Pairing scent tracking canines and high-resolution mass spectrometry to discover novel chemical tracers of wastewater contamination in surface waters

Alex D. Gipe

A thesis

Submitted in partial fulfillment of the

Requirements for the degree of

Master of Science

University of Washington

2018

Committee

Christopher Simpson

C. Andrew James

Program Authorized to Offer Degree

Environmental and Occupational Health Sciences

University of Washington

Abstract

Pairing scent tracking canines and high-resolution mass spectrometry to discover novel chemical tracers of wastewater contamination in surface waters

Alex D. Gipe

Chair of the Supervisory Committee
Christopher Simpson
Department of Environmental and Occupational Health Sciences

The Centers for Disease Control and Prevention reported 6,939 deaths and 477,000 emergency department visits annually during 2009 to 2013 were attributed to 13 waterborne pathogens. The annual cost of hospitalization and treatment associated with these diseases totaled $3.8 billion. To combat water body impairment and limit the spread of disease federal, state, and local agencies work to identify and remediate sources of microbial contamination in surface waters through microbial source tracking (MST) utilizing microbial and chemical tracers. The objective of this study is to identify novel chemical tracers of microbial contamination which could be utilized in MST efforts, furthering the ability of public health professionals to identify and remediate impaired waters. High-resolution mass spectrometry was used to analyze water samples from nine surface water sites within the Samish watershed which had been classified by a scent tracking canine regarding their presence of wastewater contamination. Chemicals unique to canine positive sites and those detected at a two-fold greater peak abundance in these sites were identified at varying confidence levels. A group of eight chemicals, including 2-mercaptobenzithizole, which were detected in three of five canine positive sites were identified as potential chemical tracers of wastewater.

Table of Contents

Figures

Tables

Introduction

In 2016 an estimated 829,000 deaths were attributed to diarrhea caused by waterborne pathogens, which represents 1.9% of the global burden of disease.[1] Beyond the global health burden, waterborne disease has an estimated economic cost of 12 billion dollars per year. [2] Though typically thought of as an issue of developing nations, the United States of America (USA) faces its own disease burden caused by waterborne pathogens. The Centers for Disease Control and Prevention reported 6,939 deaths and 477,000 emergency department visits annually during 2009 to 2013 were attributed to 13 waterborne pathogens.[3] The annual cost of hospitalization and treatment associated with these diseases totaled $3.8 billion.[4]

Waterborne pathogens are bacteria, virus, protozoa, and helminths which cause illness in humans by contact with contaminated water. Many waterborne pathogens are associated with fecal contamination and enter the environment from natural sources such as wildlife feces, and anthropogenic sources which include direct and indirect sources of sewage discharge. Frequently anthropogenic sources of pathogens enter the environment from inadequately treated sewage, storm water, illicit sewage discharges, and combined sewer overflows (CSO).[5] Once in the environment pathogens are transmitted to humans through ingestion, inhalation, or contact with water containing the infectious agent. Common pathogens transmitted by fecal contaminated waters are *Cryptosporidium*, *Giardia*, *Escherichia Coli* (*E. Coli*) O157:H7, *Vibrio cholerae*, *Salmonella typhimurium*, and *Norovirius*.[6] Typical illnesses associated with waterborne pathogens include diarrhea, vomiting, gastrointestinal disease, and cholera.

Depending on the extent of water contamination, outbreaks associated with waterborne pathogen exposures can cause tens to thousands of cases of illness. In 2000, 2300 gastrointestinal illness cases in Walkerton, Ontario were linked to the presence of *E. Coli* O157:H7 in the city’s drinking water supply.[7] Many factors contribute to waterborne disease outbreaks (WBDOs). The sewage treatment infrastructure which collects, conveys and treats wastewater can be a major contributor to WBDOs due to failing or leaking pipes or inadequate treatment.[6] On-site septic systems are also a contributing factor due to failing systems which leak untreated sewage into ground water. Drinking water sources and treatments also contribute to WBDOs due to inadequate filtration or disinfection of surface and ground water. From 2011 to 2012, drinking water associated WBDOs were responsible for 431 cases of illness and 14 deaths in the USA.[8] Weather also plays a crucial role in the process of contamination entering water sources. Flooding and heavy rainfall events wash contaminates into surface waters. In addition to this, increased stormwater can overwhelm wastewater treatment plants causing CSOs, where untreated sewage is discharged to waterways.[6] This is mainly the case in combined sewer systems where a mixture of stormwater and sewage is conveyed to a wastewater treatment plant through a common infrastructure.

The Safe Drinking Water Act, which regulates drinking water pollutant concentration and treatment, specifies a maximum microbial pathogen concentration of zero in drinking water.[9] This demonstrates the seriousness of pathogen contamination, and establishes that there is no safe lower limit to microbial pathogens in water intended for consumption. Recreational waters, those used to swim, fish, boat or perform other activities in which water contact occurs, are regulated by the Clean Water Act. The Clean Water Act sets the water quality criteria for two microbial indicator species, *Enterococci* and *E. Coli*, which are used to monitor water bodies for the presence of fecal microbial contamination.[10] Criteria for recreational waters are set at a higher (less protective) threshold than drinking water due to their design to protect a person from infection due to primary contact which includes limited ingestion.

Though drinking water and its sources are of primary concern for pathogen contamination and have the highest impact on outbreak potential, recreational waters in the USA are highly impaired by microbial pollution. The Clean Waters Act section 303(d), lists impaired waters and requires the development of Total Maximum Daily Loads (TMDLs) for pollutants to reduce sources in a watershed. Pathogen contamination is the leading cause for impairment for 303(d) listed water bodies.[11] The US Environmental Protection Agency (US EPA) reports that approximately 190,000 miles of rivers, streams and shorelines as well as 575,000 acres of lakes and wetlands are impaired due to microbial pollution in the USA.[11] In the state of Washington where this study takes place, 649 miles of rivers and streams, 35,978 acres of lakes and ponds, and 63.1 miles of ocean coastline are impaired due to microbial pollution.[11] These statistics demonstrate the scope and range of recreational water impairment by microbial pathogens in the USA. Recreational waters support a multitude of uses, providing economic and health benefits. While costs associated with pathogen contamination are difficult to quantify, Breen et al. found evidence that recreational demand of water sites increased with better water quality.[12] The economic loss can more readily be determined due to losses in tourism, aquaculture, and other economic benefits. For example, Weiskel et al. reported a more than 75 million dollar yearly economic loss in Massachusetts due to pathogen contamination of waterbodies.[13]

Microbial Source Tracking

To combat water body impairment federal, state, and local agencies work to identify and remediate sources of microbial contamination. Identification of microbial pollution in waterbodies is typically performed using microbial indicators such as *Enterococci*, *E. Coli*, and fecal coliforms. These bacteria, or classes of bacteria, are commonly found in the mammalian gastrointestinal tract which makes them excellent indictors of fecal waste. The US EPA recreational water quality criteria sets a threshold geometric mean concentration for *Enterococci* in fresh and marine waters of 35 colony forming units (cfu) per 100 mL, or *E.Coli* concentration of 126 cfu/100mL for fresh waters, for determination of impairment.[10] Once impairment is determined remediation requires the identification of microbial contamination source(s). Microbial source tracking (MST) utilizes a tool set of microbial and chemical tracers to locate and confirm a source of microbial contamination in the environment.[14] MST techniques are divided into two categories: 1) molecular and biochemical tracers, and 2) chemical tracers.

Molecular and Biochemical Tracers

Molecular and biochemical tracer techniques use genetic methods to identify fecal bacteria that are specific to a given host.[14] These techniques are divided into two categories: library-dependent and library-independent. Library-dependent techniques utilize fecal source fingerprinting, based upon the unique molecular and biochemical patterns of microbes isolated from potential sources. Molecular techniques such as ribotyping, are based on the determination of genetic sequences which uniquely identify a bacterium to its host. Identification requires a fecal source library with comparative information on host specific bacterial molecular and biochemical features. Fecal coliform, *E. Coli*, and *Enterococcus* species are commonly used for these techniques.[15] Ribotyping, which is one of the most commonly used methods, is based upon the detection of genetic difference within the 16S and 23S rRNA genes which are highly conserved. Other common library-dependent techniques used to identify specific DNA and RNA sequences are Pulsed Field Gel Electrophoresis and Repetitive Palindromic Polymerase Chain Reaction.[15-17] In contrast, biochemical techniques identify host specific organisms based on observable physical and biochemical characteristics. Such techniques include Antibiotic Resistance Analysis, Carbon Utilization profile, and Nutrient Utilization Pattern.[15] Library-dependent techniques are limited by the size and specificity of the bacterial libraries, which require a large number of bacterial isolates.[15,18] In addition to this, due to genetic changes over time, geography, and environment, libraries are most accurate when collected and applied to local areas.[19] This limits their application and comparability. Cost and time are also limiting factors as analysis procedures are extensive and require complex instrumentation.

Library-independent techniques do not require a fecal source library, instead relying upon host specific bacteria and viruses. These techniques require culturing or isolation of genetic material from specific microorganisms in environmental waters. One of the most promising host specific bacteria are of the genus Bacteroides, which are commonly found in the intestinal tract of mammals. There are current host-specific markers for human, ruminant, dog, pig, cattle, and some bird species, giving a wide availability of source determination. Other bacteria species used as human fecal markers are *Bididobacterium*, *Methanovibacter*, and *Rhodococcus coprophilus*. These techniques are also used to detect pathogenic protozoa such as *Cryptosporidium* and *Giardia*. Host specific viruses can also be used to discern between sources of fecal contamination. Human specific adenoviruses and enteroviruses as well as human polyomaviruses have been used to distinguish human fecal pollution in environmental waters.[15] Many of these detection methods for genetic identification do not discern between viable and non-viable cells, limiting their ability to indicate infectious risk. Viability can be determined by methods which target rRNA from live cells for amplification.[20,21] Methods are also limited by the presence of analytical inhibitors, such as metals and humic acid in water samples.[6] Cost and time are also limiting factors, although the reduction in PCR analysis equipment costs and published SOPs for human associated Bacteroides HF 183 and human polyomavirus makes these organisms promising tracers of human fecal contamination.[22]

Chemical Tracers

Chemical methods of MST rely on the detection of chemicals associated with a specific source of fecal pollution.[14] Chemical tracers could represent human and agricultural sources of fecal pollution. Four general classes of chemicals have been identified as promising tools for source tracking of fecal bacteria. 1) Fecal sterols/stanols 2) Pharmaceuticals and personal care product (PPCPs) 3) Optical Brighteners/fluorescent whitening agents (OBs) 4) Caffeine and miscellaneous analytes.[14,23]

Sterols are natural occurring steroids present in a wide variety of animals. When metabolized in the gut they form stanols which vary in concentration between animals due to their differences in diet and intestinal flora.[24] In humans the dominant stanol is the metabolite of cholesterol, 5β-stanol (coprostanol), which represents approximately 60% of stanols found in the gut.[25] The high ratio of coprostanol to other stanols found in human guts distinguishes their fecal material from that of other animals which have much lower levels.[26] Due to coprostanol specificity to humans and sensitivity of detection methods, stanols represent a promising method of tracking human fecal pollution. Some drawbacks of the use of stanols as chemical tracers are a lack of standardized analytical methods, and natural production of stanol analogs in the environment.[23]

PPCPs represent a broad class of compounds and other chemicals used by humans and found in wastewater. Common pharmaceuticals such as ibuprofen, acetaminophen, carbamazepine, and antibiotic agents and their metabolites are excreted in human feces and urine. The specificity of these chemicals to human wastewater and their broad use makes them ideal chemical tracers of human fecal pollution. Care must be taken to choose compounds with similar environmental fate and transport as bacterial contaminants. Though detection methods are accurate and sensitive, a lack of standardized methods makes comparing results difficult. In addition to this, dilute concentration in large bodies of water could make reliable detection difficult in areas of smaller populations where chemical use is less prevalent.[23] This could factor in to the ability to identify fecal contamination from failing on-site septic systems which serve single households.

Optical Brighteners are organic compound additives in laundry detergents and household products used to whiten clothing. The chemicals are easily detectable due to their absorption of ultraviolet light which is reemitted as blue light in the visible spectrum. OBs are indicators of sewage and septic system wastewater due to the mixing of gray and wastewater in household plumbing systems.[27,28] Due to the lack of natural sources of OBs their specificity as indicators of human wastewater is high. Analysis methods however can greatly effect this specificity due to other fluorescing compounds in the environment.[29] In addition, OBs can be difficult to detect in rivers and bays due to dilution, but show promise in smaller water bodies and for detection of sewer infrastructure failures.[29]

Caffeine and its metabolite paraxanthine represent another chemical tracer of human fecal contamination due to its prevalence of consumption and excretion in human waste. Caffeine which naturally occurs in a number of plant species, is an ingredient in coffee, tea, other drink products, and chocolates. While only 3% of consumed caffeine is excreted unmetabolized in urine, the average daily caffeine intake in the USA is 210 mg, therefore caffeine is detected in wastewater at µg/L concentrations.[30,31] While naturally occurring sources of caffeine exist, few to no plant species that produce caffeine are native to the USA. Caffeine's primary metabolite paraxanthine comprises approximately 84% of the breakdown products released following caffeine uptake. Paraxanthine has no known natural source, therefore its presence in the environment strongly indicates human wastewater contamination.[23]

Other metabolites and chemicals associated with human wastewater have been studied and proposed for use in MST. Several studies have examined the use of cotinine, a major metabolite of nicotine as a chemical tracer of human fecal contamination.[32,33] Although these studies detected cotinine in wastewater samples and environmental sites suspected to be contaminated by wastewater, further evaluation of cotinine's environmental fate and correlation to fecal bacteria need to be performed. Artificial sweeteners, such as sucralose and acesulfame, represent another group of chemical tracers. Due to their ubiquitous use in food products and poor metabolism in humans, artificial sweeteners have potential to be chemical tracers of wastewater. Several studies have detected sucralose and acesulfame in environmental samples associated with human wastewater contamination.[33,34] However, artificial sweeteners may persist longer than pathogens in the environment which can limit their use in indicating the presence of wastewater contamination.

Determination of Novel Chemical Tracers

Detection of novel chemical tracers of human fecal contamination relies on the ability to analyze environmental water samples for their chemical constituents in an untargeted approach.

Previously identified chemical tracers, described above, use analytical methods which limit detection to a predetermined list of target compounds, which when present indicate the presence of wastewater. These targeted methods are limited to detecting chemicals which are known to be present in wastewater and environmental waters contaminated by wastewater. In contrast, novel chemical tracer identification occurs by using analytical methods which capture vast amounts of data on chemical constituents through non-targeted approaches. This approach utilizes high-resolution mass spectrometry (HRMS) to detect thousands of organic chemical features in an environmental sample based on their accurate mass.[35] Accurate mass represents the mass of a chemical ion which has been experimentally measured to a high degree of accuracy and precision.[36] HRMS instruments acquire hundreds to thousands of accurate mass measurements in full scan acquisition mode while providing high mass accuracy.[35] Each accurate mass measurement represents a chemical feature in the environmental sample.

The non-target identification procedure begins with automated peak detection of accurate mass features detected in the sample. Software employing a recursive algorithm groups peaks together based on a specific mass to charge (m/z) range and retention time window. Elemental formulas are then assigned to features of interest based upon their accurate mass and isotopic patterns. Plausible chemical matches are then determined by searching the elemental formulas in chemical databases.[37] Database searches consider accurate mass and isotope pattern to score possible matches. Further confidence in correct identification is performed by reanalyzing samples to collect MS/MS spectra which are matched to spectral libraries.[37] Final confirmation can be made using reference standards if available.

HRMS non-target techniques have recently been employed in environmental settings for the detection and identification of unknown chemical contaminants. A 2015 study in the Rhine River (central Europe) reported identifying two compounds, the muscle relaxant Tizanidine and 1,3-dimethyl-2-imidazolidinone, an industrial solvent, which had never before been detected in the river. Through spatial sampling of the river’s tributaries and main stream, the point source of the Tizanidine was determined to be within a 60km stretch of the river.[38] Non-target approaches have also been utilized to compare receiving waters of wastewater sources. McEachran et al. reported that overall chemical compositions of samples from two sources of wastewater and their receiving waters were distinct from one another.[39] This demonstrates the ability of these techniques to distinguish between sources based on their chemical composition providing the ability to uniquely identify and track sources within a waterbody. Similar methods have been used to determine possible aquatic toxicants in highway runoff. Du et al. reported detecting chemicals associated with stormwater in water samples and tissue samples of coho salmon (*Oncorhynchus kisutch*) exposed to highway runoff. The chemical acetanilide, used in rubber vulcanization, was identified in the gill and liver samples of exposed fish and may cause possible adverse toxicological effects.[40] Further research within this field identified several chemical groups associated with tire wear particles in water samples lethal to coho salmon.[41] These studies demonstrate the ability of HRMS non-target approaches to identify unknown chemical signatures in the environment, and track them spatially within water bodies and ecological systems allowing for source tracking of contaminants.

Linking Human Microbial Contamination and Novel Chemical Tracers

The demonstrated ability of HRMS non-target techniques to identify unknown chemicals in environmental waters can be utilized to detect novel chemical tracers of microbial contamination. By identifying environmental sites with known human-sourced microbial contamination, water samples can be collected and analyzed for possible chemical tracers. This method is reliant on the identification of environmental sites with known wastewater contamination, which can subsequently be compared to sites with no contamination. While microbial source classification can be performed with microbial or chemical indicators described above, an alternative method of identifying wastewater contamination at environmental sites is the use of scent tracking canines who have been trained to indicate the presence of wastewater.

Scent tracking canines have been utilized in other demanding environments to detect illegal substances such as explosives and narcotics. With proper training canines can detect a wide range of chemical odorants at part-per-billion (ppb) and part-per-trillion (ppt) concentrations.[42] These scent tracking methods have been expanded to include the detection of wastewater and detergents to assist in the source tracking of illicit discharges. A 2014 study examined the use of scent tracking canines to detect human wastewater in storm drains in Santa Barbara, CA. The response of two scent tracking canines was compared at 24 sites to several human specific microbial tracers and two chemical indicators of wastewater. The dogs indicated 70% and 100% positively at sites confirmed to be contaminated by wastewater by one or more of the assays. Sites negatively identified by both dogs did not contain human waste markers.[43] To the authors knowledge this is the only comparative study of scent tracking canines and wastewater markers, although several public health and water districts have utilized the capabilities of scent tracking canines in source tracking programs. One such program, in Skagit County, WA, performed scent testing in a controlled setting, and at field sites of environmental waters suspected to be contaminated with human sewage. The canine positively indicated the presence of sewage in quality assurance samples collected from a septic system and did not indicate on control samples of distilled water, horse and cow manure in distilled water, and UV treated septic effluent.[44] Although the use of scent tracking canines in microbial source tracking is relatively new and understudied, it shows promise as a tool to identify and track discharges of wastewater in surface waters.

Study Objective and Significance

This study aims to use HRMS non-target analysis to identify novel chemical tracers of wastewater in water samples obtained from sites which have been indicated to contain human wastewater by scent tracking canines. By utilizing canines trained to identify human wastewater, surface water sites can be classified as positive or negative in regard to their presence of human wastewater contamination. Water samples from canine positive and negative sites are analyzed using HRMS non-target techniques to identify possible chemical tracers associated with the presence of wastewater in canine positive sites. The objective of this work is to identify novel chemical tracers of microbial contamination which could be utilized in MST efforts, furthering the ability of public health professional to identify and remediate impaired waters.

Methods

Study Setting

The study was performed in Skagit County, Washington USA, which is located approximately 50 miles north of Seattle, WA. Skagit County is 1,920 square miles and has an approximate population of 125,600.[45] Within Skagit County the Samish River Watershed drains an area of 139 square miles from the foothills of the Cascade mountain range to Samish Bay. The Samish River and Bay have a historical trend of high fecal bacteria which have impacted water quality and limited use of productive shellfish beds in the Bay. Several areas within the watershed have been listed as impaired due to fecal bacteria.[46]

Site Locations

Sampling sites were selected from previously sampled locations along the Samish watershed. A site map is shown in Figure 1 with sampling locations' geographical coordinates listed in Table 1

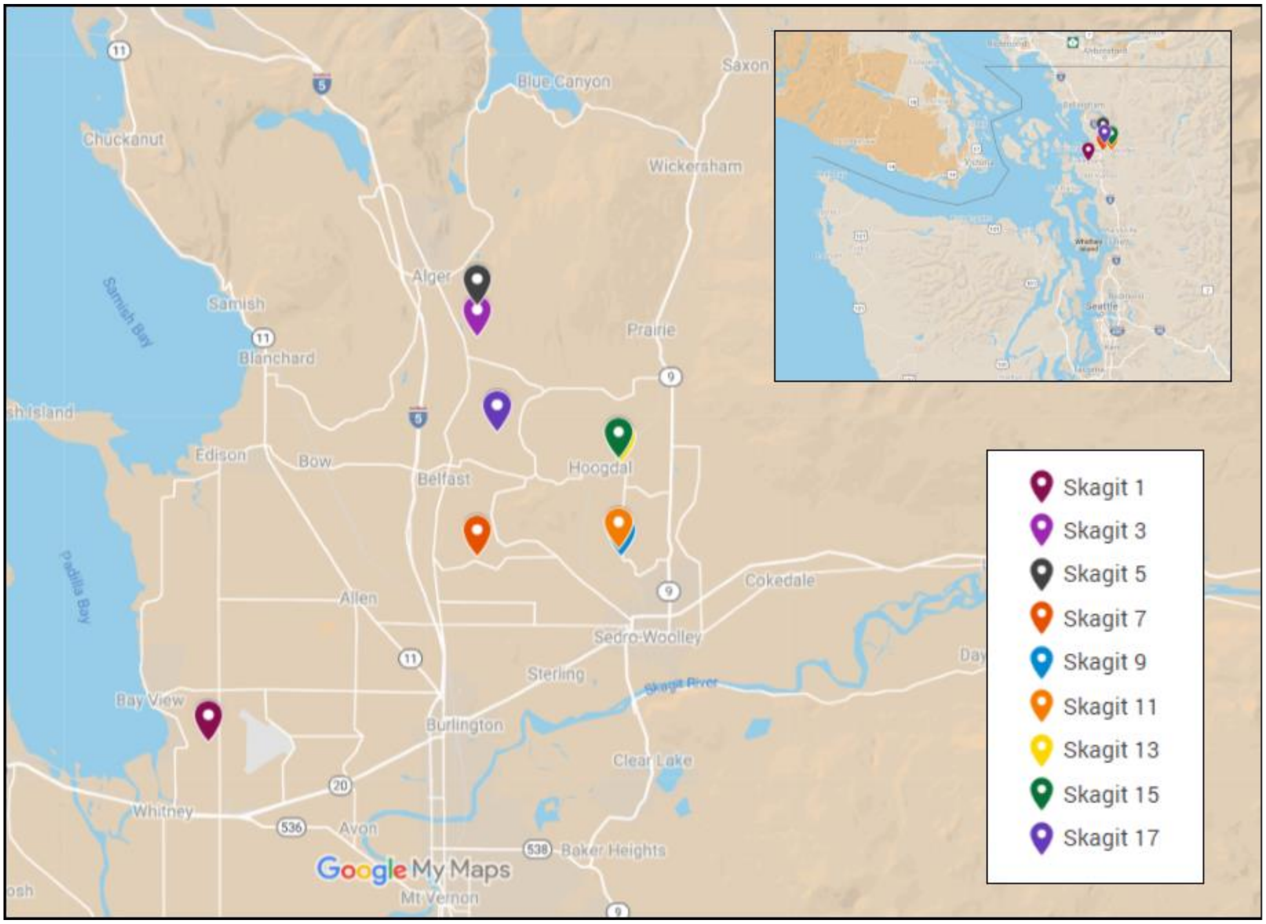

Figure 1 Site Map of water sampling sites within the Samish Watershed.

Table 1 University of Washington Sample ID, Skagit County PIC site name, geographical location for study sampling sites, and canine classification of wastewater contamination

Sample ID	site name	Latitude	Longitude	Canine
Skagit 1	DYKBV	48.47	-122.45	Positive
Skagit 3	BUT-BCR1	48.6	-122.32	Negative
Skagit 5	BUT-BCR8	48.61	-122.32	Negative
Skagit 7	BCCR	48.53	-122.32	Positive
Skagit 9	WILL	48.53	-122.25	Negative
Skagit 11	CULLDITCH	48.53	-122.25	Positive
Skagit 13	SWEDETRIBN	48.56	-122.25	Positive
Skagit 15	SWEDETRIBS	48.56	-122.25	Negative
Skagit 17	SKAR	48.57	-122.31	Positive

Canine Classification and Water Sampling

Water samples from surface water sites were collected by Skagit County Public Works staff and brought to a neutral location (parking lot of public works building) for classification by a scent tracking canine. Water samples were placed in sterile open topped containers and spaced amongst the parking lot to prevent cross contamination of scent. The canine was presented to each container where it provided a positive or negative indication to the presence of human wastewater. The canine's indications were interpreted by the canine handler and recorded by Skagit County staff. Positive and negative control samples consisting of distilled water (DW) (negative control), cow/horse manure diluted in DW (negative control), and sewage diluted in DW (positive control) were used to verify canine indications. After classification, five canine positive sites and four canine negative sites were chosen for resampling for chemical analysis (Table 1). The following day 8L of water were collected from each site in a solvent rinsed amber glass bottle. Water samples for fecal coliform analysis were also collected at this time. Water samples for chemical analysis were transported on ice to the University of Washington Center for Urban Waters Laboratory in Tacoma, WA. Samples were stored a 4°C until being processed within 48hrs.

Water Sample Preparation

Water samples from each site were split into triplicate 1.5L samples. Sample processing followed standardized methods outlined in detail in Du et al. and Peter et al..[40,41] Briefly, water samples were processed by solid phase extraction using Infinity SPE cartridges (100mg 3mL, ABS Materials, Wooster, OH, USA) preconditioned with 50/50 (v/v) methanol/deionized water (3mL) and deionized water (25mL). Water samples (1.5L) were loaded on to cartridges without prefiltration. Following loading, cartridges were rinsed with deionized water (10mL), dried with nitrogen (15min), and eluted with methanol (2 x 2.5mL). Eluent was concentrated to 1mL with nitrogen. Internal Standards consisting of seven isotopically labeled compounds were added to final concentrated samples before analysis. Method blanks were prepared following the same methods as field samples but used deionized water.

High Resolution Mass spectrometry Analysis

HRMS analysis was performed using an Agilent 1290 UHPLC (Santa Clara, CA, USA) for chromatographic separation and Agilent 6530 Quadrupole Time-of-Flight (QTOF) for detection. Compound separation was obtained using a reverse phase C18 column (Agilent ZORBAX Eclipse Plus 2.1 x 100mm, 1.8 µm particle size) and 5mM ammonium acetate plus 0.1% acetic acid in each methanol and water. Flow, mobile phase gradient, and source conditions are described in detail in Peter et al.[41] HRMS spectra were detected in ESI positive and ESI negative across 100-1700 m/z range for MS and 50-1700 m/z for MS/MS in 2 GHz Extended Dynamic Range mode.

Data Analysis

(data analysis method is outlined in figure 5 in the appendix)

Initial HRMS data extraction was performed using Agilent MassHunter Profinder (B.08.00) software to extract, group, and align features within field replicates (figure 5 step 1). Profinder employs a recursive algorithm to isolate peaks within replicates and assign unique exact mass-retention time pairs, designating a feature. Features were extracted following methods described in detail in Du et al. for ESI positive and negative data.[40] Briefly, features with a peak height count greater then 5000 were extracted as positive adducts (M+H, M+Na, and $M+NH_4$) or negative adducts (M-H and $M+CH_3OO$). Replicate feature alignment used mass/charge (m/z +/- 20 ppm) and retention time (+/- 0.3 min) windows to identify and group features.

Following feature extraction subsequent data analysis was performed using Agilent Mass Profiler Professional (MPP) software. MPP is a statistical software used to filter HRMS data and explore relationships amongst study conditions. In this study MPP was used to filter HRMS data to remove features of non-interest and determine features present in sites indicated to contain human wastewater by the scent tracking canine (canine positive sites).

Initial filtration of replicate samples was performed to remove features not present in 100% of replicates for each sample site (figure 5 step 2). Features which did not meet this criterion were excluded from further analysis. The remaining features for each site were then compared to method blanks to remove features present due to sample preparation (figure 5 step 3). Features in field sample replicates and method blanks were compared using a moderated T-test (the default option in the MPP software). Statistically significant features (P-value > 0.1) present in field sites at two-fold greater peak abundance than the average method blanks were retained. Features present in canine negative sites after blank subtraction were consolidated into a single list representing features present in any canine negative site, here after referred to as "canine negative site features." (figure 5 step 4).

Canine positive site features were identified with two approaches: Venn diagrams and Fold change. In the first approach, features were identified by comparing each canine positive site's features to canine negative site features with the use of Venn diagrams (figure 5 method 1). Within the Venn diagram, features present only in the canine positive sites were retained. Unique features for each canine positive site were then compared between sites by Venn diagram to determine features present in two of five, and three of five canine positive sites. The selected detection frequency of features of interest found in canine positive sites were based on studies proposing the use of pharmaceuticals and personal care products as tracers of wastewater whose detection frequencies range from 50% to 80%.[32,47]

Further analysis of canine positive site features was performed by fold change analysis to identify features present at a greater peak abundance in canine positive sites in comparison to canine negative sites (figure 5 method 2). For each canine positive site, features present at two-fold greater peak abundance than in canine negative sites were retained. The two-fold greater peak abundance threshold was chosen based on similar methodology employed by Peter et al. to distinguish between features of interest found in environmental sites and blanks at a five-fold peak abundance threhold.[41] A lower fold change threshold value was used for this study to increase the number of features identified as possible tracers of wastewater. These fold change features were then compared between the canine positive sites by Venn diagram to identify features present in two of five, and three of five positive sites.

Suspect screening of positive sites' unique and fold change features of interest were performed in Agilent MassHunter ID Browser (B.07.00) to identify possible chemical matches (figure 5 step 5). Feature screening was performed using in-house and commercial (Norman MassBank[48] and Agilent Technologies MassHunter Forensics & Toxicology PCDL) databases containing approximately 48,000 chemicals. Feature accurate masses, isotopic spacing, and abundance were used to calculate chemical formulas which were matched to possible chemical(s) within databases. When no database match was made, formula generation was utilized and limited to carbon, hydrogen, oxygen, nitrogen, sulfur, and chlorine. The chemical adducts which were allowed were M+H, M+Na, and $M+NH_4$ for positive ions, and M-H and $M+CH_3OO$ for negative ions. Chemical matches and formula were scored on the bases of their exact mass match, and isotope spacing and abundance. Chemical matches which scored 80 or greater and had multiple database hits were manually examined to identify the most correct match based on available predicted adduct matches and isotope abundance. Formulas generated for chemicals without a database match which scored 90 or greater were similarly examined to confirm correct formula assignment. Confirmed identifications were compiled to eliminant duplicates.

Further confirmation of chemical matches were performed by MS/MS analysis (figure 5 step 6). Chemicals identified by suspect screening were searched in the Metlin (www. metlin.scripps.edu) and mzCloud (www.mzcloud.org) online spectra libraries by chemical formula to identify possible MS/MS spectra. Spectral library searches were focused on identified chemical formulas which contain elements beyond carbon, hydrogen, and oxygen. Priority was placed on these chemicals as their chemical formulas have fewer possible chemical structures, limiting their possible identities. Chemicals with available experimental or *in silico* spectra were prioritized for MS/MS analysis. After MS/MS analysis, spectra data for target chemicals were compared to the available library spectra for confirmation. Spectra were matched based on the presence and abundance ratio of precursor and product ions. Precursor and product ion accurate mass were deemed a match for masses within 5ppm of those reported in library spectra. At least two product ions were used to match spectra where available. Ratios of product ions were confirmed if matching collision energy spectra were available. If matching collision energy spectra were not available, product ion confirmation was made on the basis of mass accuracy alone.

Confidence in chemical identification was assigned following the criteria outlined by Schymanski et al.[49] The highest confidence match, confirmed structure (S1), is assigned to chemicals whose retention time and MS/MS spectra were matched to a reference standard. No compounds detected in this study were confirmed at a S1 confidence. A probable structure match is achieved when chemical MS/MS spectra is matched to library experimental (S2A) or *in silico* spectra (S2B). If spectrum data is unavailable, fragment elucidation can be used to predict formula at lower confidence (S3). Confidence of chemical match is further reduced for matches based on molecular formula only (S4), or exact mass only (S5).

Results

After alignment and initial replicate filtration 1,958 features were found across the nine sample locations (figure 2). 1,131 of these features were present in field sites at a two-fold greater peak abundance than in method blanks. Of these features, 139 were found in two of five canine positive sites at a two-fold greater peak abundance then in canine negative sites. 96 of these features were unique to canine positive sites. 15 features were found in three of five canine positive sites at a two-fold greater peak abundance then canine negative sites. Eight of these features were unique to canine positive sites.

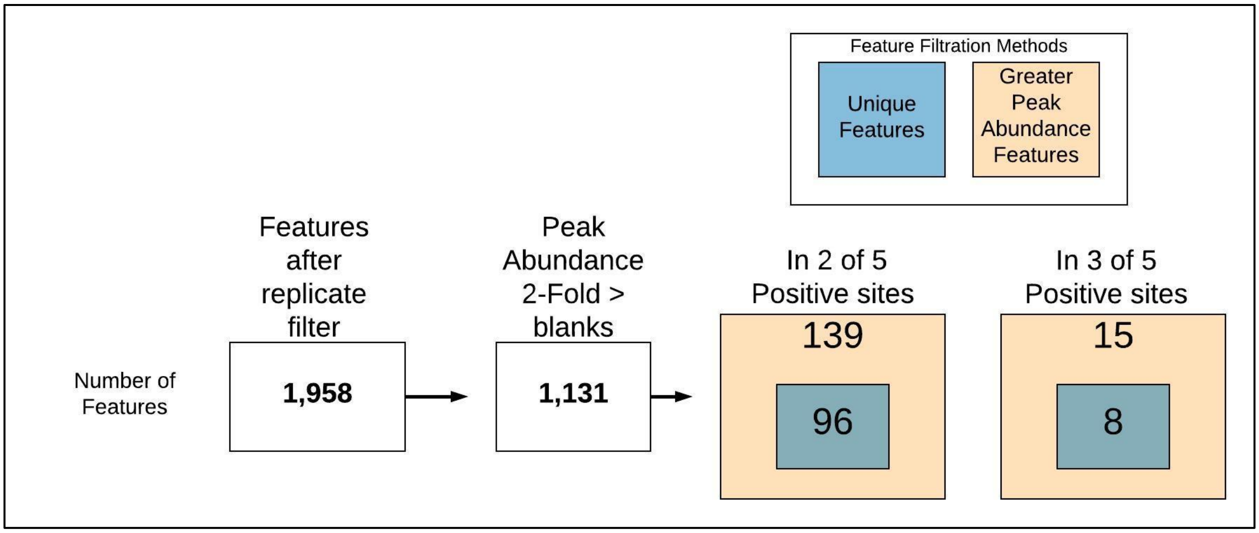

Figure 2 Number of Features detected in each step of feature reduction methods.

Fecal coliform concentrations for field sites are given in Table 2. Canine negative sites had a mean fecal coliform concentration of 20.2 cfu/100mL with a standard deviation of 18.8 cfu/100mL. Skagit 1, a canine positive site, had the highest concentration of fecal coliform, 350 cfu/100mL. The only other canine positive site with fecal coliform concentrations greater than the mean concentration in canine negative sites was Skagit 13 (79 cfu/100mL). Canine positive sites, Skagit 7, 11, and 17 had concentrations less than the mean concentration at canine negative sites.

Table 2 Fecal coliform Concentrations in Study Sites

Sample ID	site name	Canine	Fecal Coliform (cfu/100mL)
Skagit 1	DYKBV	Positive	350
Skagit 3	BUT-BCR1	Negative	17
Skagit 5	BUT-BCR8	Negative	0.9
Skagit 7	BCCR	Positive	7.8
Skagit 9	WILL	Negative	17
Skagit 11	CULLDITCH	Positive	7.8
Skagit 13	SWEDETRIBN	Positive	79
Skagit 15	SWEDETRIBS	Neagtive	46
Skagit 17	SKAR	Positive	23

The analytical results of the sites were compared by hierarchical cluster analysis (HCA) for ESI positive and ESI negative features (Figures 3 and 4). Sites were clustered based on their features' presence and abundance. The HCA for ESI positive divided the sample sites into 4 main clusters with Skagit 1, 5, and 13 each clustering on their own while the remaining sites formed several subclusters (Figure 2). Skagit 1 and 13 demonstrated a similar cluster pattern for ESI positive and negative features and shared the closest branch in comparison to other sites (figure 3 and 4). Within the subclusters, both canine positive and negative sites clustered together without discernable chemical pattern. Skagit 5 for ESI positive features and Skagit 11 for ESI negative features neither resembled Skagit 1 and 13 nor the other subclustered sites.

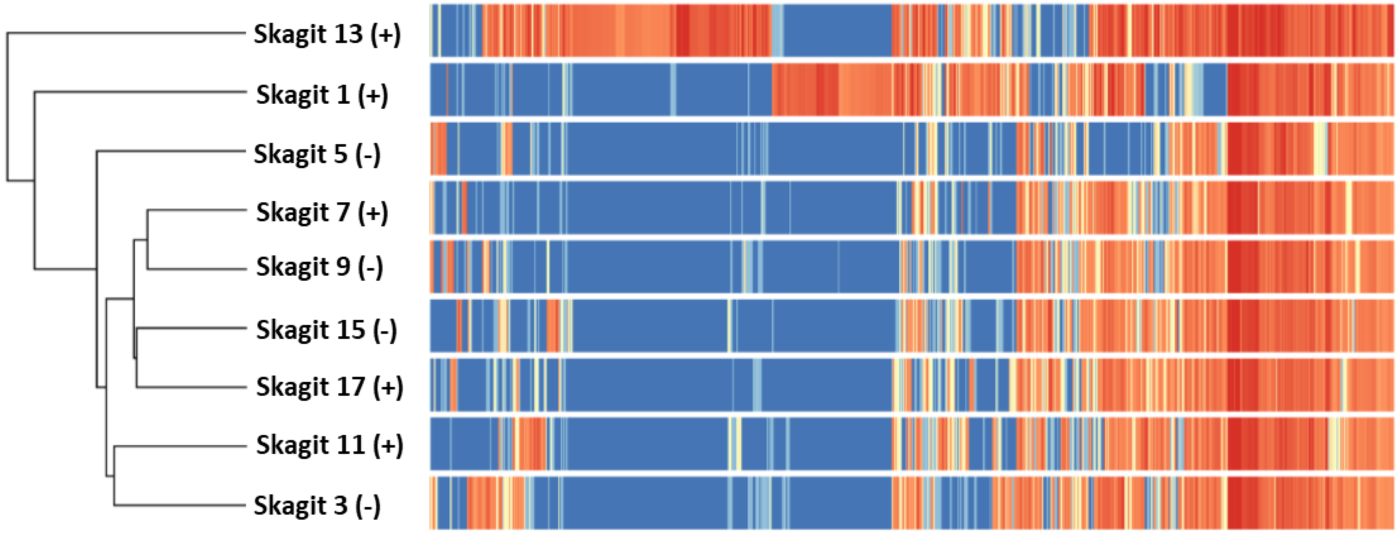

Figure 3 Hierarchical Cluster Diagram for ESI Positive features

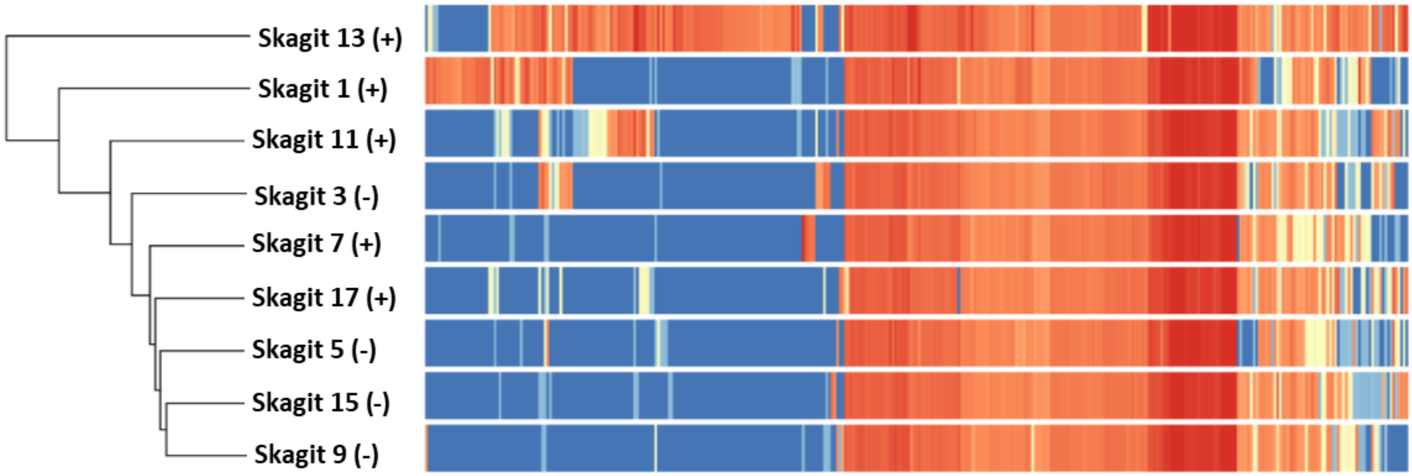

Figure 4 Hierarchical Cluster Diagram for ESI Negative features

Chemical identities, chemical formulae, adduct, nominal mass, retention time, identification method, and confidence level are reported in Table 3 in the appendix. Of the 139 features of interest identified in positive sites, chemical formulas were assigned to 76, a S4 confidence level identification. 33% of these chemicals, consisted of only carbon, hydrogen, and oxygen, while the remaining 67% also contained phosphorus, nitrogen, sulfur, or chlorine. 62 chemicals were identified by accurate mass only (S5 confidence level). One unique feature found in three of five positive sites was identified at a S2A confidence level as 2-mercaptobenzothiazole.

Eight compounds were prioritized for MS/MS analysis due to available library spectra. Of these compounds five were triggered upon during MS/MS analysis. Four of these compounds' measured MS/MS spectra did not match available library spectra. 2-Mercaptobenzothiazole (MBT) identity was confirmed by MS/MS analysis. The MBT experimental spectra was obtained from the mzcloud spectral library and compared to MS/MS spectra data for m/z 165.979 from site Skagit 13. Comparisons were made for precursor ion (m/z 165.979) and product ions (m/z 134.007 and 57.975) at a collision energy of 20v. Precursor and product ion masses from Skagit 13 were within 5 ppm of reported values. Ion ratios were not used for matching as collision energy for library spectra was unavailable. Experimental and environmental spectra can be seen in Figure 5 in the appendix.

Discussion

The objective of this study was to identify chemicals which could be utilized in MST efforts by comparing HRMS data from canine positive and negative sites. The performance of a chemical tracer is based, in part, on its source specificity.[23] The chemical should be specific to sites where contamination occurs and detected in high frequency at these sites.[23] This study's feature reduction methodology provides four stringency levels of selection for chemicals of interest as tracers of wastewater. Selection stringency is based on the chemical's specificity and detection frequency at canine positive sites. Features' unique to canine positive sites demonstrate greater specificity than two-fold greater peak abundance features which were detected in both canine positive and negative sites. Detection frequency of features is measured by their detection in two of five (40%), or three of five (60%) canine positive sites. Unique features in three of five canine positive sites are of highest interest as tracers because of their specificity and higher detection frequency at sites indicated to contain wastewater. Eight chemicals met these criteria, and the identity of one was confirmed as 2-mercaptobenzothiazole (S2A). The other seven chemical identities ($C_{53}H_{79}N_8O_2$, m/z 424.9869, m/z 507.2642, m/z 540.4460, m/z 806.5091, m/z 806.5091, m/z 904.6040) could not be confirmed beyond chemical formula or accurate mass (S4 and S5). The 60% detection frequency of these compounds is comparable to study results proposing the use of five compounds (benzophenone, ethyl citrate, galazolide, tributyl phosphate, and triclosan) as tracers of wastewater which found detection frequencies of 50% to 70%.[32] While the eight chemicals detected by this method were found in 60% of sites, they were more importantly all detected in sites Skagit 1 and 13 which demonstrated the most evidence of wastewater contamination.

An additional 88 chemicals were uniquely identified in two of five positive sites. The lower detection frequency of 40% greatly increased the number of possible tracers. While these chemicals were detected at a lower frequency than other proposed chemical tracers, fecal coliform concentration and HCA suggests chemicals found in Skagit 1 and 13 are most clearly related to wastewater contamination. For this reason, the chemicals found in Skagit 1 and 13 are an additional priority for further investigation as chemical tracers of wastewater. It is important to note that though all but one chemicals' identity was not confirmed beyond chemical formula or exact mass, these chemicals are still usable tracers as they are distinguishable in the environment by these identities.

Chemicals identified by the lower stringency method of two-fold greater peak abundance represent a wider group whose abundance in the environment may be altered by the presence of wastewater. This method netted 31% additional chemical tracers than those present only in canine positive sites. Chemical identities were confirmed at S4 and S5 confidence levels in two of five and three of five positive sites. As these compounds are found in both canine positive and negative sites further characterization of background concentrations needs to be performed to determine if the elevated concentrations found in canine positive sites represents wastewater contamination.

Of the compounds identified for further investigation, chemicals containing more than carbon, hydrogen, and oxygen were prioritized as their chemical formulas provide fewer molecular options. Eight chemicals with available spectra were prioritized for MS/MS analysis. Of these, one chemical confirmation was made as MBT. MBT is from the benzothiazoles (BTs) chemical class, which are high production chemicals used in a multitude of natural and synthetic rubber products. Due to their use in the production of car tires BTs commonly enter surface waters from storm water containing tire abrasion particles.[50] Municipal and industrial wastewater streams are also sources of BTs entering surface waters.[51-53] MBT is a production intermediate and hydrolysis product of BTs and can co-occur with other BTs in waste streams.[54] BTs have been detected in house hold wastewater at 50% to 80% of the concentration found in municipal wastewater.[54] BTs use in household consumer products has not been specifically determined, however their wide spread use in the production of rubber and plastics suggests their inclusion in various home goods.[54,55] While MBT has not specifically been detected in household waste streams, BTs significant contribution to municipal wastewater concentrations suggests their emission from household products is a source into the environment.[54] BTs from storm water, municipal wastewater, and household wastewater show reportedly different compositions, suggesting sources could be distinguished in the environment.[54] The ability to distinguish between sources is important for the use of BTs and MBT as chemical tracers, this should be further investigated.

The HCA for ESI positive and negative features show a lack of distinct grouping for the canine positive and negative sites. However, some canine positive sites demonstrate unique clustering which distinguishes them from other sites. Sites Skagit 1 and 13, each cluster separately, demonstrating their unique chemical signatures which discern them from other sites (Figure 3 and 4). The unique chemical clustering of Skagit 1 and 13 may be indicative of wastewater contamination which is supported by elevated fecal coliform concentrations at these sites. The remaining canine positive sites do not have discernable patterns of clustering which distinguishes them from canine negative sites. In addition, bacterial contamination is absent at these sites in concentrations greater then canine negative sites.

The scent tracking canine's indication of wastewater contamination was not supported by fecal coliform and chemical clustering across all canine positive sites. Canine positive sites, Skagit 1 and 13, had elevated fecal coliform concentrations and demonstrated unique chemical clustering. This chemical and biological evidence that distinguishes these sites from canine negative sites suggests possible wastewater contamination. Sites, Skagit 7, 11 and 17, did not have elevated fecal coliform levels and limited to no distinguishable unique chemical signature. Taking into account this biological and chemical evidence, three of five sites indicated to contain wastewater by the canine did not show a pattern which differentiates them from canine negative sites. Other studies' results utilizing scent tracking canines show varying levels of agreement between canine indication and fecal bacteria levels. One study found sites indicated by canines to contain wastewater contamination had high levels of fecal indicator bacteria. [43] In contrast to that study, past field studies within Skagit County utilizing scent tracking canines have shown low fecal coliform concentrations at sites indicated to contain wastewater, even when nearby septic failure was confirmed by dye testing.[56] Furthermore, positive and negative control samples testing the canine's ability to discern human wastewater from cow/horse manure, all were correctly classified. These results suggest that the chemical scent canines identify as wastewater behaves differently in the environment compared to the fecal coliform and chemicals detected in this study. Though the chemical scent is unknown, volatile organic chemicals, which are typically detected by canines, may not be amenable to the liquid chromatography mass spectrometry (LC/MS) analysis used in this study. Therefore the chemicals which make up the scent canines are responding to in wastewater contaminated samples may not be detected using our analysis approach.[42]

Study Limitations

The methods used in this study to identify chemical tracers may be limited by the lack of bacterial and chemical indicators of wastewater at all canine positive sites. The highest priority for chemical identification was placed on chemicals present in three of five canine positive sites. While fecal coliform concentrations and chemical clustering only provided evidence of the presence of wastewater at Skagit 1 and 13, the fact that all chemicals which met these criteria were found in these two sites provides sufficient justification for prioritizing chemicals found in three of five canine positive sites. The approach to identify tracers may have been limited by water sample collection a day after canine screening took place. Though no rain event was recorded between screening and the resampling, and samples were collected at the same time of day which may mitigate variance associated with daily patterns, the chemical composition and bacterial concentration could change between samplings which may limit this study's ability to detect common chemical signatures at canine positive sites. The sample prep (Infinity SPE cartridges) and analysis method (LC/MS) are inherently selective for specific chemical properties (hydrophobic, semi-volatile, non-volatile, ionizable) because of this, other types of viable chemical tracers of wastewater would not be detected by this study's approach. The study was also limited by the small geographic region and limited number of sampling stations used to determine possible chemical tracers. Chemical tracers identified by this study may be limited to this region or specifically to wastewater contamination in rural areas.

Future Steps

The chemicals identified in this study should be further examined in field programs to confirm their use as indicators of wastewater contamination in surface waters. Highest prioritization should focus on those uniquely detected in canine positive sites. Further confirmation of chemical applicability should take place at sites of known and unknown contamination as well as in wastewater. Additionally, MBT and the composition of BTs in surface waters should be explored as tracer of wastewater. Care should be taken to distinguish sources of MBT and BTs in surface water to determine that their presence is due to wastewater not stormwater.

Conclusions

This study demonstrated the novel use of HRMS and non-target techniques to identify chemical tracers of wastewater contamination in surface waters. A group of eight compounds uniquely found in three of five canine positive sites were identified as highest potential chemical tracers of wastewater. Chemicals were identified at varying confidence levels (accurate mass, chemical formula, chemical identity). 2-Mercaptobenzothiazole was identified at a high confidence (S2A) and has been shown to be present in wastewater effluent. Further field investigation should be performed to confirm the chemical's presence in wastewater and wastewater impacted surface waters.

Of the nine sites investigated in this study, two sites, Skagit 1 and 13, had both canine and bacterial evidence consistent with wastewater contamination. Three other canine positive sites did not show evidence of wastewater contamination. The low concurrence of canine indication and bacterial evidence may in part be due to sampling conditions and analysis techniques, but this inconclusive evidence limited the ability of this study to identify chemicals present in wastewater contaminated surface waters. The use of scent tracking canines to classify sites of interest for HRMS analysis is novel, but classic approaches of human specific bacterial indicators are better suited to classify sites for the detection and identification of novel chemical tracers of wastewater.

Appendix

Table 3 Compound name, chemical formula, detection confidence, retention time (RT), adduct, neutral mass, and detection method for chemicals identified in 2 of 5 and 3 of 5 positive sites.

The methods section describes the identification method and number of sites the chemical was detected in. Chemicals only found in either two or three of five sites are labeled as Unique 2/5 or 3/5. Chemicals detected at a two-fold greater peak abundance in two or three of 5 positive sites in comparison to negative sites area labeled Fold Change 2/5 or 3/5. Compounds detected in both Skagit 1 and 13 are indicated as such.

Compound	Formula	Detection Confidence	RT (min)	Adduct	Neutral Mass	Method	Present in Skagit 1 & 13
2-Mercaptobenzothiazole	C7 H5 N S2	S2A	5.698	[M-H]	166.9858	Unique 3/5	x
m/z 101.0084		S5	0.609	[M+Na]	78.0212	Unique 3/5	x
m/z 126.0224		S5	0.606	[M+H]	125.0152	Unique 3/5	x
m/z 310.1770		S5	5.102	[M+NH4]	146.0736	Unique 3/5	x
m/z 184.0471		S5	0.745	[M+H]	183.038	Unique 3/5	x
m/z 211.1356		S5	4.575	[M+H]	193.1018	Unique 3/5	x
m/z 197.1529	C12 H20 O2	S4	7.961	[M+H]	196.1459	Unique 3/5	x
m/z 227.0085		S5	0.607	[M+H]	226.0012	Unique 3/5	x
m/z 227.1641		S5	9.14	[M-H]	228.1714	Unique 2/5	
m/z 239.1988	C10 H26 N2 O4	S4	11.337	[M+H]	238.1913	Unique 2/5	
m/z 242.1171	C13 H21 S2	S4	0.667	[M+H]	241.109	Unique 2/5	
m/z 252.1936		S5	4.758	[M+H]	251.1857	Unique 2/5	x
m/z 251.2008		S5	14.611	[M-H]	252.2082	Unique 2/5	x
m/z 253.2168	C16 H30 O2	S4	15.424	[M-H]	254.224	Unique 2/5	x
m/z 255.1592	C12 H22 N3 O3	S4	8.21	[M-H]	256.1666	Unique 2/5	x
m/z 265.1296	C12 H26 O2 S2	S4	10.7	[M-H]	266.1369	Unique 2/5	x
m/z 265.1797		S5	11.936	[M-H]	266.187	Unique 2/5	
m/z 272.9378		S5	0.623	[M+H]	271.931	Unique 2/5	x
m/z 275.2007		S5	14.521	[M-H]	276.208	Unique 2/5	x
m/z 277.2163	C16 H28 N3 O	S4	15.133	[M-H]	278.2237	Unique 2/5	x
m/z 282.0139	C19 H5 O S	S4	0.609	[M+H]	281.0069	Unique 2/5	x
m/z 283.0602	C14 H10 N3 O4	S4	9.046	[M-H]	284.0674	Unique 2/5	
m/z 285.0762		S5	9.045	[M+H]	284.069	Unique 2/5	
m/z 285.2266		S5	3.387	[M+H]	284.219	Unique 2/5	x
m/z 289.2180	C12 H30 N6 S	S4	14.972	[M-H]	290.2252	Unique 2/5	x
m/z 295.2263	C16 H30 N3 O2	S4	12.999	[M-H]	296.2336	Unique 2/5	x
m/z 355.2851		S5	16.812	[M+CH3COO]	296.2704	Unique 2/5	x
m/z 307.0584	C14 H8 N6 O3	S4	7.188	[M-H]	308.0656	Unique 2/5	x
m/z 315.0867	C17 H14 O6	S4	9.195	[M+H]	314.0794	Unique 2/5	
m/z 327.2338	C22 H32 O2	S4	15.625	[M-H]	328.2413	Unique 2/5	x
m/z 401.1288	C10 H20 N3 O10	S4	0.682	[M+CH3COO]	342.115	Unique 2/5	x

m/z 365.1050	C12 H22 O11	S4	0.681	[M+Na]	342.1159	Unique 2/5	x
m/z 345.1371	C16 H26 O6 S	S4	9.38	[M-H]	346.1446	Unique 2/5	
m/z 198.4836		S5	0.612	[M+Na]	350.9892	Unique 2/5	x
m/z 351.2506	C17 H32 N6 O2	S4	16.561	[M-H]	352.258	Unique 2/5	x
m/z 358.9618		S5	0.596	[M+H]	355.9745	Unique 2/5	
m/z 357.2792		S5	13.212	[M+H]	356.2686	Unique 2/5	x
m/z 381.2982		S5	16.579	[M+Na]	358.3087	Unique 2/5	x
m/z 366.3731		S5	17.528	[M+H]	365.3585	Unique 2/5	x
m/z 373.0915	C17 H16 N3 O7	S4	13.86	[M-H]	374.0988	Unique 2/5	
m/z 375.2362	C15 H32 N6 O5	S4	12.332	[M-H]	376.2432	Unique 2/5	x
m/z 385.3781	C21 H46 N5 O	S4	16.038	[M+H]	384.3697	Unique 2/5	x
m/z 399.3598	C26 H44 N3	S4	18.898	[M+H]	398.3523	Unique 2/5	
m/z 414.3584	C24 H47 N O4	S4	15.05	[M+H]	413.3511	Unique 2/5	
m/z 421.3289	C22 H42 N6 O2	S4	18.237	[M-H]	422.3366	Unique 2/5	
m/z 424.9869		S5	0.606	[M+H]	423.9919	Unique 2/5	x
m/a 423.4196	C26 H54 N3 O	S4	20.931	[M-H]	424.4269	Unique 2/5	
m/z 429.3728		S5	17.642	[M+H]	426.3603	Unique 2/5	
m/z 429.3736	C22 H48 N6 S	S4	18.035	[M+H]	428.3659	Unique 2/5	x
m/z 439.3583	C19 H47 Cl N8 O	S4	16.954	[M+H]	438.3555	Unique 2/5	
m/a 441.3733		S5	16.28	[M+H]	440.3673	Unique 2/5	x
m/z 441.1908	C25 H30 O7	S4	9.425	[M-H]	442.1984	Unique 2/5	
m/z 441.3725		S5	18.891	[M-H]	442.3736	Unique 2/5	
m/z 445.1834	C22 H26 N3 O7	S4	13.32	[M+H]	444.1764	Unique 2/5	
m/z 443.3528		S5	18.882	[M-H]	444.36	Unique 2/5	
m/z 453.3368	C28 H44 N3 O2	S4	16.703	[M-H]	454.3441	Unique 2/5	
m/z 455.3512	C26 H44 N6 O	S4	15.679	[M-H]	456.3581	Unique 2/5	
m/z 455.3519	C28 H46 N3 O2	S4	16.182	[M-H]	456.3592	Unique 2/5	
m/z 474.3944	C30 H48 O3	S4	14.778	[M+NH4]	456.3601	Unique 2/5	
m/z 472.4114	C26 H53 N3 O4	S4	15.05	[M+H]	471.4038	Unique 2/5	
m/z 490.1720		S5	9.018	[M+NH4]	472.1396	Unique 2/5	x
m/z 471.3833	C29 H50 N3 O2	S4	19.723	[M-H]	472.3906	Unique 2/5	
m/z 492.4035	C19 H46 N12 S	S4	14.608	[M+NH4]	474.3683	Unique 2/5	
m/z 476.9847		S5	0.62	[M-H]	477.9938	Unique 2/5	
m/z 543.2798		S5	11.432	[M+CH3COO]	484.2663	Unique 2/5	x
m/z 485.3996	C25 H52 N6 O S	S4	17.18	[M+H]	484.3923	Unique 2/5	
m/z 487.4144	C30 H54 N3 O2	S4	19.726	[M-H]	488.4217	Unique 2/5	
m/z 499.4146	C31 H54 N3 O2	S4	20.867	[M-H]	500.422	Unique 2/5	
m/z 499.4153	C33 H56 O3	S4	20.653	[M-H]	500.4224	Unique 2/5	
m/z 503.4092	C30 H54 N3 O3	S4	19.966	[M-H]	504.4166	Unique 2/5	
m/s 507.2642		S5	6.281	[M+H]	506.2582	Unique 2/5	x
m/z 515.4457	C32 H58 N3 O2	S4	20.865	[M-H]	516.453	Unique 2/5	
m/z 540.4460		S5	15.962	[M+NH4]	522.4205	Unique 2/5	x

m/z 549.4856		S5	17.869	[M+Na]	526.4963	Unique 2/5	x
m/z 546.3993		S5	15.163	[M+NH4]	528.3651	Unique 2/5	
m/z 533.3833	C30 H52 N3 O5	S4	14.601	[M-H]	534.3907	Unique 2/5	
m/z 561.3954		S5	16.234	[M+Na]	538.4041	Unique 2/5	
m/z 534.4240		S5	16.065	[M+H]	542.4098	Unique 2/5	x
m/z 564.3425	C31 H45 N6 O4	S4	17.482	[M-H]	565.3499	Unique 2/5	x
m/z 573.3639	C35 H48 N4 O S	S4	13.027	[M+H]	572.3549	Unique 2/5	
m/z 573.4100	C27 H56 N8 O S2	S4	13.977	[M+H]	572.4017	Unique 2/5	
m/z 583.0272		S5	0.63	[M-H]	584.0445	Unique 2/5	x
m/z 587.4709		S5	19.205	[M+H]	586.4645	Unique 2/5	x
m/z 593.4482		S5	16.256	[M+H]	592.4457	Unique 2/5	
m/z 603.0205	C34 H19 Cl N O2 S3	S4	0.611	[M-H]	604.028	Unique 2/5	x
m/z 605.4779	C37 H66 O6	S4	18.96	[M-H]	606.4858	Unique 2/5	x
m/z 327.9712		S5	0.613	[M+Na]	609.9667	Unique 2/5	x
m/z 618.4750	C36 H67 N4 S2	S4	17.342	[M-H]	619.4805	Unique 2/5	x
m/z 643.3849		S5	15.067	[M+Na]	620.3965	Unique 2/5	
m/z 631.4132	C32 H62 N4 O2 S3	S4	13.029	[M+H]	630.403	Unique 2/5	
m/z 676.4922		S5	18.378	[M+Na]	653.5069	Unique 2/5	x
m/z 688.4921	C26 H53 N23	S4	18.27	[M+H]	687.4853	Unique 2/5	x
m/z 688.4906	C39 H67 N3 O7	S4	18.682	[M-H]	689.4981	Unique 2/5	x
m/z 706.5611		S5	18.314	[M+H]	705.5535	Unique 2/5	x
m/z 716.5257	C43 H73 N O5 S	S4	18.789	[M+H]	715.5204	Unique 2/5	x
m/z 741.4686		S5	17.569	[M+Na]	718.4832	Unique 2/5	x
m/z 719.4851	C40 H68 N2 O9	S4	17.865	[M-H]	720.4927	Unique 2/5	x
m/z 732.5729	C46 H75 N4 O S	S4	18.529	[M+H]	731.5656	Fold Change 3/5	x
m/z 762.5967		S5	18.023	[M+Na]	739.597	Fold Change 3/5	x
m/z 741.0157		S5	0.627	[M-H]	742.0332	Fold Change 3/5	x
m/z 743.4816		S5	17.55	[M+H]	742.4757	Fold Change 3/5	x
m/z 760.5125		S5	17.849	[M+NH4]	742.4793	Fold Change 3/5	x
m/z 767.4714		S5	17.548	[M+Na]	744.4839	Fold Change 3/5	x
m/z 746.5442		S5	18.152	[M+H]	745.5369	Fold Change 3/5	x
m/z 769.4869		S5	17.703	[M+Na]	746.499	Fold Change 2/5	x
m/z 778.4948	C39 H68 O14	S4	14.239	[M+NH4]	760.4607	Fold Change 2/5	x
m/z 762.5161		S5	17.551	[M+H]	761.5198	Fold Change 2/5	x
m/a 767.4714	C43 H64 N3 O9	S4	17.55	[M+H]	766.4641	Fold Change 2/5	x
m/z 782.5704	C51 H75 N O5	S4	18.621	[M+H]	781.5657	Fold Change 2/5	x
m/z 784.5830	C46 H77 N3 O7	S4	17.669	[M+H]	783.5759	Fold Change 2/5	x
m/z 784.5941		S5	19.062	[M+H]	783.5905	Fold Change 2/5	x
m/z 786.6062	C47 H77 N8 S	S4	19.517	[M+H]	785.5982	Fold Change 2/5	x
m/z 806.5091		S5	17.165	[M+NH4]	788.4779	Fold Change 2/5	x

m/z 803.5427	C45 H74 N2 O10	S4	17.552	[M+H]	802.535	Fold Change 2/5	x
m/z 803.4935		S5	17.547	[M-H]	804.5028	Fold Change 2/5	x
m/z 805.5594	C50 H76 O8	S4	17.758	[M+H]	804.5541	Fold Change 2/5	x
m/z 805.5089	C48 H68 N7 O2 S	S4	17.757	[M-H]	806.5159	Fold Change 2/5	x
m/z 814.6277	C44 H87 N5 O4 S2	S4	18.434	[M+H]	813.6202	Fold Change 2/5	x
m/z 816.6410		S5	18.773	[M+H]	815.6375	Fold Change 2/5	x
m/z 815.4965	C41 H74 N3 O11 S	S4	17.583	[M-H]	816.5037	Fold Change 2/5	x
m/z 858.6222	C58 H83 N O2 S	S4	18.866	[M+H]	857.6151	Fold Change 2/5	x
m/z 860.6387	C53 H79 N8 O2	S4	19.252	[M+H]	859.6316	Fold Change 2/5	x
m/z 882.5079	C46 H66 N5 O11	S4	15.168	[M+H]	864.4781	Fold Change 2/5	
m/z 878.5680		S5	18.316	[M+H]	877.5624	Fold Change 2/5	x
m/z 909.5577	C50 H80 N O12	S4	17.747	[M+Na]	886.5684	Fold Change 2/5	x
m/z 893.5409	C52 H74 N7 O2 S2	S4	20.649	[M+H]	892.5338	Fold Change 2/5	x
m/z 904.6040		S5	17.742	[M+H]	903.5173	Fold Change 2/5	x
m/z 907.5224		S5	18.979	[M+H]	906.5154	Fold Change 2/5	x
m/z 929.5240	C49 H78 O15	S4	17.188	[M+Na]	906.5347	Fold Change 2/5	x
m/z 931.5397		S5	17.348	[M+Na]	908.5514	Fold Change 2/5	x
m/z 937.5874	C42 H76 N17 O4 S	S4	18.155	[M+H]	914.5979	Fold Change 2/5	x
m/z 934.6400		S5	17.284	[M+NH4]	916.6057	Fold Change 2/5	x
m/z 955.5431		S5	17.305	[M+NH4]	932.5552	Fold Change 2/5	x
m/a 951.5949		S5	20.635	[M+H]	950.5877	Fold Change 2/5	x
ma/ 951.5468	C60 H72 N8 O S	S4	20.631	[M-H]	952.5536	Fold Change 2/5	x
m/z 965.5742	C56 H88 N2 O3 S4	S4	18.986	[M+H]	964.5677	Fold Change 2/5	x
m/z 965.5270	C60 H78 N4 O S3	S4	18.977	[M-H]	966.5347	Fold Change 2/5	x
m/z 997.5526		S5	20.034	[M-H]	998.5595	Fold Change 2/5	x
m/z 1011.5316		S5	18.609	[M-H]	1012.5403	Fold Change 2/5	x

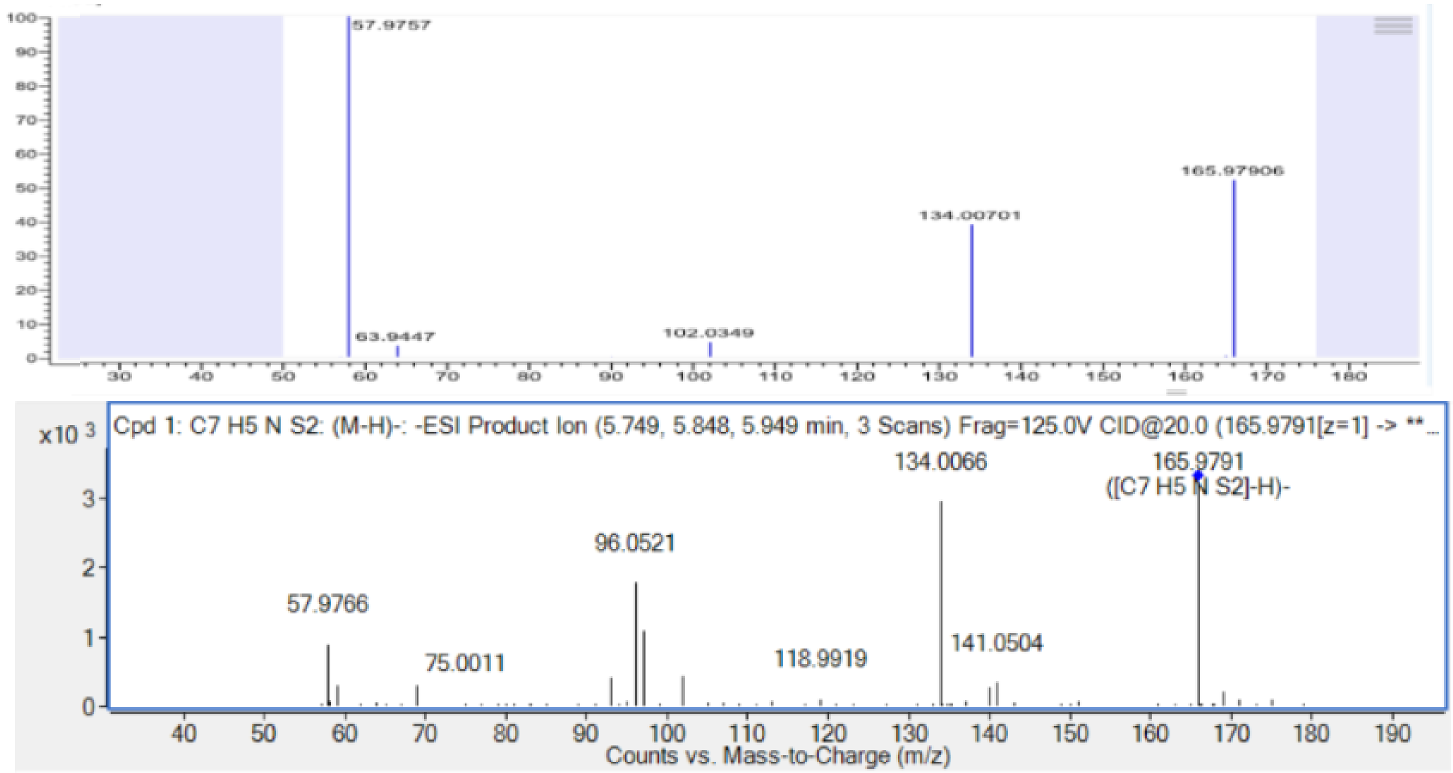

Figure 5 MS/MS spectra comparison for 2-mercaptobenzothiazole

The MS/MS spectra are shown for 2-mercaptobenzothiazole. Database spectra from mzcloud.org (top) and spectra measured in Skagit 13 (bottom) were matched on precursor ion m/z 165.979, and fragment ions m/z 134.007 and m/z 57.975.

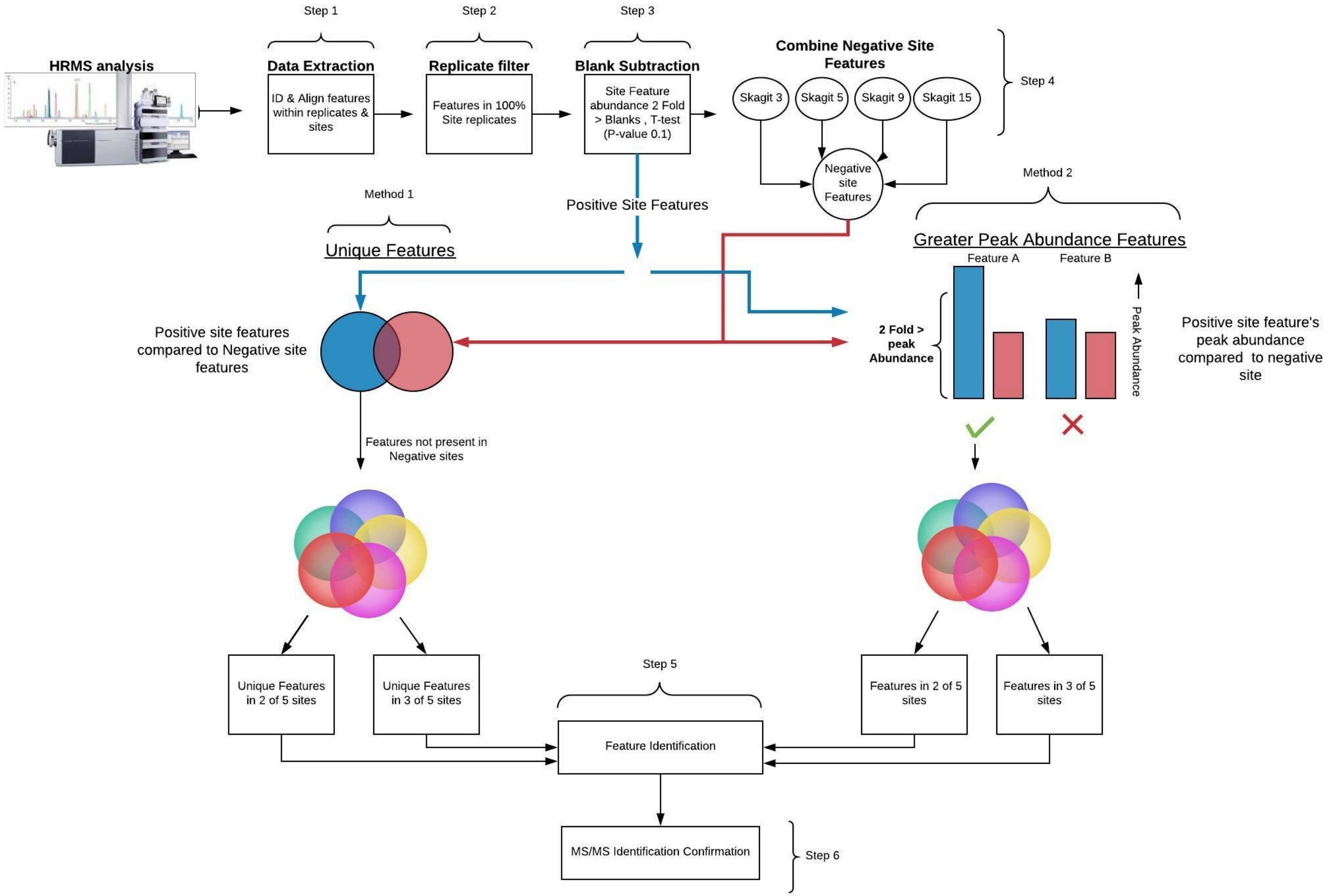

Figure 6 Data Analysis Flow Chart

References

1. Organization WH. Mortality and burden of disease from water and sanitation. http://www.who.int/gho/phe/water_sanitation/burden_text/en/. Published 2016. Accessed Sept. 29th, 2018.
2. Alhamlan F, Al-Qahtani AA, Al-Ahdal M. Recommended advanced techniques for waterborne pathogen detection in developing countries. In. *J. Infect. Dev. Ctries.* Vol 92015:128-135.
3. Gargano JW, Adam EA, Collier SA, Fullerton KE, Feinman SJ, Beach MJ. Mortality from selected diseases that can be transmitted by water - United States, 2003-2009. *Journal of water and health.* 2017;15(3):438.
4. Adam E, Collier S, Fullerton K, Gargano J, Beach M. Prevalence and direct costs of emergency department visits and hospitalizations for selected diseases that can be transmitted by water, United States. *Journal of Water and Health.* 2017;15(5):673-683.
5. Arnone RD, Walling JP. Waterborne pathogens in urban watersheds. *Journal of water and health.* 2007;5(1):149.
6. Ramírez-Castillo FY, Loera-Muro A, Jacques M, et al. Waterborne Pathogens: Detection Methods and Challenges. In: LeChevallier MW, ed. *Pathogens.* Vol 4.2015:307-334.
7. Hrudey SE, Payment P, Huck PM, Gillham RW, Hrudey EJ. A fatal waterborne disease epidemic in Walkerton, Ontario: comparison with other waterborne outbreaks in the developed world. *Water science and technology : a journal of the International Association on Water Pollution Research.* 2003;47(3):7.
8. Beer K, Gargano JW, Roberts VA, et al. Surveillance for Waterborne Disease Outbreaks Associated With Drinking Water — United States, 2011–2012. *American Journal of Transplantation.* 2015;15(12):3260-3267.
9. Agency USEP. National Primary Drinking Water Regulations. https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations. Published 2009. Accessed Sept. 29th, 2018.
10. U.S. Environmental Protection Agency. 2012 Recreational Water Quality Criteria. https://www.epa.gov/sites/production/files/2015-10/documents/rec-factsheet-2012.pdf. Published 2012. Accessed Sept. 29th, 2018.
11. U.S. Environmental Protection Agency. National Summary of Impaired Waters and TMDL information. Published 2018. Accessed Sept 29th, 2018.
12. Breen B, Curtis J, Hynes S. Water quality and recreational use of public waterways. *Journal of Environmental Economics and Policy.* 2018;7(1):1-15.
13. Division, of Marine Fisheries. *Report at Mid-Decade; Commonwealth of Massachusetts,.* Boston, MA1985.
14. Washington State Department of Ecology. Focus on Microbial Source Tracking. In:2012.
15. Sargeant D, Kammin W, Collyard S. Review and Critique of Current Microbial Source Tracking (MST) Techniques. In: Program EA, ed: WA State Dept. of Ecology; 2011.
16. Furukawa T, Yoshida T, Suzuki Y. Application of PFGE to source tracking of faecal pollution in coastal recreation area: a case study in Aoshima Beach, Japan. *Journal of applied microbiology.* 2011;110(3):688.
17. Mohapatra BR, Broersma K, Nordin R, Mazumder A. Evaluation of Repetitive Extragenic Palindromic-PCR for Discrimination of Fecal Escherichia coli from Humans, and Different Domestic- and Wild-Animals. *Microbiology and Immunology.* 2007;51(8):733-740.

18. Kuntz RL, Hartel PG, Godfrey DG, Segars WI, McDonald JL, Gates KW. Targeted Sampling Protocol as Prelude to Bacterial Source Tracking with Enterococcus faecalis. *Journal of Environmental Quality.* 2003;32(6):2311-2318.
19. Stoeckel DM, Harwood VJ. Performance, Design, and Analysis in Microbial Source Tracking Studies. *Applied and Environmental Microbiology.* 2007;73(8):2405.
20. Bergeron P, Oujati H, Catalán Cuenca V, Huguet Mestre JM, Courtois S. Rapid monitoring of Escherichia coli and Enterococcus spp. in bathing water using Reverse Transcription-quantitative PCR. *International Journal of Hygiene and Environmental Health.* 2011;214(6):478-484.
21. Eichmiller JJ, Hicks RE, Sadowsky MJ. Distribution of genetic markers of fecal pollution on a freshwater sandy shoreline in proximity to wastewater effluent. *Environmental science & technology.* 2013;47(7):3395.
22. Harwood et al. Microbial Source Tracking Standard Operating Procedures Binary PCR Protocol for Human-Associated Bacteroides, Methanobrevibacter smithii and Human Polyomaviruses. In: University of Southern Mississippi; 2010.
23. Hagedorn C, Weisberg S. Chemical-based fecal source tracking methods: current status and guidelines for evaluation. *Reviews in Environmental Science and Bio/Technology.* 2009;8(3):275-287.
24. Leeming R, Ball A, Ashbolt N, Nichols P. Using faecal sterols from humans and animals to distinguish faecal pollution in receiving waters. *Water Research.* 1996;30(12):2893-2900.
25. MacDonal I, Bokkenheuser V, Winter J, McLernon A, Mosbach E. Degradation of fecal sterols in the human gut. *J Lipid Res.* 1983;24:675-694.
26. Bull ID, Lockheart MJ, Elhmmali MM, Roberts DJ, Evershed RP. The origin of faeces by means of biomarker detection. *Environment International.* 2002;27(8):647-654.
27. Boving T, Meritt D, Boothroyd J. Fingerprinting sources of bacterial input into small residential watersheds: fate of fluorescent whitening agents. *International Journal of Geosciences.* 2004;46(2):228-232.
28. Poiger T, Field JA, Field TM, Siegrist H, Giger W. Behavior of fluorescent whitening agents during sewage treatment. *Water Research.* 1998;32(6):1939-1947.
29. Dickerson JW, Hagedorn C, Hassall A. Detection and remediation of human-origin pollution at two public beaches in Virginia using multiple source tracking methods. *Water Research.* 2007;41(16):3758-3770.
30. Seiler RL, Zaugg SD, Thomas JM, Howcroft DL. Caffeine and Pharmaceuticals as Indicators of Waste Water Contamination in Wells. *Ground Water.* 1999;37(3):405-410.
31. Nicklas P. Screening for non-regulated organic compounds in municipal wastewater in Göteborg, Sweden. *Water Science and Technology.* 1996;33(6):9-15.
32. Glassmeyer ST, Furlong ET, Kolpin DW, et al. Transport of chemical and microbial compounds from known wastewater discharges: potential for use as indicators of human fecal contamination. *Environmental science & technology.* 2005;39(14):5157.
33. James CA, Miller-Schulze JP, Ultican S, Gipe AD, Baker JE. Evaluating Contaminants of Emerging Concern as tracers of wastewater from septic systems. *Water Research.* 2016;101:241-251.
34. Buerge IJ, Buser H-R, Kahle M, Müller MD, Poiger T. Ubiquitous occurrence of the artificial sweetener acesulfame in the aquatic environment: an ideal chemical marker of domestic wastewater in groundwater. *Environmental science & technology.* 2009;43(12):4381.
35. Hernández F, Sancho J, Ibáñez M, Abad E, Portolés T, Mattioli L. Current use of high- resolution mass spectrometry in the environmental sciences. *Analytical and Bioanalytical Chemistry.* 2012;403(5):1251-1264.
36. Brenton A, Godfrey A. Accurate mass measurement: Terminology and treatment of data. *Journal of the American Society for Mass Spectrometry.* 2010;21(11):1821-1835.

37. Krauss M, Singer H, Hollender J. LC–high resolution MS in environmental analysis: from target screening to the identification of unknowns. *Analytical and Bioanalytical Chemistry.* 2010;397(3):943-951.

38. Ruff M, Mueller MS, Loos M, Singer HP. Quantitative target and systematic non-target analysis of polar organic micro-pollutants along the river Rhine using high-resolution mass-spectrometry – Identification of unknown sources and compounds. *Water Research.* 2015;87:145-154.

39. McEachran A, Hedgespeth M, Newton S, et al. Comparison of emerging contaminants in receiving waters downstream of a conventional wastewater treatment plant and a forest-water reuse system. *Environmental Science and Pollution Research.* 2018;25(13):12451-12463.

40. Du B, Lofton JM, Peter KT, et al. Development of suspect and non-target screening methods for detection of organic contaminants in highway runoff and fish tissue with high-resolution time-of-flight mass spectrometry. *Environ Sci: Processes Impacts.* 2017;19(9):1185-1196.

41. Peter KT, Tian Z, Wu C, et al. Using High-Resolution Mass Spectrometry to Identify Organic Contaminants Linked to Urban Stormwater Mortality Syndrome in Coho Salmon. *Environmental science & technology.* 2018;52(18):10317.

42. Hayes JE, McGreevy PD, Forbes SL, Laing G, Stuetz RM. Critical review of dog detection and the influences of physiology, training, and analytical methodologies. *Talanta.* 2018;185:499-512.

43. Van De Werfhorst LC, Murray JLS, Reynolds S, Reynolds K, Holden PA. Canine scent detection and microbial source tracking of human waste contamination in storm drains. *Water environment research : a research publication of the Water Environment Federation.* 2014;86(6):550.

44. Skagit County Public Works. *Skagit County/Environmental Canine Services: Human Sewage Detection Project.* 2014.

45. US Census Bureau. *Skagit County Washington 2017 estimated population.* 2017.

46. WA Dept. of Ecology. Focus on Samish Watershed. In: Program WQ, ed2006.

47. Fenech C, Rock L, Nolan K, Tobin J, Morrissey A. The potential for a suite of isotope and chemical markers to differentiate sources of nitrate contamination: A review. *Water Research.* 2012;46(7):2023-2041.

48. Schulze T. et al. NORMAN MassBank - Towards a community-driven, open-access accurate mass spectral database for the identification of emerging pollutants. 2012:- 9.

49. Schymanski EL, Jeon J, Gulde R, et al. Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environmental science & technology.* 2014;48(4):2097.

50. Reddy C, Quinn J. Environmental chemistry of benzothiazoles derived from rubber. *ENVIRON SCI TECHNOL.* 1997;31(10):2847-2853.

51. Kloepfer A, Gnirss R, Jekel M, Reemtsma T. Occurrence of benzothiazoles in municipal wastewater and their fate in biological treatment. *Water Science and Technology.* 2004;50(5):203-208.

52. Jungclaus G, Games L, Hites R. Identification of trace organic compounds in tyre manufacturing plant waste waters. *Analytical Chemistry.* 1976;48(13):1894-1896.

53. Fiehn O, Reemtsma T, Jekel M. Extraction and analysis of various benzothiazoles from industrial wastewater. *Analytica Chimica Acta.* 1994;295(3):297-305.

54. Kloepfer A, Jekel M, Reemtsma T. Occurrence, sources, and fate of benzothiazoles in municipal wastewater treatment plants. *Environmental science & technology.* 2005;39(10):3792.

55. Liao C, Kim U, Kannan K. A Review of Environmental Occurrence, Fate, Exposure, and Toxicity of Benzothiazoles. In. *Environ. Sci. Technol.* Vol 522018:5007-5026.

56. Skagit County Public Works. Human Bacteria Source Tracking with Trained Dogs Skagit County, 2015-2016. In:2016.