A. butzleri*, *A. cryaerophilus*, and *A. skirrowii* (Kayman, 2012). *A. butzleri* contains the ability to survive and grow in the environment. Further, it can survive at lower temperatures (15-30˚C) than *Campylobacter* spp.; and can grow in the presence of oxygen (Van Driessche & Houf, 2005) (Wesley, 2000). *Arcobacter* spp. are considered a zoonotic pathogen, and one study identified it as the fourth most frequent bacteria isolated from humans with acute enteric disease (Levican, 2013).

The objective this study was to identify and determine the prevalence of the putative enteric bacterial pathogen *A. butzleri* in urban stormwater ponds located in western Canada, as there is currently limited information as to the critical role of enteric bacterial pathogens, in particular *A. butzleri*, in outbreaks associated with stormwater reuse. Furthermore, we sought to resolve the source of the enteric bacterial pathogens by using microbial source tracking techniques, addressing overall water quality by assessing FIB contamination, and analyzing environmental parameters.

# Materials and Methods

## Sampling

To determine what select enteric bacterial pathogens present, stormwater samples were collected from stormwater ponds southern Alberta, Canada. Semiweekly sampling began as soon as stormwater ponds were fully thawed (i.e., May 9th, 2017), and ended just before freezing (i.e., September 25th, 2017) at three stormwater ponds in Calgary, Alberta, Canada: McCall Lake, Country Hills Stormwater Facility, and Inverness Stormpond. At each pond, four (i.e., McCall Lake and Inverness) or five (i.e., Country Hills) locations were sampled 41 times. During sampling, 500 mL of water was collected into two sterile 250 mL Nalgene polyethylene bottles (Nalgene, Rochester, NY, USA) (Systems Plus, Baden, Ontario, Canada) by hand using sterile gloves. Following collection, were sent for pathogen analysis to the University of Alberta (Edmonton, AB, Canada) within 24 hours.

## Culture Methods for Bacterial Water Quality Indicators

Traditional water quality indicators were analyzed using defined substrate culture methods (IDEXX, Westbrook, ME, USA) for *E. coli* and total coliform detection using Colilert Quantitray-2000®. All samples were analyzed according to guidelines from the manufacturer. The Quantitrays® (IDEXX, Westbrook, ME, USA) were sealed and incubated for 24 hours at 35˚C. Following incubation, most probable numbers (MPN) were determined by scoring the number of positive wells (Colilert® – yellow wells = total coliforms, yellow and fluorescent = *E. coli*) and transforming results into quantitative estimates using IDEXX MPN charts. If the Quantitrays® were positive for *E. coli*, they were set aside for further analysis.

Enumeration of fecal coliform bacteria was performed by the Provincial Laboratory in Calgary, Alberta, Canada, in which water (10 mL) was filtered onto a 15x16 mm membrane that was placed on a Membrane Fecal Coliform (mFC) plate (Dalynn Biological, Calgary, AB, Canada), then incubated at 44.5 ˚C for 24 hours. Fecal coliform bacteria were enumerated according to standard operating procedures for the Provincial Lab of Public Health and in alignment with current practices of recreational water assessment in the province of Alberta, Canada.

## Molecular Methods

DNA extraction. Immediately upon receipt of stormwater samples in the laboratory, 20 mL of water was filtered onto 0.4 micron polycarbonate filters by EPA method 1611 (Environmental Protection Agency, 2012) by the Provincial Laboratory in Calgary. Filters were then stored at -80˚C until they were shipped to Edmonton, Alberta, Canada on dry ice.

Controls. In order to ensure amplification, plasmids or positive samples were used as positive controls. IAC was used as a control. DNA free water was used as a negative control, prepared in the same manner as water samples to ensure no carryover.

qPCR Assays. qPCR assays included targets for bacterial indicators (i.e., *Enterococcus* spp.), microbial source tracking markers (i.e., human (HF183 and HumM2), cattle (Rum2Bac), seagull (LeeSg), Canada goose (CG01), dog (dog3), and muskrat (MuBac)), as well bacterial pathogen-related genes (i.e., *Campylobacter* spp. (Van Dyke), *Salmonella* spp. (InvA), and *A. butzleri* (HSP60)). Two markers were used for human contamination to ensure that the results were not due to cross-reaction with other fecal sources (e.g., dog, turkey, and chickens)(Green, 2014). Each marker has a different limit of detection as assayed previously in the laboratory. Each assay target species, target locus, and primer/probe names and their corresponding sequences are specified.

Amplification was performed on an Applied Biosystems TaqMan 7500 fast real-time PCR system (Applied Biosystems, Foster City, CA, USA). The reaction mixture was made to a final volume of 20 µL. The calibrator control standards and plasmid controls were performed in triplicate. The negative controls, *Enterococcus* spp., and salmon DNA PCR assays were performed in duplicate. Sample assays were only performed once. All reactions were carried out in a MicroAmp Fast Optical 96-Well Reaction Plate (Applied Biosystems, Foster City, CA, USA). Samples were pulsed down at maximum speed in a centrifuge prior to thermal cycling. Thermal cycling conditions were 50°C for two minutes; 95°C for 30 seconds (holding); followed by 45 cycles of 95°C for three seconds; and 60°C for 30 seconds for all assays, except for the muskrat marker MuBac. The annealing temperature of MuBac was 57°C. The threshold cycle (CT) was set at 0.1 for Entero1 and Sketa, and 0.05 for all other targets. For further analysis, all CT values were entered into Microsoft Excel.

DNA plasmid standards for each assay were developed previously in the laboratory. Briefly, DNA targets were PCR amplified, run on a 2% agarose gels and amplicons extracted by a QIAquick (Qiagen, Hilden, Germany) gel extraction kit. The products were cloned into pCR2.1-TOPO (Thermo Fisher Scientific, Waltham, MA, USA) per manufacturer’s instructions using TOP10 F’ *E. coli* competent cells. The plasmids were mini-prepped using QIAprep spin mini prep kit (Qiagen, Hilden, Germany). Plasmids were then quantitated using the Qubit 2.0 fluorimeter (Invitrogen Carlsbad, CA, USA) and diluted to 100,000 copies/µL stocks. Thus, stocks of plasmid DNA were aliquoted and stored at -80°C. During each qPCR analysis, 10-fold serial dilutions were made from the known concentration plasmid for each individual target.

## Data import and cleaning

The data was cleaned to format the variables (e.g., DNQ, not detects, >2419.6) in order to deal with limits of detection and detection limits. In the raw data, detectable but not quantifiable (DNQ) had not been assigned a value. See supplementary material - cleaning.

## Statistical Methods

For further analysis, all CT values were entered into Microsoft Excel and volume corrected pior to data analysis using R statstical software. Data from *E. coli*, *Enterococcus*, *fecal coliforms*, and all qPCR marker concentrations were log10 transformed prior to analysis because the Shapiro-Wilk test for normality indicated the data was non-normally distributed data. Water quality needed to be evaluateded based on a variety of parameters, but most importantly temporal and spatial variability, and violations of water quality standards. Violin plots were used as they allow us to see the overal trends in the data. In addition to line graphs, which had an emphasis on temporal trends on any seasonality. Bar graphs were also used to visualized overall marker concentrations as sampling locations and sampling site, putting a greater emphasis on spatial trends. All of the aforemention plots were created using the ggplot2 package in R. A correlation matrix was used to determine which of the abundant microbial source tracking markers were found to be related to *A. butzleri* contamination, and was created using the corrplot package in R. Furthermore, this plot showed relatedness between microbial source tracking markers. Several linear models were run in order to test the hypothesis that the source of fecal material in stormwater ponds would affect *A. butzleri*. The same microbial source tracking markers that were found to be abundant sources of contamination had linear models run on them. In addition, a random forest tree was created using the caret packing in R. The tree was created using method repeated cv, and assessed by the measure accuracy. In addition, the tune length was adjusted to try to accomodate more variables.

# Results

## Water Quality

A high-level descriptive overview of the bacteriological water quality in each of these ponds, and at each of the sites, is provided in Figure 1, and is based on the percentage of samples violating water quality standards/guidelines, as evaluated against: the USEPA’s recreational water quality guideline for *Enterococcus* spp. by molecular methods (Environmental Protection Agency, 2012). Water quality was also evaluated against Alberta’s former recreational water quality standards based on thermotolerant coliform concentrations was also used to evaluated water quality (supplemental material Figure 1). A number of observations are worth noting from this high-level analysis.

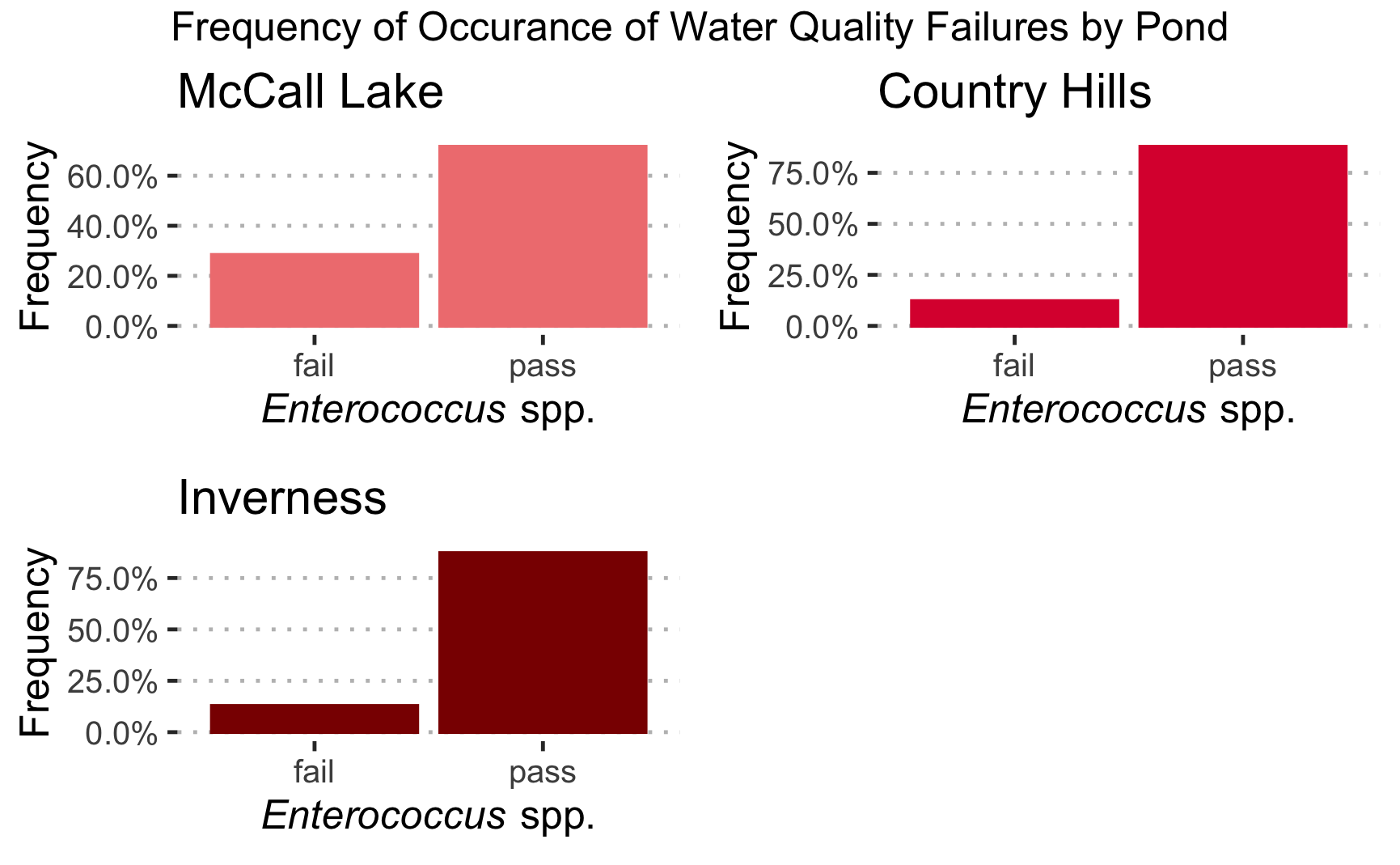


Figure 1: Frequency of occurrence of samples positive based on water quality indicator *Enterococcus* spp. in stormwater ponds in Calgary, Alberta.

Firstly, considerable spatial variation was observed with respect to the frequency of water quality failures among the urban stormwater ponds, with McCall Lake appearing to be the most contaminated of the three storm ponds. This result was true regardless of the bacterial water quality indicator chosen for analysis (i.e., *Enterococcus* or thermotolerant coliforms [see supplemental material, Figure 2]). Approximately 29% of all water samples taken at McCall Lake failed water quality guidelines for *Enterococcus* at the recommended STV or geomean values set out in the guidance documents. Inverness stormpond had the fewest water quality violations among the three ponds, also based on all bacteriological indicators examined, and therefore was considered to have the best water quality overall.

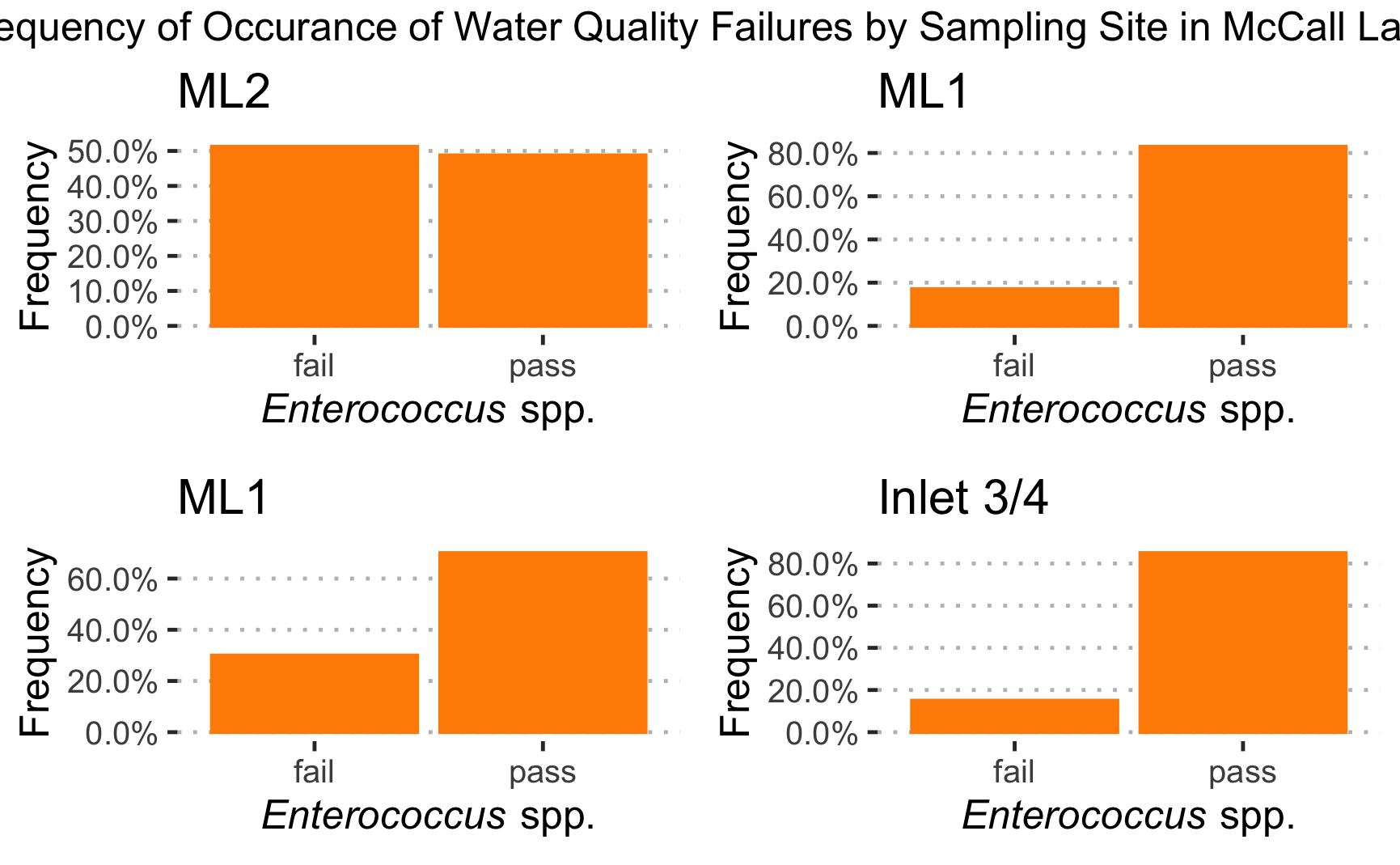


Figure 2: Frequency of occurrence of samples positive based on water quality indicator *Enterococcus* spp. broken down by sampling site at McCall Lake in Calgary, Alberta.

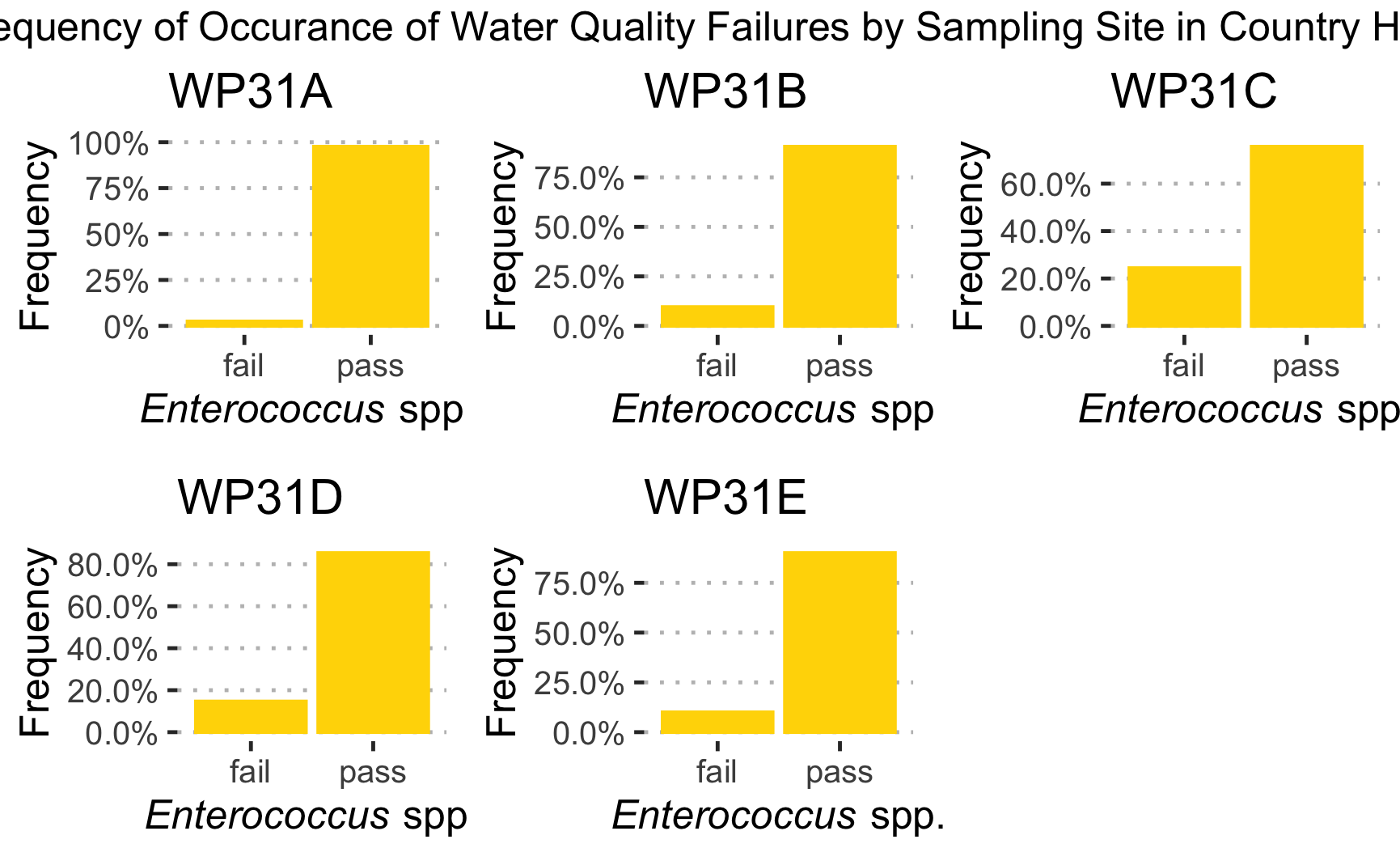


Figure 3: Frequency of occurrence of samples positive based on water quality indicator *Enterococcus* spp. broken down by sampling site at Country Hills in Calgary, Alberta.

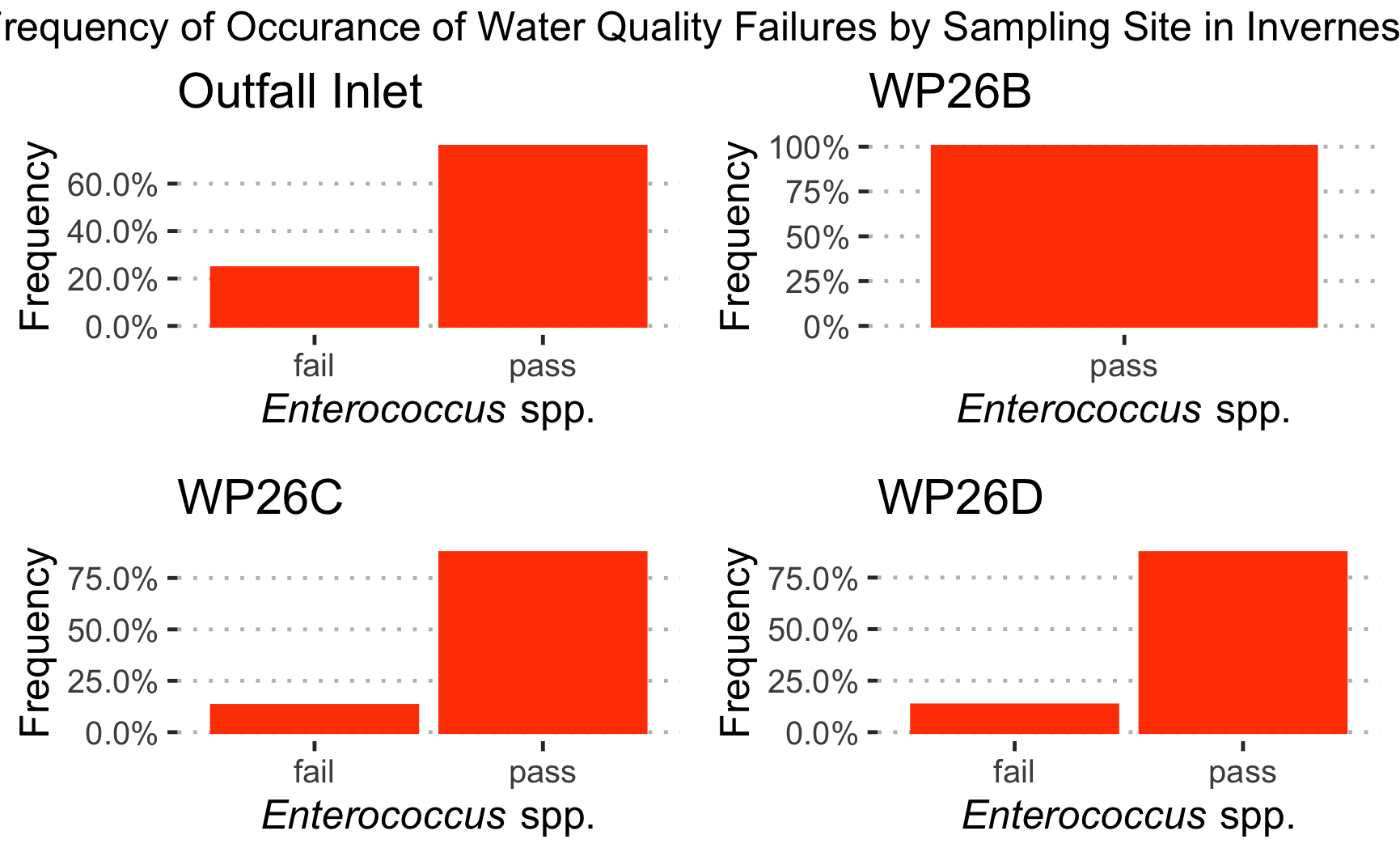
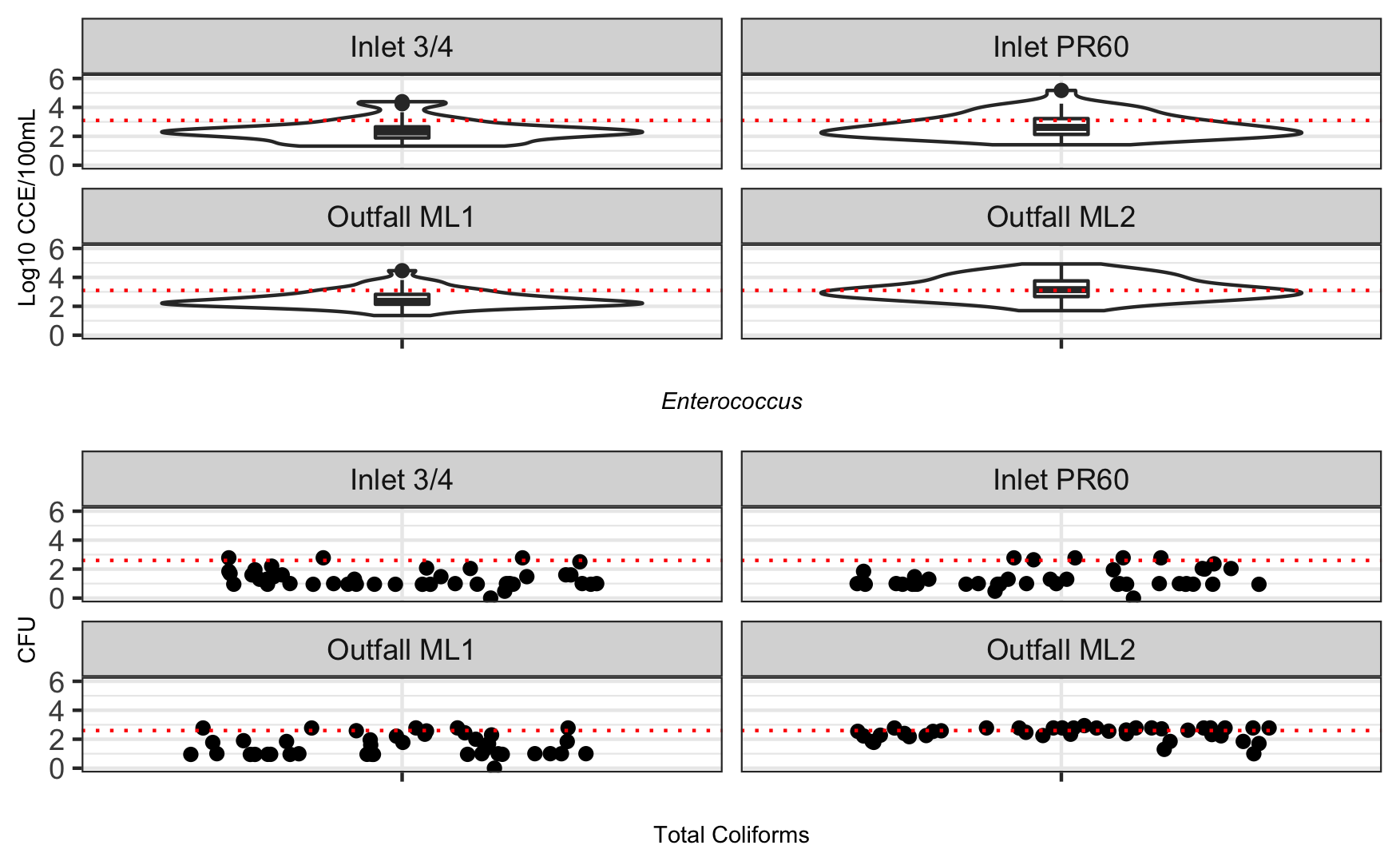


Figure 4: Frequency of occurrence of samples positive based on water quality indicator *Enterococcus* spp. broken down by sampling site at Inverness in Calgary, Alberta.

However, variation in bacteriological water quality was also observed among sampling sites within a single pond. The most contaminated site across all stormwater ponds examined was site ML2 at McCall Lake, with upwards of 51% of all samples failing the US EPAs Guidelines for Recreational Water Quality for Enteroccus spp. STV (Figure 2). This site had the poorest water quality irrespective of the bacterial indicator used in the analysis (supplemental material figure 3). It is important to note, however, that ML2 was an above-grade outfall, thereby potentially explaining the more frequent bacteriological failures at this site as due to the fact that water samples were directly collected from the outfall and not after dilution into the pond.

Bacteriological water quality indicators were further assessed for all sampling sites (see supplemental material figures 1-4), however due to the poor water quality at McCall Lake this assessment is included in the report.

 Median levels of *Enterococcus* at the ML2 site approximated 3.1 log10 CCE/100 mL, whereas at all other sampling sites in McCall Lake (i.e., ML1, Inlet 3/4, and PR60), the median occurrence was almost an order of magnitude lower (~ 2.3 log10 CCE/100 mL) (Figure 5). Incidentally, ML2 also had the largest overall interquartile variation in the concentration of *Enterococcus* during the study season (Figure 5). Concentrations of thermotolerant coliforms were also high at this site and followed a similar trend to that of *Enterococcus* (supplemental material).

It is important to note that in most cases for *Enterococcus* at sites other than ML2, there were several outliers in the data set (Figure 5).Although outliers may reflect recent localized contamination events not necessarily reflective of overall water quality in the stormwater pond (e.g., aquatic birds in one area of the pond), their occurrence could also reflect the periods of peak contamination in stormwater ponds, and for which this effect may be contingent on temporal variables associated with water quality (e.g., first flush from storms, to be discussed later). Specifically, outliers for *Enterococcus* concentrations were represented by values higher than ~3.5 log10 for Inlet 3/4, ~3.75 log10 for ML1, and ~4 log10 for PR60.

This analysis revealed that stormwater ponds in southern Alberta, Canada, do not often meet existing guidelines or standards as laid out by USEPA and Alberta Environments and Parks in terms of recreational water quality, surface water quality and/or irrigation water quality (Figure 5). For this reason, we sought to determine the potentiasl sources of poor water quality using microbial source tracking assays for humans, ruminants, dogs, gull, geese and muskrat.

## Microbial Source Tracking

Calgary stormwater ponds were mainly impacted by human and gull feces (Figure 6-7, see supplemental material for exact numbers). The human specific markers, HF183 and HumM2, were detected at 42% and 16%, of samples, respectively in McCall Lake (Figure 6, see supplemental material for exact numbers). On the lower end, they were dtected in 14% and 3% of samples at Inverness. The gull specific marker (i.e., LeeSg) was found in 15% of samples at McCall Lake (Figure 7, see supplemental material for exact numbers). Of these, the more dominant source of fecal pollution was from humans (Figure 6, see supplemental material for exact numbers). All other host-specific markers (i.e., dog, Canada geese, muskrat, and ruminants) were detected in <5% of pond samples (supplemental material figures 9-11).

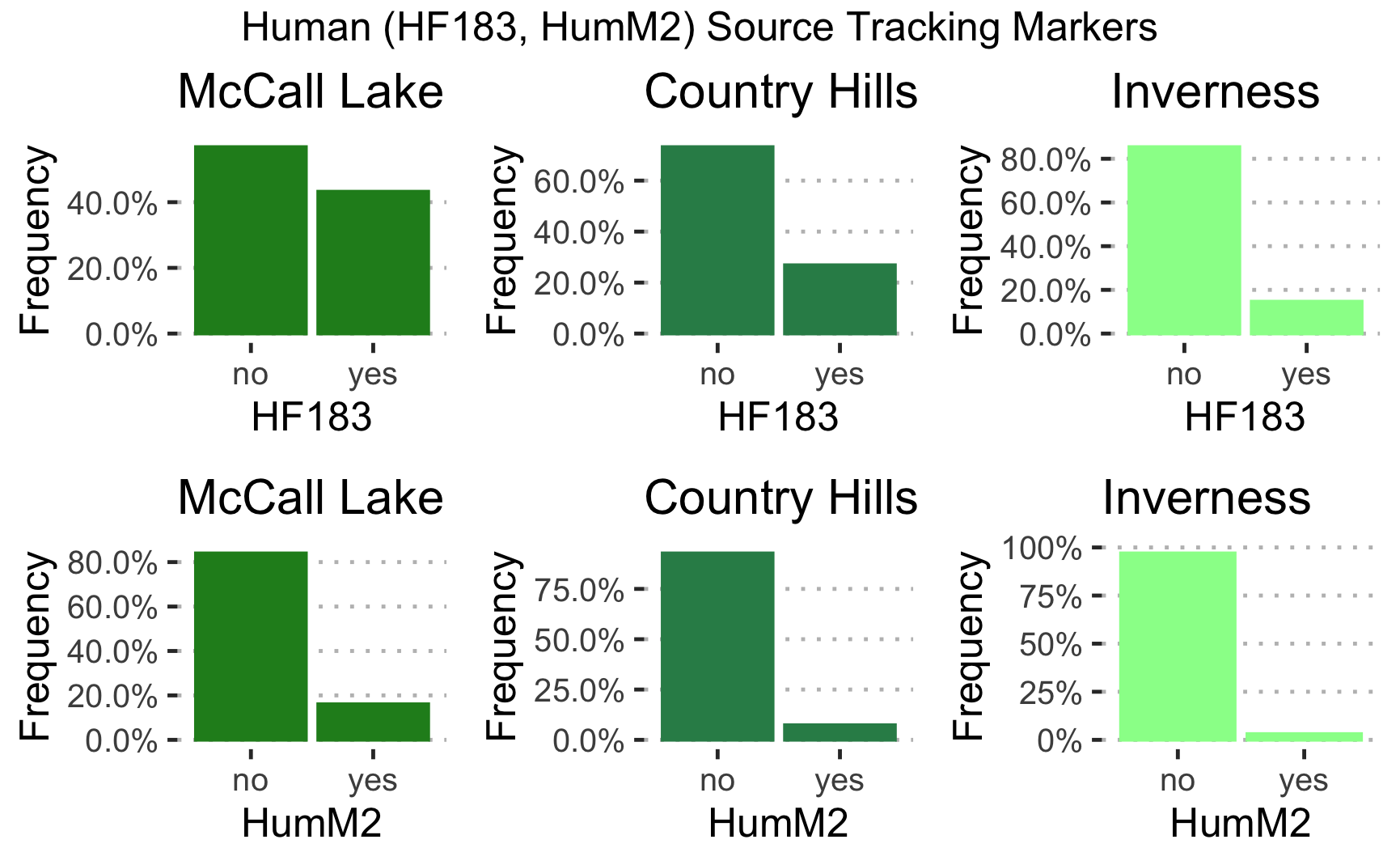


Figure 6: Frequency of occurrence of samples positive based on MST marker HF183 and HumM2 for human fecal contamination broken down by sampling site in Calgary, Alberta.

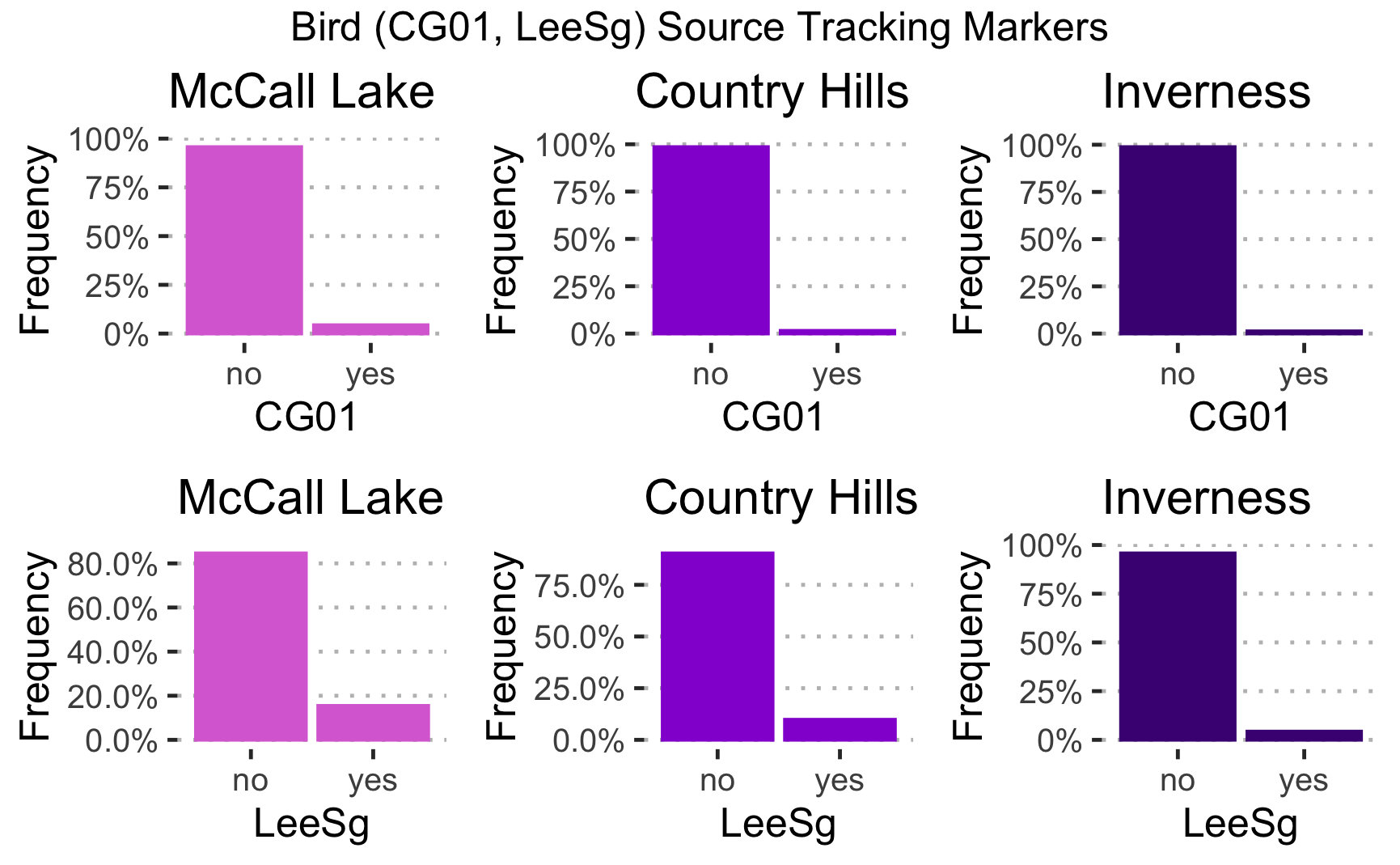


Figure 7: Frequency of occurrence of samples positive based on MST marker CG01 and LeeSg for human fecal contamination broken down by sampling site in Calgary, Alberta.

In congruence with finding that ML2 at McCall Lake was the most frequently contaminated site with human feces (Figure 6), this site also had the greatest median concentration of the human fecal marker HF183 (i.e., 4.2 log10 copies/100 mL) observed across all three stormwater ponds and sampling sites in these ponds (Figure 8). In comparison, all other McCall Lake sampling sites had a median concentration of HF183 at ~3.4 log10 copies/100 mL (i.e., close to the quantification limit of the assay (red dotted line)) (Figure 8). Specifically, at ML2, there was a single outlier in the data set for HF183, represented by a value of 6.0 log10 copies/100 mL (Figure 8). However, although ML2 represented the most consistently contaminated sampling site with human fecal contamination at McCall Lake, all other sites in this thesis appeared to be at risk for human fecal contamination (supplemental material).

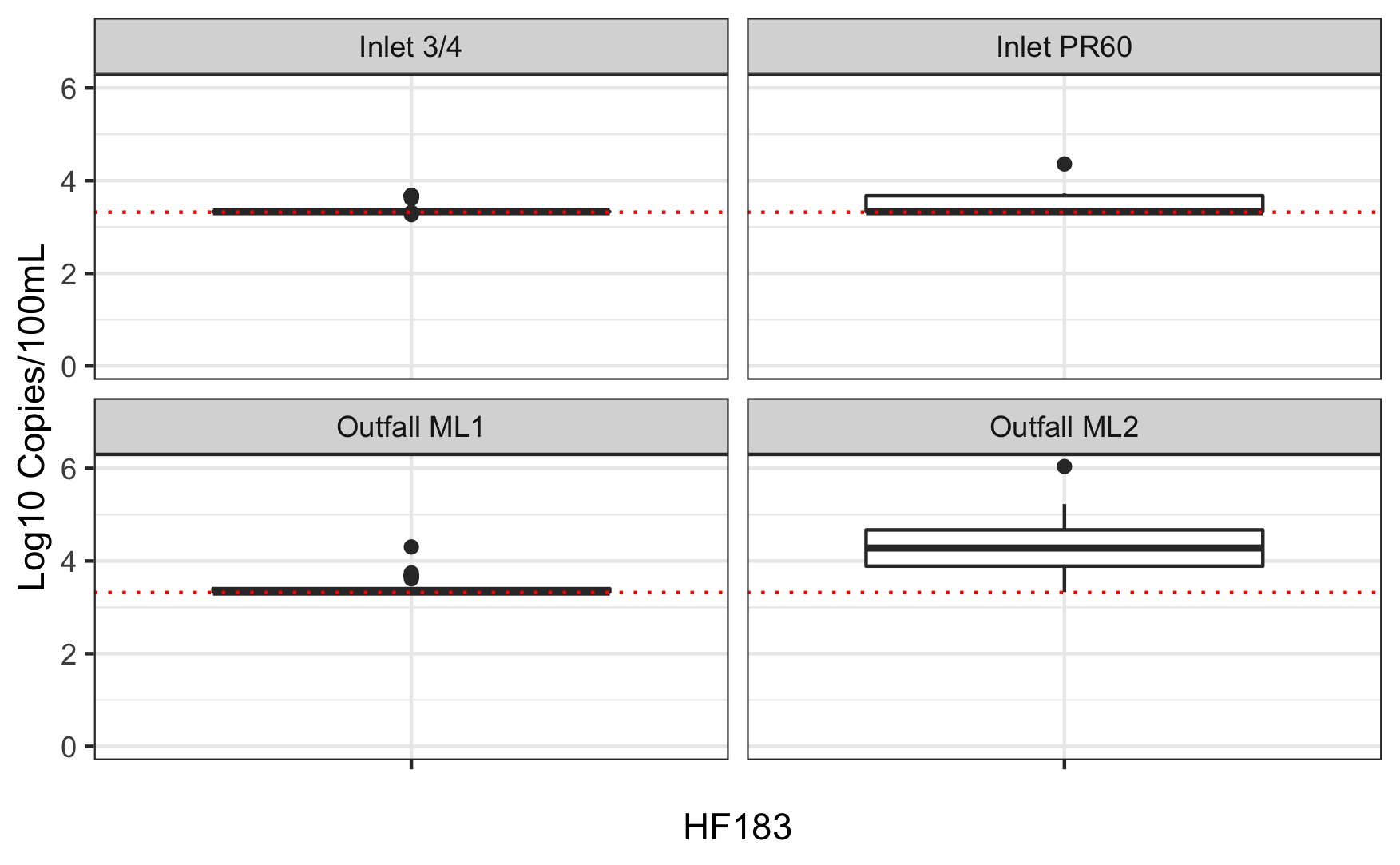


Figure 8: Box and Whisker Plot of HF183 levels by sampling site in McCall Lake (ML2 n=38, ML1 n=6, PR60 n=13, Inlet ¾ n= 5). The outer edges of the box represent the 25th and 75th percentiles (i.e., interquartile range), and the line within the box represents the median. The location of median indicates the skew of the data. The whiskers represent the interquartile range\*1.5. The outliers are determined by being greater or less than 1.5 times the upper of lower interquartile ranges as represented by circles.

Temporal fluctuations in human fecal pollution markers were noted between the stormwater ponds, and among the sampling sites within a stormwater pond (Figure 9).

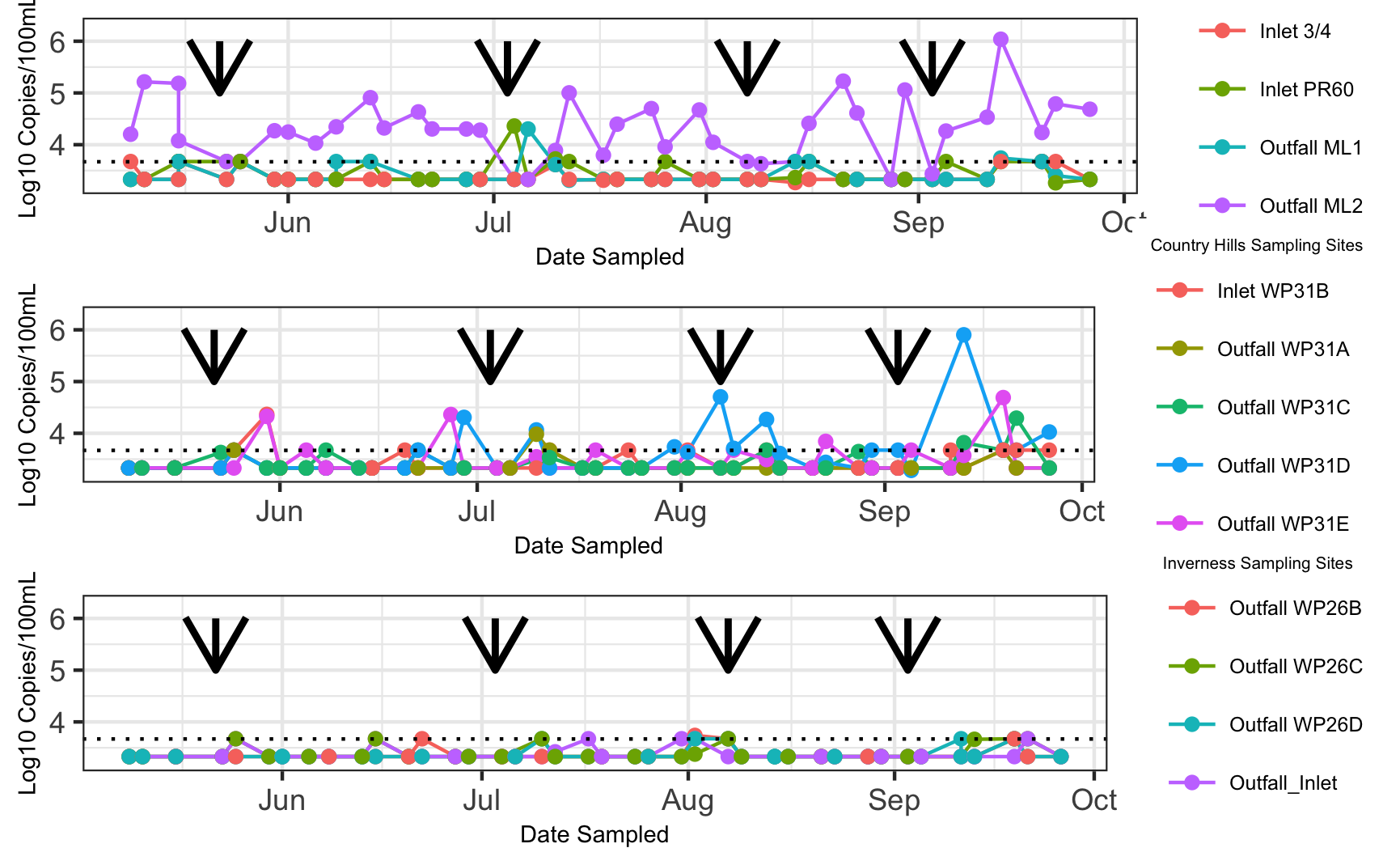


Figure 9: Temporal pattern of occurrence HF183 log10 concentrations at all sampling sites in McCall Lake (top), Country Hills (middle), and Inverness (botton) over the 21-week sampling season. The limit of quantification95 (LOQ95) as a black dotted line. The black arrows represent long holiday weekends.

Of all stormwater pond sampling sites, ML2 at McCall Lake experienced the most consistent temporal pattern of human fecal contamination throughout the sampling season. For example, within the 41 sampling dates, over the 21-week sampling season, there were only two sampling dates in which we did not detect HF183 at ML2 (i.e., July 4th and August 28th) (Figure 9). However, there were other sampling dates when levels of HF183 decreased to a non-quantifiable level at ML2 (i.e., May 23rd, May 25th, August 8th, and August 14th). Interestingly, this pattern tended to occur after long weekends (i.e., holidays occurring on the following Mondays: May 22nd, July 3rd, August 7th, and September 3rd), and three of these long weekends corresponded to decreases in human fecal contamination markers on the following day of sampling (i.e., May 23rd, July 4th, and August 8th, which were Tuesdays). This suspicious temporal pattern of contamination suggested that the levels of human fecal contamination may have been related to industrial/commercial activities, as the levels of human fecal contamination decreased during times when industries/commercial premises may have been closed for the holidays.

Human fecal contamination at the sampling sites was often highly variable between sequential sampling dates. For example, at Inlet PR60 in McCall Lake, within a two-week span, biweekly HF183 values fluctuated between undetectable levels (i.e., June 29th and July 6th) and 4.3 log10 copies/100mL (i.e., July 4th) and 3.5 log10 copies/100mL (i.e., July 10th). This high variability in human fecal contamination markers over sequential sampling dates, elicits potential concerns regarding the sporadic nature of contamination and the stability of water quality in the urban stormwater ponds.

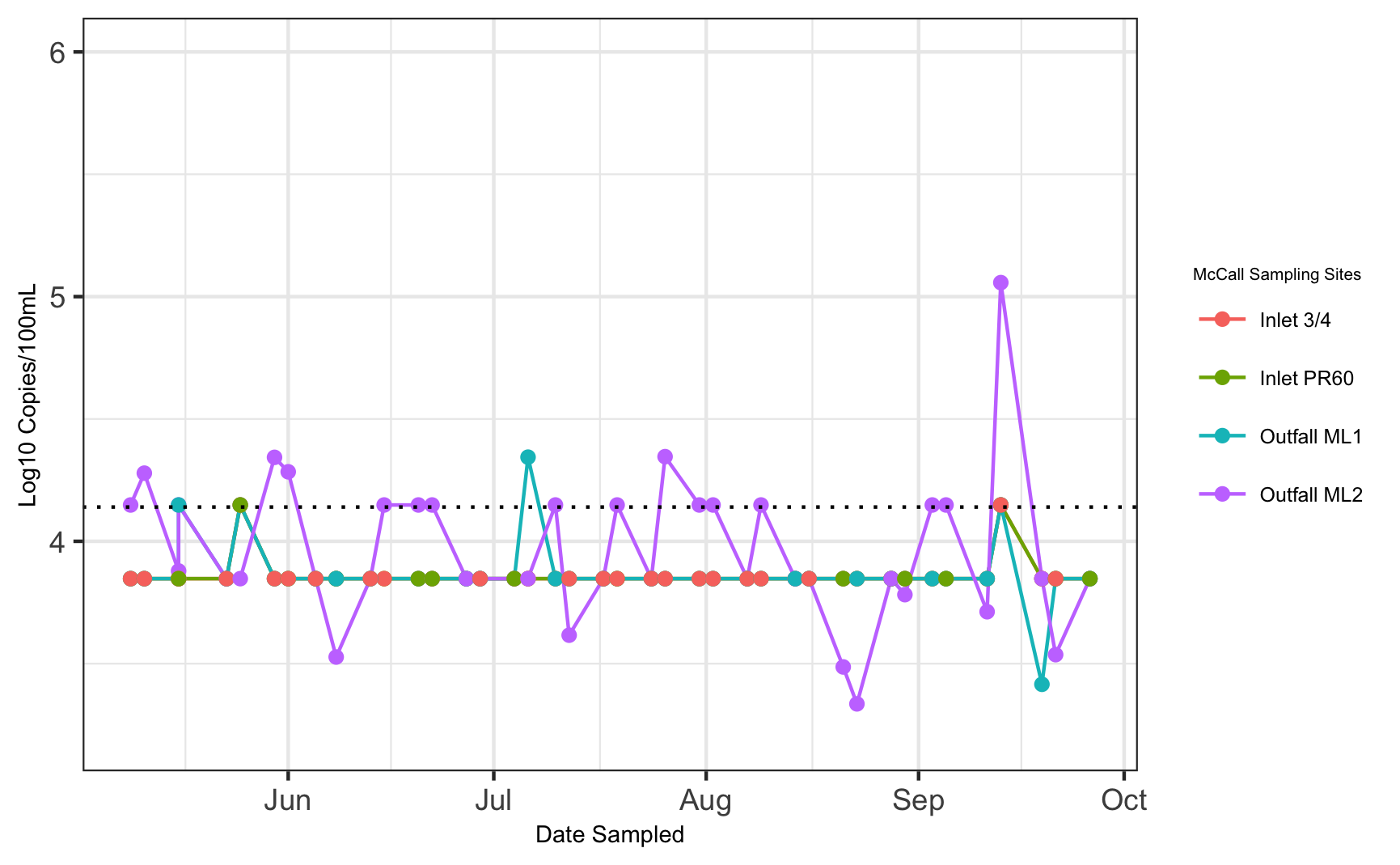


Figure 10: Temporal pattern of occurrence HuMm2 log10 concentrations at all sampling sites in McCall Lake over the 21-week sampling season. The limit of quantification95 (LOQ95) as a black dotted lines.

HumM2 was detected less frequently and at lower concentrations in all of the urban stormwater ponds tested (supplemental material). ML2 had the highest occurrence of HumM2 detections of all McCall Lake sampling sites, which corresponded with the findings with the human fecal contamination marker HF183 (Figure 10). Furthermore, Inlet ¾ had the lowest occurrence of HF183 in McCall Lake, and was also tied for the lowest occurrence of HumM2 in McCall Lake.

In addition to seeing temporal and spatial trends in human fecal contamination, temporal and spatial trends in seagull contamination were also noted. In McCall Lake, seagull fecal contamination was considered to be a sporadic, highly variable, source of pollution. Seagull contamination was first noted in McCall Lake at the end of June, and tended to be episodic (Figure 11). For example, at ML2, seagull fecal contamination was detected on July 12th and at a level of 4.1 log10 copies/100 mL, and then it was not detected at quantifiable levels again until August 8th (i.e., 3.7 log10 copies/100 mL) (Figure 11). Although ML2 was most frequently positive for detection of seagull fecal contamination among all sampling sites, this pattern of sporadic, highly variable findings were also noted at the other McCall Lake sites (i.e., PR60, ML1, and Inlet ¾).

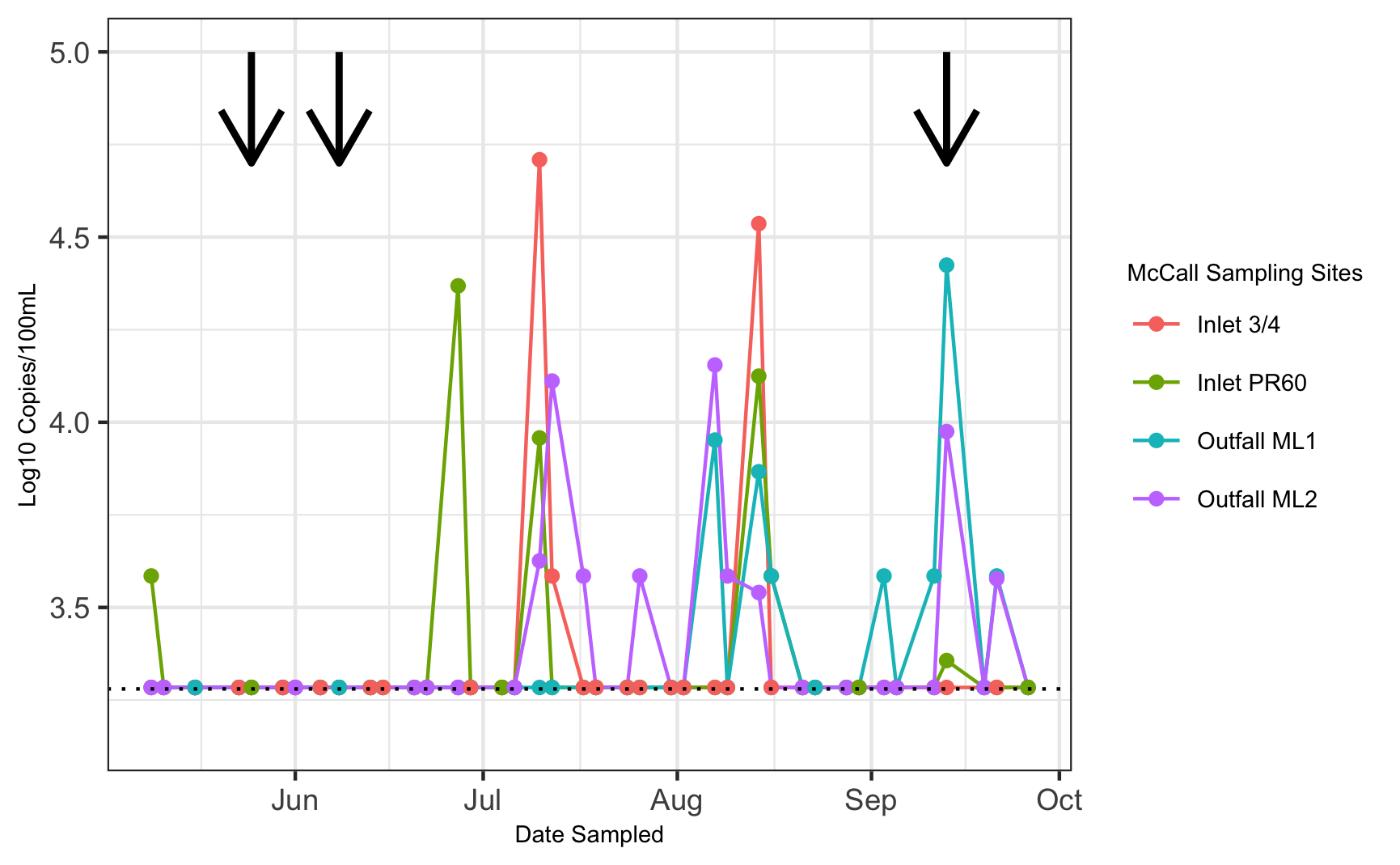


Figure 11: Temporal pattern of occurrence of LeeSg concentrations at all sampling sites in McCall Lake over the 21-week sampling season. The limit of quantification95 (LOQ95) as a blue dotted line. The black arrows represent greater than 10 mm of rain in the previous 72 hours.

Overall, there were patterns of similarity in seagull contamination regarding temporal trends. These patterns were similar across the sampling sites in McCall Lake, which was noted for two key reasons. Firstly, there were three instances where seagull fecal contamination occurred concurrently at three or more McCall Lake sites (i.e., July 10th, August 14th, and September 13th). Secondly, on the aforementioned sampling dates, the levels of seagull fecal contamination detected were all within one order of magnitude of each other. These patterns suggested that a potential environmental variable linked the contamination along the sampling sites at McCall Lake. One potential environmental variable examined was antecedent rainfall. Only three dates (i.e., May 25th, June 8th, and September 13th) had greater than 10 mm of rain. Seagull fecal contamination was detected on only one of the sampling dates (September 13th), though at three sampling sites on this date.

A high-level overview of the frequency of several enteric bacterial pathogens (i.e., *A. butzleri*, *Campylobacter* spp., *Salmonella* spp., and STEC) in each of the Calgary urban stormwater ponds (i.e., McCall Lake, Country Hills Stormwater Facility, and Inverness Stormpond), and at each sampling site within the ponds is provided in Figure 12. The most frequently detected bacterial pathogen found in stormwater ponds was *A. butzleri*, detected in 36% of samples at McCall Lake, 24% of samples at Inverness and 18% of samples at Country Hills (Figure 12, see supplemental materials for exact percents). The second most common pathogen detected was *Campylobacter* spp. (4% at Country Hills) and *Salmonella* spp. (1% at McCall Lake) (Figure 12).

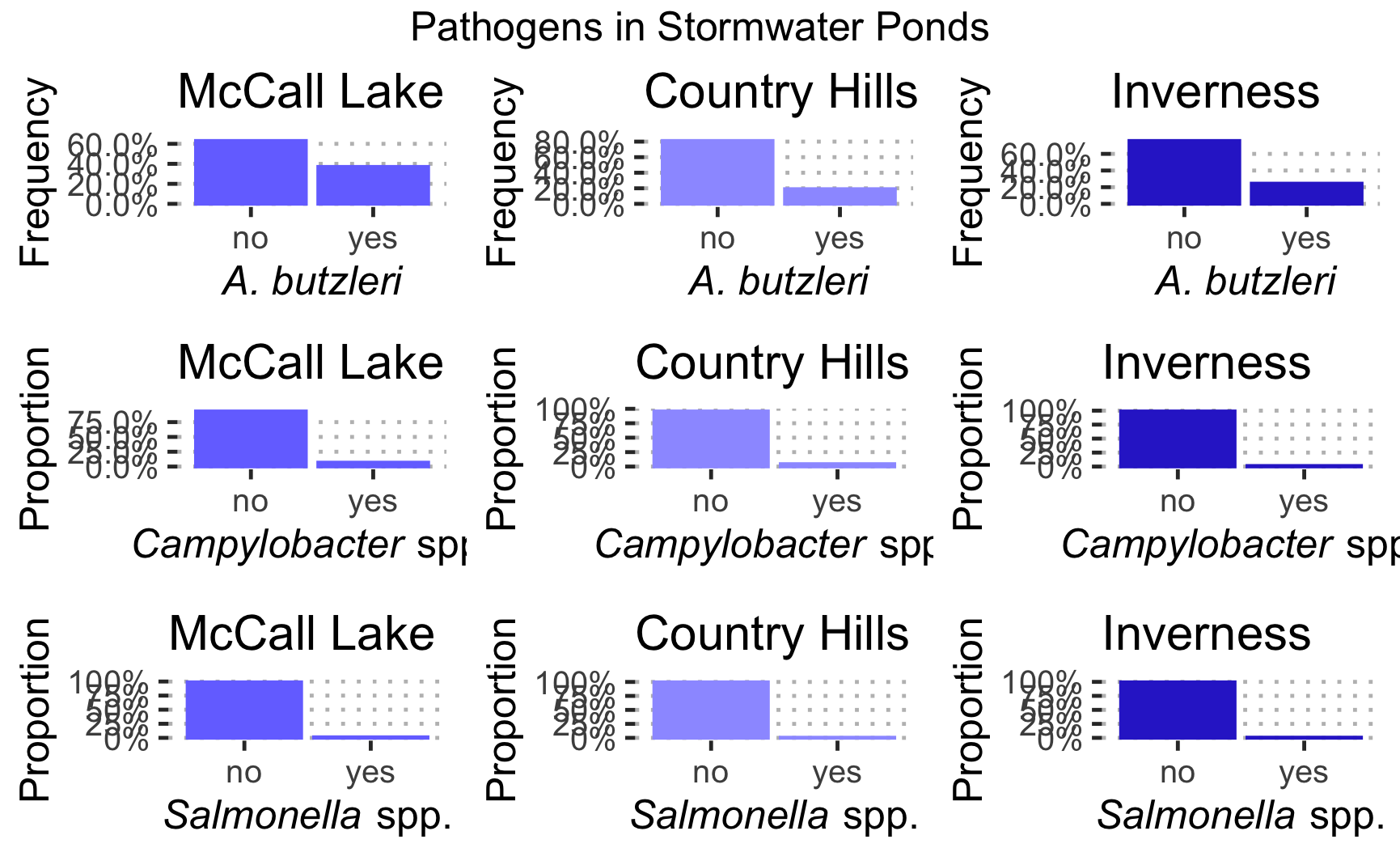


Figure 12: The frequency of occurence of the enteric bacterial pathogen *A. butzleri* *Campylobacter* spp., *Salmonella* spp., in the Calgary, Alberta stormwater ponds.

To better understand temporal variation, we further examined patterns of occurrence based on molecular qPCR results. Notable temporal fluctuations in *A. butzleri* were observed between the urban stormwater ponds, and among the sampling sites within a pond (*BELOW*). We found that at Inlet ¾, in McCall Lake, considerable temporal fluctuations were detected in the levels of *A. butzleri* between sequential sampling dates. Within a two-week time period (i.e., four sequential sampling dates, June 20th – June 29th), the concentration of *A. butzleri* varied from being not detected (i.e., below the limit of quantification of 3.5 log10 copies/100 mL) on June 20th, then spiking to 3.9 log10 copies/100 mL on June 22nd, to be not detected on June 27th, and spiking again to 4.3 log10 copies/100 mL on June 29th (Figure 13).

We tracked environmental variables that could contribute to temporal fluctuations in *A. butzleri* concentrations (e.g., antecedent rainfall data, temperature, etc.). Of note, we recorded three sampling dates that had rainfall greater than 10 mm (i.e., May 25th, June 8th, and September 13th, Figure 5 1). We noted that *A. butzleri* was detected at all McCall Lake sampling sites on several sampling dates, July 10th, August 14th, August 16th, September 13th, of which September 13th had significant rainfall (Figure 5 1). However, on another rainfall date (i.e., May 25th) *A. butzleri* was not observed at any of the sampling sites, and on June 8th, *A. butzleri* concentrations reached detectable levels only at the outfalls (i.e., ML1 and ML2) (Figure 13).

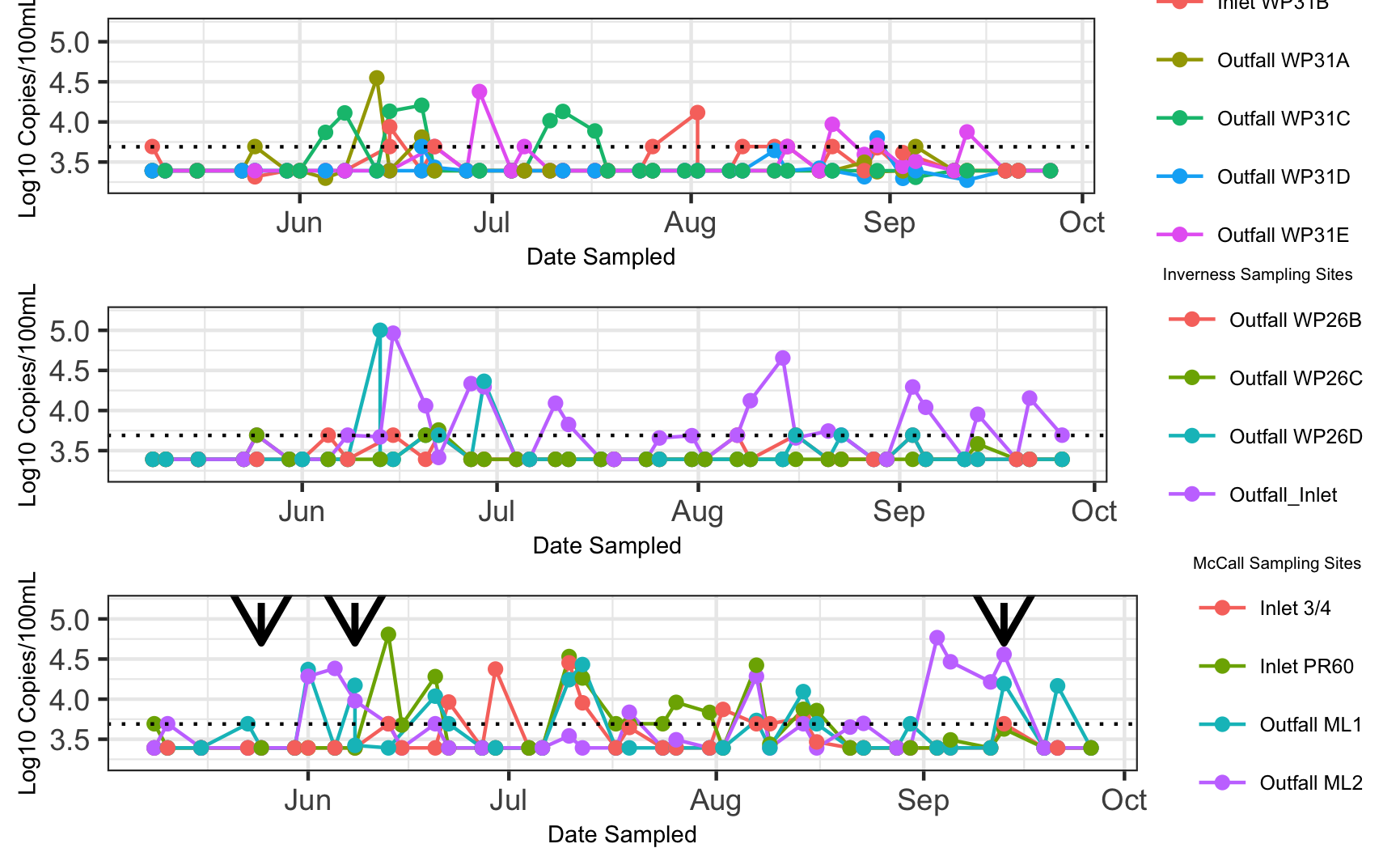
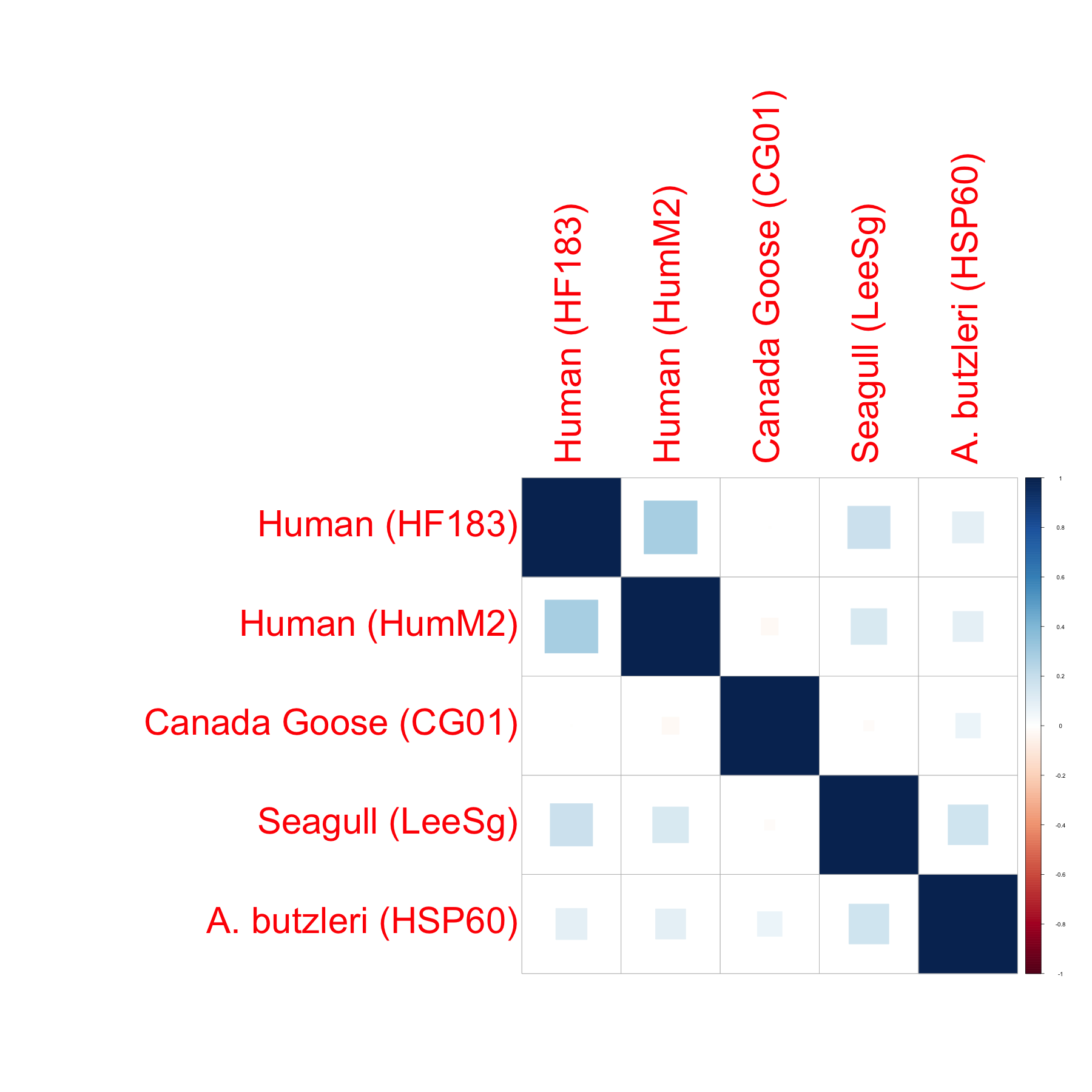


Figure 13: Temporal pattern of occurrence *A. butzleri* log10 concentrations at all sampling sites at Country Hills (top), Inverness (middle), and McCall Lake (bottom), over the 21-week sampling season. The limit of quantification95 (LOQ95) as a black dotted line. The black arrows represent rainfall greater than 10 mm.

Due to the prevalence and abundance of *A. butzleri* contamination, we sought to determine the potential sources of its contamination. Water samples were analyzed by identifying which microbial source tracking markers occurred most often with *A. butzleri* detections. We found that the most common source of pollution co-occurring with *A. butzleri* detection was human fecal pollution. The human marker HF183 was the most frequently found marker with *A. butzleri*. (Figure 14). The second most dominant source of fecal pollution was seagull (i.e., LeeSg) (Figure 14).

 Several linear models were run in order to test the hypothesis that the source of fecal material in stormwater ponds would affect *A. butzleri*. Based on previous analyses, linear models were only run ont the most abundant microbial source tracking markers (i.e., HF183, HumM2, LeeSg, and CG01) (Figure 15). The linear model for human fecal contamination (i.e., HF183 and HumM2), rejects the null hypotheses based on p-value. This indicates thats human fecal material is a signifant predictor of our enteric bacterial pathogen, *A. butzleri*. Combining this with the correlation plot previously presented, and temporal occurance plots previously shown, we can suggest that HF183 and HumM2 affect *A. butzleri* contamination, which be discussed later on. In addition, we reject the null hypothesis for LeeS, which also aligns with our correlation plot and temporal occurance plots. However, we were not able to reject the null hypothesis for CG01 as the p-value is greater than .05. This also aligns with our previous analysis, as this outcome was least associated to our predictor *A. butzleri*.

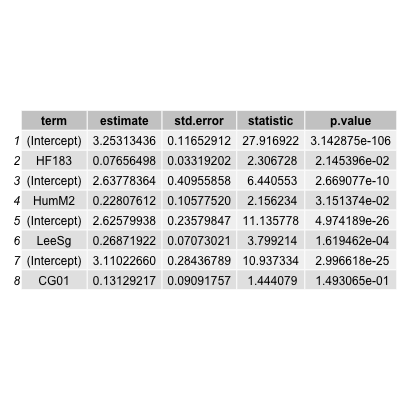


Figure 14: Linear model for predicting A. butzleri, based on HF183 or HumM2 (i.e., humfan fecal contamination), LeeSg (i.e., seagull fecal contamination) and CG01 (i.e.,Canada Goose fecal contamination

For this paper, two random forest analyses were performed. The first of which aims to predict the outcome *A. butzleri* (Figure 16). For this tree, we will treat week 0 as a null value, as if the pond is always contaminted with *A. butlzeri* we would not need to look any further. If *Enterococcus* spp. is less than 3.2 than we *A. butlzeri* is no. However if it is greater than 3.2 log10 CCE/100mL, rainfall is more 1 mm and also has *E. coli* greater than 1.7 log10 CFU/100mL than *A. butlzeri* may be positive. This is reflective of what FIB (i.e., *E. coli*, *Enterococcus* spp.) should be doing, is being in greater numbers when the pathogen is present (i.e., *A. butlzeri*). However, this tree has been influenced by the number of samples in this study (~700) as majority of random forest based on water quality data encompass ~10,000 samples (Tesorieo, 2017).

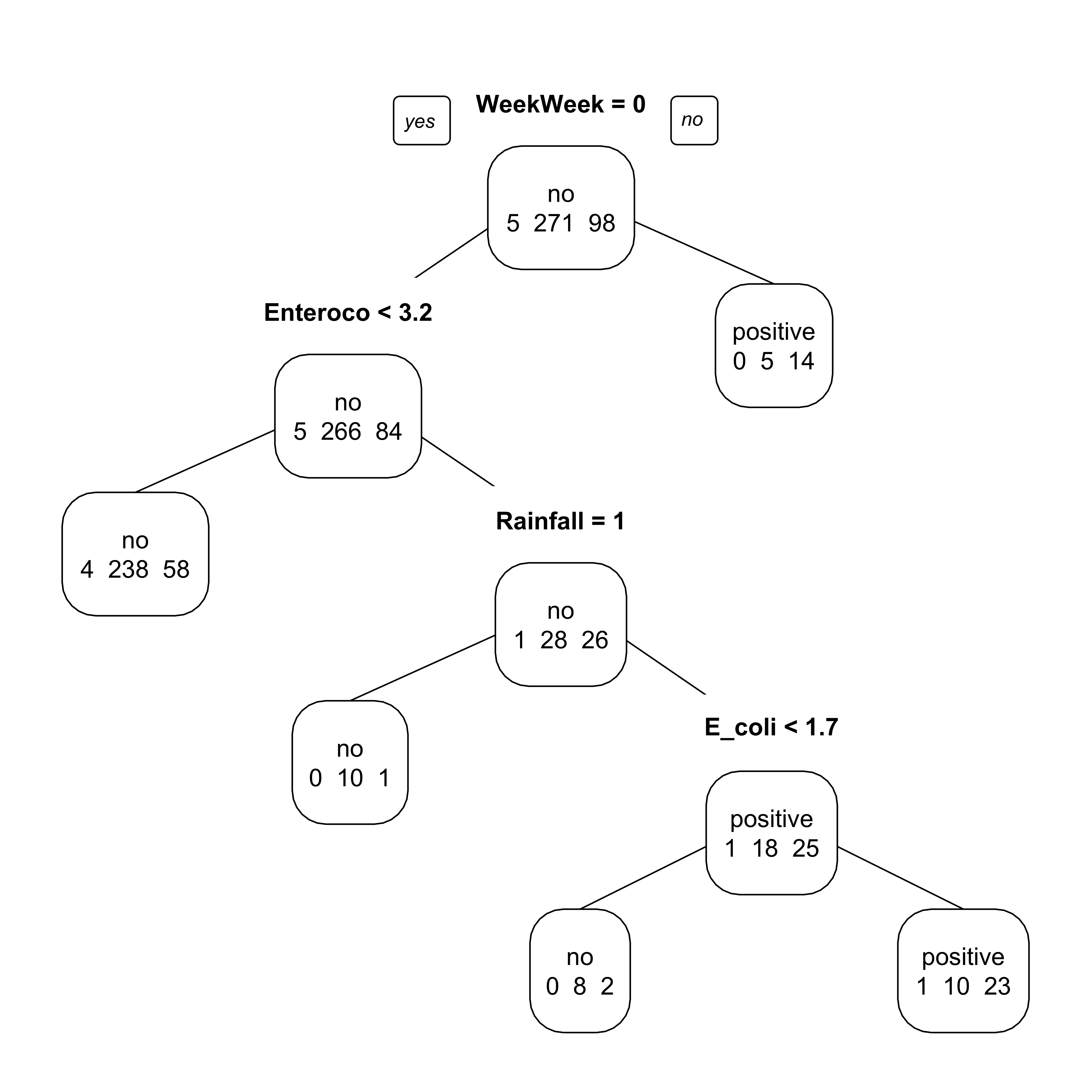


Figure 15: Random forest based on the outcome of *A. butlzeri*.

To further our understanding of contamination by stormwater pond, a random forest analysis was performed. Our outcome variable was the pond, and our predictors were all other variables excluding sampling site. This tree is helpful in that other scientists can use it to help assess the water quality of their study sites, based off of this tree (Figure 17). Furthermore, it allows us to gain a greater understanding of the sources of conatimantion effecting each stormwater pond we tested. This tree tells us if HF183 is greater than 3.9 log10 copies/100mL than the pond is McCall Lake. Based on our earlier analysis of how McCall Lake is the most conaminted sampling location by HF183, this is accurate. If HF183 is less than 3.9 log10 copies/100mL and *E. coli* is greater than or equal 0.87 CFU/100 mls than the pond is also contaminted with *A. butzleri*, otherwise the next predictor is the sampling date. If the sample has greater than 3.6 log10 copies/100mL of *A. butzleri* and *Enterococcus* levels higher than 2.5 log10 CCE/100ml than the sampling location is Country Hills. This is also reflective of the aforementioned data, as Country Hills is more contaminated than Inverness but less contaminated than McCall Lake. In addition, if *A. butzleri* is less than 3.6 log10 copies/100mL, than it may be contamined with seagull fecal material. This is also reflective of the aforementioned data, as we found *A. butzleri* to have a higher co-occurance with human fecal markers (i.e., HF183 and HumM2) than bird fecal markers (i.e., LeeSg or CG01). Following this point, our tree is further broken down by all the water quality indicators (i.e., *Enterococcus*, *E. coli*, thermotolerant coliforms). Moving back towards the top of tree we note that the third decision point is based off of date sampled, and then thermotolerant coliforms. All of these decision points lead to Inverness. This is reflective of the previous data presented as Inverness was the least contaminated of all sampling locations. Furthermore, we find that the following decision point is total coliforms. This is unsurprising, as total coliforms are found to be abundant in recreational waters and are no longer considered to be preditors of gastroitenstinal illness when assessing water quality. As such, the US EPA has moved away from using total coliforms. Additionally, it should be noted that this half of the tree soley relies on water quality indicators and does not create nodes based off microbial source tracking markers or pathogen specific qPCR markers. This is further reflective of the results that were previously presented, as Inverness has the best water quality of all stormwater ponds tested and did not consistently have high levels of any microbial source tracking marker tested or pathogen specific qPCR marker.

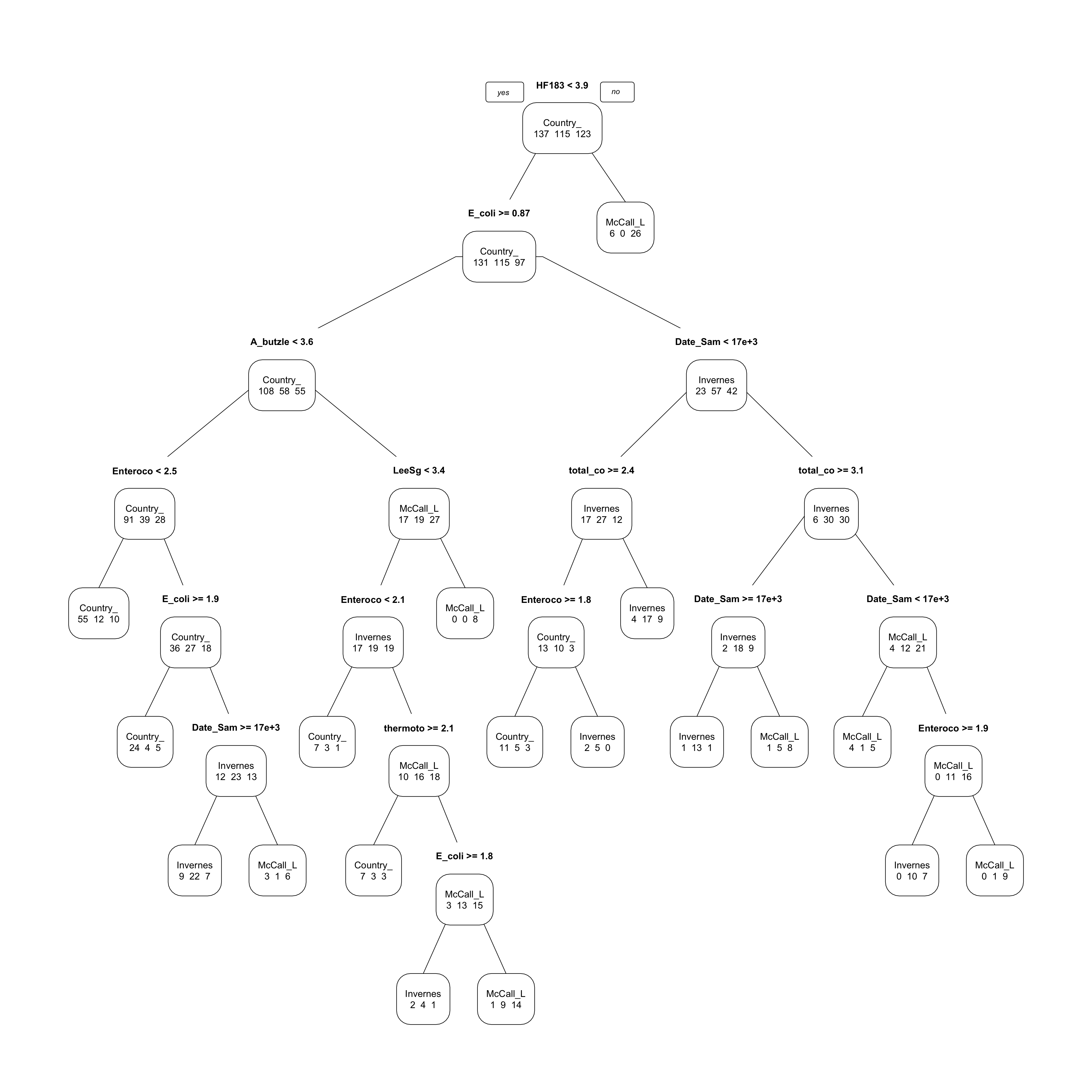


Figure 16: Random forest based on the outcome of pond (i.e., McCall Lake, Inverness, Country Hills.)

# Discussion

## Summary and Interpretation

To the best of our knowledge, this is the first report on the occurrence of *A. butzleri* in stormwater ponds. A select group of stormwater ponds were sampled in Calgary, Alberta, Canada, based on their opportunities for water reuse. Based on direct molecular testing, *A. butzleri* was detected in 25% of all water samples. In addition, microbial sources of fecal pollution, other select enteric bacterial pathogens, and FIB were assessed through molecular methods and analyzed.

Currently, there is limited knowledge on *A. butzleri* in stormwater. However it has been detected in many different types of water, ranging from rivers and wells (Wesley, 2000) (Fong, 2009) (Van Driessche & Houf, 2005) to saltwater lakes and coastal seawater (Wesley, 2000) (Fong, 2009) (Van Driessche & Houf, 2005), and even drinking water reservoirs (Wesley, 2000) (Van Driessche & Houf, 2005). (Banting, 2016) detected *A. butzleri* in 54% of irrigation water samples in Alberta, Canada, through an MPN-qPCR assay, using hsp60 as their target. In addition, Webb et al. (Banting, 2016) found that raw sewage had the highest density of *A. butzleri* in two wastewater treatment facilities in southwestern Alberta, Canada. (Collado, 2010) tested 12 sampling sites along the Llobregat River in Catalonia, Spain; and at nine of the sampling sites, *Arcobacter* spp. was detected in 100% of samples, and at one site it was not detected in any of the samples. (Collado, 2008), in their study of a fecally-contaminated freshwater stream, found the highest amount of *Arcobacter* spp. (i.e., 3.7 x 105 MPN/100mL) to be at the sampling location closest to the wastewater treatment.

At individual sampling sites, concentrations of pathogens in stormwater samples occurred over a narrow range of values through culture-based methods. In our study, *A. butzleri* measured through culture-based methods at ML2 ranged from 0.9-93 MPN/300mL. As previously mentioned, there is limited research on the *Arcobacter* spp. detections through culture-based methods in stormwater; however, prior studies have found culturable levels of *A. butzleri* to be as high as 105 MPN/100mL in raw sewage (Banting, 2016), which is considerably higher than the levels detected in our study. The finding of *Arcobacter* spp. in sewage or water impacted by raw sewage is not uncommon [McLellan2010] (Collado, 2008) (Khan, 2009). (Collado, 2008), in their study in Spain, found the presence of *A. butzleri* in 58% of river water samples and in 100% of sewage samples. In addition, another study detected *A. butzleri* in 100% of domestic sewage samples in the United Kingdom (Merga, 2014). Furthermore, they demonstrated that *A. butzleri* in urban sewage can survive treatment and therefore has the potential to be released into environmental bodies of water (Collado, 2010).

When direct molecular methods were used, *A. butzleri* was detected at levels as high as 4.7 log10 genome copies/100mL at sampling site ML2. (Webb, 2016) found the density of *Arcobacter* spp. to range from 101.5 to 104 log10 genome copies mL-1 in treated sewage by molecular-based methods with primers developed by (Webb, 2016). (Lee, 2012) found the levels of *A. butzleri* to range from 1 x 102.7 to 1 x 105 gene copies/100mL at one Lake Erie beach over a 3-month period (i.e., July – August) through molecular-based methods using ArcoI and ArcoII primers developed by (Bastyns, 1995). The occurrence of such ranges of enteric bacterial pathogens at individual sampling sites can pose a unique challenge to the development of a stormwater treatment facility and sampling plan.

Understanding the spatial differences at stormwater sampling sites can allow us to better determine which sampling sites would be best suited for stormwater reuse applications. Our study revealed that some urban stormwater ponds have more consistent pathogen detections than other stormwater ponds: McCall Lake had the most *A. butzleri* detections, in comparison to Inverness Stormpond and Country Hills Stormwater Facility. (Talay, 2016) tested 115 different water samples (i.e., sewage, rivers, spring water, and drinking water) by molecular-based methods from Izmir Turkey. They found that the prevalence of *Arcobacter* spp. was highest in river water (i.e., 52% of samples), and that all drinking water samples were negative (Talay, 2016). (Webb, 2016) in their testing of two different wastewater treatment plants in southwestern Alberta, Canada, found higher densities of *A. butzleri* in the Lethbridge wastewater treatment plant than in the Fort Macleod treatment plant through molecular-based methods.

Moreover, differences in methods may not be only variable affecting *A. butzleri* occurrences and concentrations. A study by (Fera, 2010) indicated that *Arcobacter* spp. may survive better at lower temperatures. (Lee, 2012) found that *Arcobacter* spp. detections in recreational water was higher in September at Lake Erie in North America, and suggested that the levels show a negative correlation with the temperature of the water through molecular methods. Conversely, (Webb, 2016) reported lower densities of *A. butzleri* in December and March from samples from wastewater treatment plants located in southwestern Alberta, Canada, by molecular-based methods. However, some studies did not find a seasonal effect on enteric pathogens (Rechenburg, 2009). That said, environmental variables are not the only factor influencing enteric bacterial pathogens in stormwater-impacted bodies of water.

## Strengths and Limitations

The analytitical methods used in this analysis were similar to those use in speciality specific journal in the field of water and health (e.g., Journal of Water and Health, Frontiers in Water, Water Quality Research, Water and Environment). This is a strength of this research study, as it makes it less complicated for scientists in the field to compare the fecal indicator bacteria (FIB), enteric bacterial pathogens, and microbials sources of fecal pollution with their findigs, which is a common practice in assessing water quality. Additionally, the analysis used to evaluate the aforementioned variables allow scientists to gain a better understanding of the variables beyond just frequency of occurance. For example, for FIB jitter plots and violin plots were made that reflect outliers, medians, and how these relate to US EPA or Alberta Environment water quality standards. Another important aspect in water quality is temporal vararation, and how environmental variables may affect any temporal fluctuations we see. Temporal variability was assessed for our main pathogen of concern (i.e., *A. butzleri*), in addition to any other variables that may affect *A. butzleri* contamination (e.g., microbial source tracking markers), while overlaying environmental variables (i.e., rainfall). The analysis enabled us to overlay all sampling locations and ponds, which is important when determing in-study comparisons. Moreover, the analysis in this study was taken one step further than majority of studies in the field by providing a comprehensive supplemental material section that provides all analyses for all ponds and sampling sites.

There are several alternative ways this data may have been analyzed. For example, when analyzing FIB we could have also provided temporal trends, however, as that was not the focus of this study. Furthermore, FIB may also have been analzyed by box and whisker plots instead of violin plots, as these two plot types are very similar. In this study temporal trends were analyzed using line graphs, however an alternative would have been to use bar charts. A lot of data has been presented in this paper in both in both table formats (in the supplemental material) and figure format in the publication, in order to provide the most information.

Models are not often used when assessing water quality data through molecular methods, and when they are far more complicated than the scope of this class (Tesorieo, 2017) (Smith, 2010) (Casanovas-Massana, 2015). In addition, majority of random forest studies in the water microbiology encompass over 10,000 samples and this study had ~700 samples. However, as modeling was a key component of this class, several models have been included in the publication. A linear model was used to predict *A. butzleri* contamination based on several microbial source tracking markers that were present in the stormwater ponds we sampled.

Our study focused on molecular-based methods for the identification of *A. butzleri*, in addition to other pathogens and microbial source tracking markers in urban stormwater ponds. Molecular-based methods are fast, less labor-intensive and can be less biased than culture-based methods. Using molecular-techniques allowed us to assess ~800 samples for a wide vareity of microbial sources of fecal polution, enteric bacterial pathogens, and water quality indicators. Furthermore, this study entailed the entire time from thaw to freeze, when stormwater poses a risk to human health and a potential reuse opporuntity. Having a long sampling season, ina addition to obtaining samples biweekely allowed us to fully assess temporal trends. Additionally, by assessing a variety of stormwater ponds we were able to gather a deeper understanding of water quality issues facing urban stormwater ponds in Alberta.

Although molecular-based methods provide a lot of infomration, these culture-independant techniques do not differentiate between living and dead cells. As such, other studies have found molecular methods to underestimate *A. butzleri* prevelance by as much as 50% (Beaudry et al., unpublished) when comparing to culture-based methods. As such, the prevelance of *A. butzleri* in this paper may be widely underestimated, and may be as high as 75%. Furthermore, qPCR methods do not provide us with any information on the physiological state or activities of the bacterium, which could be further studied using metatranscriptomics or culture-methods. In addition, as *A. butzleri* is a putative pathogen future work needs to be done in order to determine the bacteriums pathogenic potential. For this research specifically, any isolates of *A. butzleri* obtained could be tested for virulence genes using homologs of those found in *Campylobacter* spp.

## Conclusions

Our study found that *A. butzleri* was the most common pathogen present; and a growing number of researchers are suggesting that the clinical prevalence of the *Arcobacter* species is probably underestimated since the species is not routinely tested for (Levican, 2013) (Douidah, 2011) and primers that have been used to identify *Campylobacter* spp. cross react with *Arcobacter* spp. (Banting, 2016). Further, our study reflected that the *A. butzleri* found in stormwater ponds harbors many virulence genes regardless of source of fecal contamination, and should therefore be treated as pathogenic. The prevalence and levels of *A. butzleri* need to be taken into consideration when developing an urban stormwater sampling plan and stormwater treatment.

# References

Adams, B., N. (2016). Shiga toxin producing escherichia coli o157, england and wales, 1983-2012. *Emerging Infectious Diseases*, *22*, 590–597.

Banting, B., G. (2016). Evaluation of various campylobacter- specific quantitative pcr (qPCR) assays for detection and enumeration of campylobacteraceae in irrigation water and wastewater via a miniaturized most-probable-number-qPCR assay. *Applied and Enviornmental Microbiology*, *82*, 4743–4756.

Bastyns, C., K. (1995). A variable 23S rDNA region is a useful discriminating target for genus-specific and species-specific pcr amplification in arcobacter species. *Applied Microbiology*, *218*, 353–356.

Bower, S., P. A. (2005). Detection of genetic markers of fecal indicator bacteria in lake michigan and determination of their relationship to escherichia coli densities using standard microbiological methods. *Applied and Environmental Microbiology*, *7*, 8305–8313.

Casanovas-Massana, G.-D., A. (2015). Predicting fecal sources in waters with diverse pollution loads using general and molecular host-specific indicators and applying machine learning methods. *Jounral of Environmental Management*, *51*, 317–325.

Collado, I., L. (2008). Presence of arcobacter spp. In environmental waters correlates with high levels of fecal pollution. *Environmental Microbiology*, *10*, 31635–31640.

Collado, K., L. (2010). Occurrence and diversity of arcobacter spp. Along the llobregat river catchment, at sewage effluents and in a drinking water treatment plant. *Water Research*, *82*, 3696–3702.

Craun, C., G.F. (2005). Outbreaks associated with recreational water in the united states. *International Journal of Environmental Health Research*, 243–262.

Curtis, &. T., K. (2016). Examining the colonization and survival of e. Coli from varying host sources in rainage basin sediments and stormwater. *Environmental Contamination Toxicology*, *71*, 183–187.

Douidah, Z., L. (2011). Occurence of putative virulence genes in arcobacter species isolated from humans and animals. *Journal of Clinical Microbiology*, *3*, 735–741.

Fera, G., M. (2010). Specific detection of arcobacter spp. In es- tuarine waters of southern italy by pcr and fluorescent in situ hybridization. *Letters in Applied Microbiology*, *50*, 65–70.

Fong, M., T.-T. (2009). Massive microbiological groundwater contamination associated with a waterborne outbreak in lake erie, south bass island, ohio. *Environmental Health Perspectives*, *37*, 856–863.

Green, H., H.C. (2014). Improved hf183 quantitative real-time pcr assay for characterization of human fecal pollution in ambient surface water samples. *Applied and Environmental Microbiology*, *80*, 3086–3094.

Hafliger, H., D. (2013). Outbreak of viral gastroenteritis due to sewage-contaminated drinking water. *International Journal of Food Microbiology*, *54*, 123–126.

Harwood, S., V. J. (2014). Microbial source tracking markers for detection of fecal contamination in environmental waters: Relationships between pathogens and human health outcomes. *Federation of European Microboligical Societies Microbiology Reviews*, *38*, 1–40.

Kayman, A., T. (2012). Emerging pathogen arcobacter spp. In acute gastroenteritis: Molecular identification, antibiotic susceptibilities and genotyping of the isolated arcobacter. *Journal of Medical Microbiology*, *61*, 1439–1444.

Khan, L., I. (2009). Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. *Environmental Microbiology*, *12*, 378–492.

Krometis, C., L. (2010). Comparison of the presence and partitioning behavior of indicator organisms and salmonella spp. In an urban watershed. *Water Research*, *8*, 44–59.

Layton A., W. D., McKay L. (2006). Development of bacteroides 16S rRNA gene taqman-based real-time pcr assays for estimation of total, human, and bovine fecal pollution in water. *Applied and Environmental Microbiology*, *72*, 4214–4224.

Lee, A., C. (2012). Arcobacter in lake erie beach waters: An emerging gastrointestinal pathogen linked with human- associated fecal contamination. *Journal of Clinical Microbiology*, *78*, 5511–5519.

Levican, A., A. (2013). Adherence to and invasion of human intestinal cells by arcobacter species and their virulence genotypes. *Applied Environmental Microbiology*, *79*(16), 4951–4957.

McCarthy, &. G., N. (2001). Incidence of guillain-barré syndrome following infection with campylobacter jejuni. *American Journal of Epidemiology*, *153*, 610–614.

Meng, C., Z. (2018). Stormwater constructed wetlands: A source or sink for campylobacter spp. *Water Research*, *131*, 218–227.

Merga, R., J. (2014). Arcobacter spp. Isolated from untreated domestic effluent. *Letters in Applied Microbiology*, *108*, 974–984.

Moore, C., J. (2001). Occurrence of campylobacter spp. In water in northern ireland: Implications for public health. *International Journal of Environmental Health Research*, *70*, 102–207.

Rangel, S., J. (2005). Epidemiology of escherichia coli o157:H7 outbreaks, united states, 1982-2002. *Emerging Infectious Diseases*, *11*, 603–609.

Rechenburg, &. K., A. (2009). Sewage effluent as a source of campylobacter spp. In a surface water catchment. *Environmental Health Research*, *19*, 239–249.

Shanks, K., O. (2009). Quantitative pcr for genetic markers of human fecal pollution. *Applied Environmental Microbiology*, *75*, 5507–5513.

Smith, S.-B., A. (2010). Novel application of a statistical technique, random forests, in a bacterial source tracking study. *Water Resources Research*, *44*, 4067–4076.

Soller, S., J. (2017). Incidence of gastrointestinal illness following wet weather recreational exposure: Harmonization of quantitative microbial risk assessment with an epidemiological investigation of surfers. *Water Research*, *121*, 280–289.

Staley, C., Z. R. (2018). Fecal source tracking and eDNA profiling in an urban creek following an extreme rain event. *Scientific Reports*, *8*, 14390.

Talay, M., F. (2016). Isolation and identification of arcobacter species from environmental and drinking water samples. *Folia Microbiology*, *61*, 479–484.

Tesorieo, G., A. (2017). Predicting redox-snestive contamnat concentrationsin groudnwater using random forest classification. *Water Resources Research*, *53*, 25–32.

Vandenberg, D., O. (2004). Arcobacter butzleri infections report diarrhea associated with abdominal pain. *Water Research*, *109*, 1053–1066.

Van Driessche, E., & Houf, K. (2005). Survival capacity in water of arcobacter species under different temperature conditions. *Vetinary Microbiology*, *105*, 149–154.

Webb, B., A. L. (2016). Comparative detection and quantification of arcobacter butzleri in stools from diarrheic and nondiarrheic people in southwestern alberta, canada. *Journal of Clinical Microbiology*, *54*, 1082–1088.

Wesley, W., I. (2000). Fecal shedding of campylobacter and arcobacter spp. In dairy cattle. *Applied and Enviornmental Microbiology*, *66*, 4241–4246.

Wuertz, W., S. (2011). Library-independent bacterial source tracking methods. *Microbial Source Tracking: Methods, Applications, and Case Studies*.