Development and Demonstration of a Fecal Source Reference Catalog and Methodology for Improved Molecular Source Tracking Analysis for Water Quality Management

Final Report
July 2019

Report prepared for: Washington State Conservation Commission
Grant Funding Focus area: Shellfish, Whatcom County, Puget Sound

Contributing report authors:
Nichole Embertson[1], Meagan Harris[1], Scarlett Graham[1], Ryan McLaughlin[2], Jenna Brooks[3], Casey Schlenker[3], Kent Oostra[3]

[1]*Whatcom Conservation District*
[2]*Practical Informatics*
[3]*Exact Scientific Services*

Project Contacts

Lead Project Manager/Contact
Nichole Embertson, Ph.D.
Whatcom Conservation District
6975 Hannegan Rd.,
Lynden, WA 98264
P: (360) 526-2381
F: (360) 354-4678
E: nembertson@whatcomcd.org

Kent Oostra
Exact Scientific Services, Inc.
1355 Pacific Place, Suite 101
Ferndale, WA 98248
P: (360) 733-1205
E: k.oostra@exactscientific.com

Ryan McLaughlin
Practical Informatics, LLC
1721 Electric Avenue
Bellingham, WA 98229
P: (814) 758-4570
E: mclaughlin@practical-informatics.com

Table of Contents

Tables

Figures

1. Background and Purpose

Water quality pollution is a concern in many areas in Washington State whether for recreation, consumption, habitat, or shellfish harvest. One area of intense focus in the State is Whatcom County. Since 1998 a variety of water resource management stakeholders, local and state agencies, and tribal governments have worked together to characterize water quality issues as well as to identify potential management solutions. Fecal coliform bacteria has served as the predominant water quality health measure in Whatcom County and shellfish harvest areas due to the impact of pathogenic fecal microorganisms from human sewage, livestock, and other wild/domestic animals to human populations through contaminated water.

In western Washington, particularly Whatcom County, shellfish harvest for recreational, commercial, and tribal purposes is under constant threat by incoming water pollution from upstream sources. Seasonal closures have led to lost revenues of commercial harvesters and tribes and has spurred lawsuits and local action (i.e., Portage Bay Partnership, Pollution, Identification and Correction program). To try and address that issue, millions of dollars have been spent in Washington State on water quality monitoring, mitigation, planning, and protection, but problems still persist. Locally, the Whatcom Clean Water Program (WCPW) conducts routine water quality monitoring to track fecal coliform across the County lowlands and into the bays. Of the 84 freshwater stations with at least three years of data, only 16 (less than 20%) met water quality standards in 2017 for fecal coliform (Douglas 2017). Elevated bacteria levels in marine waters have also led to the establishment of three shellfish protection districts in Whatcom County: Drayton Harbor, established in 1995, Portage Bay, established in 1998, and Birch Bay, established in 2009. Of these three, two have had large sections re-opened in the late 2016 (Drayton) and early 2018 (Birch Bay), but the third, Portage Bay, remains seasonally closed, causing significant financial and cultural impact.

While some progress has been made in reducing fecal coliform levels/sources and advancing towards cleaner water, current methods of water quality microbial pollution identification by culturing fecal indicator bacteria (FIB), such as fecal coliform or *E. coli*, fails to identify individual sources and has a high degree of uncertainty. A technique that identifies individual sources of fecal pollution is necessary to make further advances in water quality improvements.

The field of molecular biology offers an alternative to the culture-based FIB methods currently being employed for water quality monitoring (Wiggins 1996, Parveen et al. 1997, Hagedorn et al. 1999, Bernhard et al. 2000, Harwood et al. 2000). Molecular source tracking (MST) is a general term for methods specifically designed to provide biological source information from environmentally derived DNA. The most common methods employed are end-point polymerase chain reactions (PCR) and quantitative PCR (qPCR) to identify the presence/absence of specific FIBs. While these methods improve upon the traditional cultivation-dependent FIB method, they still have an inherent bias for a subset of microbes. Detection of fecal pollution by MST targets a finite number of genetic markers from a limited number of gut microbes, which are believed to originate from specific sources. It is unclear whether these targets are relevant across different geographical regions (Odagiri et al. 2015). These and other issues have led to uncertainty when using targeted MST to determine the causal sources of fecal contamination. Comprehensive, geographically specific fecal sources have proven more effective (Brown et al. 2017).

An alternative to the previously described MST methods is to utilize high-throughput DNA sequencing using next-generation sequencing (NGS) platforms. Two common methods are the metagenetic survey, usually a single gene such as the 16S rRNA gene, and shotgun metagenomics, both aimed at providing a less biased and comprehensive analysis of microbial communities. For metagenetic surveys and shotgun

metagenomics, DNA is extracted directly from all organisms present in a sample and prepared for sequencing without a culturing; referred to as whole sample sequencing (WSS)[1]. This allows for a more complete interrogation of the microbial community by including more organisms and genes for analyses. In the last decade, MST studies using NGS have become more frequent and it is quickly becoming a well-established, though under-utilized, method for identifying fecal contamination sources (Ahmed et al. 2015). This study utilized WSS to look at microbial community characterization and overcome the limitations of culture-based methods by providing a far more comprehensive and definitive fecal source characterization.

The goals of this project were to begin building a more comprehensive fecal source reference catalog within Whatcom County, develop standard operating procedures will be utilized in other areas, compare data from traditional FIB methods, and begin testing the application of the fecal catalog on water spiked with fecal material obtained from Whatcom County, water samples suspected of being impacted, and water samples with unknown sources. A whole sample DNA sequencing (WSS) method using a next generation DNA sequencing (NGS) platform was used to sequence all DNA obtained directly from fecal sources to provide a more complete depiction of all microbial species present in a sample. This comprehensive analysis provides two advantages over traditional targeted MST: 1) it does not rely on any particular species of microbe to indicate fecal contamination from a source, but does support the improvement of targeted MST methods by providing a more comprehensive catalog, and 2) it is relevant to the geographical region it has been developed in. Another advantage of WSS analysis is that it is entirely culture independent and thus removes culture bias (Brown et al. 2018). The improvement of fecal source characterization in water samples may lead to improved source identification, management, and prevention of pollution to waterways and long-term improvements in shellfish production areas.

2. Project Deliverables

The project had four focus areas: 1) development of a fecal source reference catalog, 2) development of a standard operating procedure for fecal source reference sample collection, 3) comprehensively characterize the microbial community present on fecal coliform culture plates, and 4) demonstrate the utility of the reference catalog to analyze water samples (see specific deliverables below). Fecal reference and water validation sampling was conducted throughout Whatcom County, with particular emphasis on the Nooksack Watershed which empties into Portage Bay, a primary shellfish growing area.

Project deliverables included:

1. **Fecal Source Reference Catalog Development** - Obtain fecal samples from a variety of sources and comprehensively characterize the microbial community of each fecal sample via whole sample DNA sequencing (WSS) utilizing a next generation DNA sequencing (NGS) platform. Sampling methodology has intention of demonstrating statistical variation within and between fecal sources by location and season. Intensive focus will be provided for dairy.
2. **Fecal Source Reference Sample Collection SOP** - Establish a standard operating procedure (SOP) for collection and handling of fecal source reference samples and share widely for better coordination and consistency of fecal source reference sample collection and characterization.

[1] For the purposes of this document “Whole Sample Sequencing” or “WSS” will refer to the combined usage of 16S single gene survey and shotgun metagenomics, unless only one is used at which point “16S” or “shotgun” will denote the specific methodology.

3. **Fecal Coliform Plate DNA Characterization** – Characterize the microbial community of colonies grown on standard fecal coliform membrane filtration culture plates, and compare to the fecal source reference catalog. Determine if this method could provide an effective method to quickly identify the pollutant source of a "high" fecal count sample without the need for additional water sample collection.
4. **Water Quality Characterization Validation/Demonstration** – Collect water samples from high priority areas in Whatcom County and analyze via WSS to compare DNA sequence data to those produced from the fecal source reference catalog. Additionally, demonstrate the effectiveness of the reference catalog for identifying known and unknown pollution sources and provide more definitive information for local water quality groups.

3. Expected Benefits and Impacts

Whatcom County has a very diverse group of land users including rural landowners, agriculture, hobby farms, urban cities, tribal lands, forestry, and aquaculture at the outflow of upland rivers. All of these stakeholders are impacted by water quality and would benefit from improvements in water pollution characterization. In particular, shellfish industries - Tribal, recreational, and commercial, - that are downstream of potential pollutant loads, will be one of the greatest beneficiaries to improved water quality via reduction of pathogenic bacteria. Natural ecological systems - fish, wildlife, and marine – will also benefit from improved water quality.

This project was designed to provide a foundation for future MST work in Whatcom County, and possibly, the larger Puget Sound. We anticipate that methods developed in this study will be applicable to all of western Washington and areas with similar water quality challenges. This project aims to provide more effective means of water pollution fecal source characterization. In this way, it can improve agency decision making for water quality monitoring, source identification, allocation of mitigation resources, targeted conservation planning, focused education and outreach efforts, and improved landowner responsibility and community engagement. Additionally, this work contributes to the development of a framework for a more holistic assessment of pollution sources for long-term improvement of water quality.

4. Materials and Methods

4.1. DELIVERABLE 1 - Fecal Source Reference Catalog Development

Obtain fecal samples from a variety of sources and comprehensively characterize the microbial community of each fecal sample via whole sample DNA sequencing (WSS) utilizing a next generation DNA sequencing (NGS) platform.

The primary deliverable of this project was the development of a comprehensive fecal source reference catalog for Whatcom County. The WSS method was used for sequencing DNA directly obtained from fecal sources as it provides a comprehensive output that can be used by any type of MST method (including qPCR) to build a method specific fecal reference catalog marker.

4.1.1. Fecal Sample Collection

Prior to fecal sampling by staff or partners, sample kits were created and distributed along with a sample intake form, sampling and processing instructions, and when applicable, a live training on proper

sample collection technique. Trainings were conducted at WCWP Field Staff meetings and at the WCD office. This was to ensure that all sample collection was consistent and samples could be confidently used by the project.

Fecal source reference samples (Table 1) were taken over the course of a year from the following sources: dairy cattle (liquid, solid, and fresh), beef cattle (fresh and grazed old patties), sheep, goat, deer, horse, pig, beaver, raccoon, dog, roofbird (predominantly starling), goose, swan, duck, seagull, chicken, wastewater treatment plant, human septic, boat pump-out (human), and natural (upper Nooksack River). The sources were chosen based on local knowledge and historical observation of primary and highly likely sources of microbial contamination into waterways in Whatcom County.

Because of presumed natural, seasonal microbial variability, additional attention was paid to dairy lagoon sources with an in-depth look into temporal, spatial, and management variations that may affect the fecal source microbial community. Comprehensive work has not been conducted on this source to see if variables such as manure management strategy, solid vs liquid vs fresh manures, feed types, operation size, or dairy cattle breed types, affect the microbial signature. Additionally, we explored if season or location affected the liquid manure storage microbial community by conducting sampling in the Fall, Winter, and Spring at the same sites and comparing each individual to the seasonal composite.

Following best available guidance (Battelle 2005, EPA 2005, Ecology 2011, and EPA 2011) and in collaboration with SOP development, each fecal reference composite sample reflects a composite of at least 10 individuals and/or 5 locations for each reference. In some cases, this criteria was not met (i.e., Beef Cattle - Old) and these samples were used as comparisons, not catalog reference samples, or are awaiting additional samples (i.e., beaver, raccoon) to be included in the catalog. A minimum of 20 grams of sample was collected into sterile containers (bag, bottle, vial) for each fecal type. To obtain the required number of composite sub-samples, individual site or species fecal samples were collected opportunistically and frozen until the necessary number were obtained to create a composite. Fecal samples were frozen one time prior to extraction with no detriment to the subsequent DNA analysis. For quality control and quality assurance of both field collection procedures and lab analysis, a duplicate and split sample were collected representing 5% of all samples. Samples were submitted to the lab with consistent nomenclature and coding for sample type identification.

Fecal sources were sampled at times of the year they were most likely to be present and/or have an impact on water quality pollution. All reference samples were collected directly from the sources without cross-contamination. Local partners assisted with fecal source collection. Standard operating procedures were followed for all fecal source collection.

Table 1. Fecal sample reference type with number of sites and individuals sampled over a date range with relevant notes

Fecal Reference Sample Type	# of Individual Sites Sampled	# of Individuals Represented in Sample	Date Range of Sample Collection	Notes
Ruminant Livestock - Dairy				
Dairy Lagoon - Fall	9	1000+	October 2018	Individual sites were also sequenced separately from the composite.
Dairy Lagoon - Winter	9	1000+	March 2019	Individual sites were also sequenced separately from the composite.
Dairy Lagoon - Spring	9	1000+	April 2019	Individual sites were also sequenced separately from the composite.

Dairy Solids	5	1000+	Sept 2018 - April 2019	Individual sites were also sequenced separately from the composite.
Dairy Fresh	1	10	April 2019	Sampled from multiple barns at one farm location that was also part of the lagoon composite.
Ruminant Livestock				
Beef Cattle	5	22	Dec 2018 - Jan 2019	No remarkable observations.
Beef Cattle - Old	1	4	December 2018	Sampled patties from one field that had been present for greater than 2 months.
Sheep	5	38	Oct 2018 - April 2019	Individual sites were also sequenced separately from the composite.
Goat	4	20	Oct 2018 - April 2019	No remarkable observations.
Deer	7	26	Dec 2018 - April 2019	Locations from multiple areas in Whatcom County.
Non Ruminant Livestock				
Horse - Winter	6	17	Dec 2018 - Jan 2019	No remarkable observations.
Horse - Spring	5	11	April 2019	Individual sites were also sequenced separately from the composite.
Pig	4	39	Dec 2018 - Jan 2019	No remarkable observations.
Wildlife and Domestic				
Beaver	4	4+	Jan - May 2019	Samples taken from bottom of pond and land.
Raccoon	5	9+	Dec 2018 - April 2019	Age of samples were 1-4 days old.
Dog	10	14	Nov - Dec 2018	Samples taken from dog park or brought in same day by individual dog owners.
Avian				
Roofbird	3	39	Dec 2018 - April 2019	Mix of Starling, Blackbird, sparrow, and other barn birds
Roofbird - Old	3	23	Dec 2018 - April 2019	Mix of Starling, Blackbird, sparrow, and other barn birds
Goose	2	16	Nov - Dec 2018	No remarkable observations.
Swan	3	27	Nov - Dec 2018	No remarkable observations.
Duck	5	84	Dec 2018 - March 2019	No remarkable observations.
Seagull	3	16	Nov 2018 - May 2019	Collected samples from surfaces.
Chicken	4	23	Nov 2018 - April 2019	No remarkable observations.
Human				
WWTP Influent - Winter	1	1000+	December 2018	2 - hour composite
WWTP Influent - Spring	1	1000+	April 2019	2 - hour composite

Septic	1	>3	November 2018	Single sample from one septic pumpout.
Boat Pumpout	3	>3	May 2019	Taken during pump out of boat septic. Age of sample was 1 day to 1 week.
Natural				
Nooksack Upper Mainstem	1		April 2019	Taken above the majority of low-land tributary influence, but still has some influence upstream.
Nooksack Upper North Fork	1		May 2019	Taken high in watershed away from any discernible fecal sources.

4.1.2. Wildlife Tracker App

To help project partners identify the location of wildlife species for fecal collection, we created a Wildlife Tracker App using ArcGIS and Survey123. This app enables the public/landowners/partners to provide a location, date, species, number of animals, notes, and photo of animal(s) seen. This tool allowed us to locate wildlife for fecal and water validation sampling, as well as enable public viewing of patterns and locations of wildlife across the County for a better understanding of wildlife annual migration, concentration, and potential impacts to water quality.

4.1.3. Sample Handling and Storage

Individual site or species fecal samples were collected opportunistically and frozen until the necessary number were obtained for compositing, processing, and DNA extraction by the laboratory. A hold time/condition study was performed and demonstrated that a single freeze step caused no statistically significant change in the microbial community.

4.1.4. Sample Receipt and Storage

Fecal samples were given a unique identifier that followed the sample through the entire study. Fresh, liquid fecal samples were stored at 4°C for a maximum of 24 hours before DNA extraction. If retained longer, 50 ml of homogenized sample were poured into a sterile vial and frozen at -20°C for no longer than 6 months. Solid samples were treated similarly; 100 g was transferred to a sterile WhirlPak® bag and stored at 4°C for a maximum of 24 hours before DNA extraction. Solid fecal samples that were frozen upon receipt were thawed at 4°C. Samples stored longer than one month were placed directly in an ultra-low temperature freezer at -80°C. Pour-off procedures are detailed in the sample homogenization methods section. All storage temperatures were monitored and recorded twice daily.

4.1.5. Fecal Thaw method

Retained frozen liquid fecal aliquots were thawed in a cold-water bath to minimize microbial community changes. Solid fecal samples were thawed in a refrigerator at 4°C. DNA extractions were done immediately upon thawing. If individual frozen fecal samples were saved for a composite, compositing and homogenizing were done immediately upon thawing with DNA extraction directly following.

4.1.6. Fecal Homogenization and Composite Methods

Solid fecal. Solid fecal samples were poured into a large volume sterile WhirlPak® bag and kneaded until homogenous. Once samples were completely homogenized, aliquots were weighed into a sterile WhirlPak® using an individually weighted calculation (see Equation 1 below) to account for the number of individuals in each sample and ensure even mass distribution within the 100-gram composite. If the total mass available was less than 100 grams, a proportional mass was used to accommodate all samples

within a composite. Scales used for composition were calibrated daily. Composites were then homogenized by shaking and/or with a stomacher.

Equation 1: $W = (G / T) * N$

where W = Weight of sample to be added to composite bag, G = Total grams in composite, T = Total number of individuals contributing to the composite, and N = Number of individuals in sample bag.

Liquid fecal. Liquid fecal samples were composited by pouring a 50 mL aliquot into a 1 L sterile container. Composites typically consisted of 10-12 individual samples and were homogenized by shaking.

4.1.7. DNA Extraction

Following sample homogenization, DNA extractions were performed using the NucleoSpin DNA Soil kit (Takara Bio). Liquid fecal samples were shaken to homogeneity and DNA was extracted from a 400 µL aliquot. Solid fecal samples were homogenized with a sterile spatula and DNA was extracted from 100 – 200 mg. The protocol provided with the extraction kit was followed without deviation. Extracted DNA was quantified using a Qubit 3.0 fluorometer (Invitrogen). DNA was diluted to <5 ng/µL for subsequent DNA sequencing steps.

4.1.8. 16S DNA Sequencing and Data Analysis

Polymerase chain reaction (PCR) was performed with primers targeting the variable V3 and V4 regions of the 16S rRNA gene, producing an approximately 550 base pair DNA sequence. This gene region is broadly recognized as diagnostic for nearly all microbes (Klindworth et al. 2013). Library preparation was performed according to the Illumina 16S Metagenomic DNA Sequencing Library Preparation (Part # 15044223 Rev. B). Libraries were quantified using a Qubit 3.0 fluorometer and quality was assessed prior to DNA sequencing by electrophoresis and qPCR. DNA sequencing was done on an Illumina MiSeq utilizing v3 reagent chemistry.

Raw DNA sequence data from the Illumina MiSeq were collected in the form of paired-end FASTQ files and uploaded onto bioinformatics computer systems for processing. The Qiime2 software suite (ver. 2019.4) was used for the major steps of sequence analysis (Bolyen et al. 2018). The paired end fastq files for each sample were converted to Qiime file format. Amplicon denoising and basic quality control were performed using the DADA2 software (Callahan et al. 2016, via q2-dada2). Representative sequence output from DADA2 were aligned using MAFFT software (Katoh et al. 2002, via q2-alignment), then masked and used to infer a rooted and unrooted phylogenetic tree (Price et al. 2010, via q2-phylogeny). Representative sequences were taxonomically classified using the sklearn Naïve Bayes classifier available in Qiime (Bokulich et al. 2018, via q2-feature-classifier) and a pre-trained version of the Silva SSU RNA database (Quast et al. 2013, ver 132).

Alpha and beta diversity metrics were calculated for each sample using the core-metrics-phylogenetics workflow in Qiime (Faith 1992, Lozupone et al. 2005, Lozupone et al. 2007). An alpha rarefaction curve was built from the DNA sequence reads produced for each sample to determine if the resulting data was sufficiently representative of microbial diversity. Principal coordinate analysis (PCoA) was utilized for visualizing the similarities of microbial communities between fecal source samples using beta diversity metrics as input. Leave-one-out (LOO, see description below) analyses were performed on all sources, rarefied to 10,000 reads, using the SourceTracker (ver. 2) software to establish the microbial community overlap between source types (Knights et al. 2011). The DESeq2 R package was used to perform variance stabilization on the representative sequence count table prior to clustering analysis (Love et al. 2014). The stats package from the R statistical environment was used to perform hierarchical clustering analysis (HCA) using the 'ward.D2' method on variance stabilized representative sequence counts.

4.1.9. Shotgun DNA Sequencing and Data Analysis

Prior to shotgun library preparation, DNA was normalized to 0.2 ng/µL and quantified using Qubit 3.0 fluorometer. Library preparation was performed with the Illumina Nextera XT DNA Library Preparation Kit, following manufacturer protocols. Resulting DNA libraries were quantified using a Qubit 3.0 fluorometer and quality was assessed by electrophoresis and qPCR. DNA sequencing was done on an Illumina MiSeq utilizing v3 reagent chemistry.

Raw sequence data from the Illumina MiSeq were collected in the form of paired-end FASTQ files. Initial quality trimming was performed on each sample using Trimmomatic (Bolger et al. 2014, ver 0.38). Metagenome assembly was performed using the MEGAHIT sequence assembler (Li et al. 2016, ver 1.0). Open reading frames (ORFs) were predicted for each sample using Prodigal (Hyatt et al. 2010). The hmmsearch script from the HMMer software suite was used to predict functional gene annotations for ORFs using a custom updated version of the Clusters of Orthologous Groups (COGs) genes (Eddy 2011, ver 3.2.1, Tatusov et al. 2000). A custom parsing script written in Python was used to extract annotations and associated ORFs to be stored as representative functional sequences for future MST research.

4.1.10. Leave-one-out Analysis

Leave-one-out (LOO) analysis quantifies how similar source microbial signatures are to each other. It is performed by iteratively removing a source from the catalog and using the removed source as a mock-sample for comparison. The mock-sample (removed source) is then compared to the entire catalog (minus itself) to see if its microbial signature is similar to anything else in the catalog. If its microbial community signature is unique it will not have a high overlap with anything in the catalog. However, it does have a high overlap with something in the catalog this can mean that either the mock-sample and the matched source are very similar to each other or that the groupings that have been chosen for the catalog are not correct and need to be modified to maximize the distinction between sources. Leave-one-out analysis was performed on the 16S fecal source reference sample data.

4.1.11. Data storage

All raw DNA sequence data were stored in-house at both Exact Scientific Services, Inc. and Practical Informatics, LLC as well as an off-site cloud storage service. Processed and abstracted data were stored in-house at Practical Informatics, LLC, as-well as an off-site cloud storage service. All DNA sequence data were backed-up on a 24-hour schedule at Practical Informatics, LLC. Practical Informatics, LLC will serve as the long-term storage service for the duration of the project.

4.2. DELIVERABLE 2 - Fecal Source Reference Sample Collection SOP

Establish a standard operating procedure (SOP) for collection and handling of fecal source reference samples and share widely for better coordination and consistency of fecal source reference sample collection and characterization.

4.2.1. Fecal Collection Standard Operating Procedure Development

No known standard operating procedure (SOP) exists for fecal source sampling for DNA reference catalog development. Fecal reference catalog samples are typically taken by different groups using different sample attainment or compositing techniques producing ambiguous results and/or a lack of confidence in comparing or sharing results. A standard protocol was developed to ensure that all fecal reference samples were taken using the same guidance for this project and for future work of this type.

Starting with guidance from current, available, and known sources (Battelle 2005, EPA 2005, Ecology 2011, and EPA 2011), this project worked towards the development of a fecal reference source standard operating procedure. Due to the short timeframe of the project, the SOP is still out for review by project partners. The development of the SOP was an adaptive process throughout the entire project based on the field procedure assessment and results of fecal reference testing.

4.2.2. Field Intake Form

An intake form was created to capture information for each fecal sample collection event specific to the sample type and site it was obtained from. This form, created specifically for this project, assisted in understanding the observable microbial community differences between and within a sample, as well as differences among sampling methods. Pictures of fecal samples and notes on sampling procedures were collected to assist with the creation of the SOP.

4.2.3. Fecal Homogenization Study

To ensure homogenization of composites, a standard homogenization technique was established and utilized. For solid fecal sample type, dairy solids were used as a model. Two solid samples were composited in equal proportion and homogenized by shaking/kneading. For liquid/slurry fecal sample types, two liquid manure samples were used to evaluate liquid manure homogenization and pour-off procedures. Each sample was mixed by inversion and tested in triplicate.

After homogenization, DNA extractions were performed using the NucleoSpin DNA Soil kit (Takara Bio) to test for consistency of homogenization technique. Polymerase chain reaction (PCR) was performed with 16S rRNA gene primers targeting the variable V3 and V4 region of the gene producing an approximately 550 bp amplicon. DNA sequencing was then performed using an Illumina MiSeq v2 reagent kit.

4.2.4. Hold Time Study

Hold Time
Initially, the time a fecal sample could be held between collection and DNA extraction with no observable change in its microbial community signature was unknown. This was of concern for sampling and transport conditions and prompted a hold-time study to determine the scope of variation in the microbial community during different sample hold periods.

The fecal source material from three species was used to create positive control samples for the hold time study; horse, duck, and dairy lagoon. These were chosen to determine if the species or texture of manure - liquid/slurry (dairy lagoon or solid (horse and duck) - would have microbial communities with different rates of change through time. The hold times chosen for this study were Day 0 (fresh, no storage), Day 3 (stored fresh for 3 days at 4°C) and Day 30 (frozen upon receipt, thawed at day 30 and analyzed that same day). All fecal samples were aliquoted as detailed under *Deliverable 1, Sample Handling and Storage*, and analyzed in triplicate to allow for calculation of statistical significance.

Hold times and temperature parameters were as follows:

- Day 0 – Sample collection and analysis on same day
- Day 3 – Stored for three days at 4°C
- Day 30 – Stored for 30 days at -20°C and thawed at 4°C

On the day specified, DNA was extracted from each sample in triplicate and DNA sequencing was performed as described in *Deliverable 1, 16S DNA sequencing protocol*.

Hold-time Statistical Analysis

Raw sequence data analysis followed the Qiime-based steps outlined in *Deliverable 1, 16S Sequence Data Analysis*. In order to test the significance of variation in microbial community structure for various hold times, a PERMANOVA statistical test was performed with 999 permutations using the beta-group-significance function within the diversity module of Qiime (Anderson 2001). Results were visualized as boxplot using the Qiime visualization interface.

4.3. DELIVERABLE 3 - Fecal Coliform Plate DNA Characterization

Characterize the microbial community of colonies grown on standard fecal coliform membrane filtration culture plates, and compare to the fecal source reference catalog.

Fecal coliform is the most commonly used fecal indicator bacteria (FIB) in Whatcom County for freshwaters and marine waters due to the focus on recovering shellfish growing areas previously closed to harvest. Unfortunately, microbial growth on fecal coliform culture plate does not correlate well with, nor identify, fecal pollution sources, nor does it prioritize detection based on actual risks to human health (Ervin et al. 2013). Given the rapidly falling cost of high-throughput DNA sequencing, producing a simple, cost-effective protocol to analyze the metagenome present on routinely collected fecal coliform culture plates is now practical. This study assessed fecal coliform plates of fecal sources, spiked water samples (positive fecal controls), and suspected and unknown water sources to assess its accuracy against other methods documented in the study report.

4.3.1. Sample Analysis Methods

A protocol was developed to comprehensively sequence the metagenomic DNA present on a fecal coliform culture plate with more than 200 colony forming units (CFU). A value of 200 CFU was chosen as it is used locally as a threshold for “high” bacteria (exceeding water quality criteria). Most of the fecal culture plates chosen were paired with a water sample that was also analyzed in *Deliverable 4, Water Quality Characterization* portion of this study.

4.3.2. Fecal Coliform Analysis and Storage

One hundred mL of each water sample obtained during this study were used for fecal coliform culture analysis. Aliquots of 1 ml, 10 ml and 25 ml were filtered by vacuum through a 0.45 µm pore membrane. Membranes were placed on mFC agar plates and incubated at 44.5^{0}C for 24 hours before the number of colony-forming units (CFU) was determined (method SM9222D). After incubation, all bacterial colonies from plates were retrieved with a moistened sterile swab and deposited into a tube containing 400 µL sterile phosphate buffered saline (PBS). Tubes containing swabbed bacteria were then stored in an ultra-low temperature freezer at -80^{0}C. DNA extraction was performed using the NucleoSpin Microbial DNA kit (Takara Bio). Prior to library preparation, DNA was quantified using a Qubit 3.0 fluorometer.

4.3.3. Fecal Plate Shotgun DNA Sequencing and Data Analysis

To sequence the fecal culture plate DNA, it was normalized to 0.2 ng/µL and the Illumina Nextera XT DNA sample library preparation kit was used following manufacturer protocols. Libraries were quantified using a Qubit 3.0 fluorometer and quality was assessed by electrophoresis and polymerase chain reaction (PCR) prior to DNA sequencing. Sequencing was performed on an Illumina MiSeq utilizing v3 reagent chemistry.

To investigate the taxonomy of organisms grown on fecal culture plates, the Metaphlan (Segata et al. 2012, ver 2.9) software suite was used along with a recently updated pangenomic clade-specific marker

database (CHOCOPhlAn ver 29). The hclust2.py Python script (from Metaphlan) was used to perform clustering and heatmap production from the output of Metaphlan. Clustering from shotgun and 16S fecal plates were compared to qualitatively assess differences in microbial structure from each type of sequence data.

4.3.4. Fecal Plate 16S DNA Sequencing and Data Analysis

PCR was performed with 16S rRNA gene primers targeting the variable V3 and V4 region of the gene, producing approximately a 550 bp amplicon. Library preparation, quantification, and DNA sequencing was performed as previously described in *Deliverable 1, 16S DNA Sequencing and Analysis* protocol.

All fecal culture plate DNA sequence data were processed using the same workflow as in *Deliverable 1, 16S DNA Sequencing and Analysis*. The SourceTracker software was used to estimate the proportions microbial communities shared between fecal sources, fecal plates, and paired water samples.

4.4. DELIVERABLE 4 - Water Quality Characterization Validation/Demonstration

Collect water samples from high priority areas in Whatcom County and analyze via WSS to compare DNA sequence data to those produced from the fecal source reference catalog.

4.4.1. Water Sample Collection

To interrogate the diagnostic utility of the fecal source reference catalog, two types of environmental water samples were obtained, one representing a "suspected" source and the other an "unknown" or unidentified source. The suspected source samples were collected downstream of an observed source such as an active beaver dam, a pond with a large duck population, or from waterways with an observable animal presence and/or identifiable feces. These samples were used to assay whether fecal sources could be detected by WSS if there was a reasonable expectation they were present in environmental water samples. An unknown sample was one from a waterway with no known or observed source, or one with many possible sources given the surrounding land uses. The unknown water samples provided an opportunity to see what fecal sources could be detected in water samples having high fecal coliform counts of unobservable origin. In addition to the environmental water samples with observed sources, deliberately spiked water samples were created with known fecal sources and analyzed as positive controls (see Quality Assurance samples).

Environmental "suspected" and "unknown" water samples were taken in conjunction with complementary projects, from identified fecal source locations, as part of routine and storm event Whatcom Clean Water Program (WCWP) sampling runs, and/or as storm samples in high priority areas. Water samples taken in conjunction with routine and storm WCWP activities are shown in

Figure **1**. Water sample collection, handling, and transport followed the methods for fecal coliform monitoring outlined by Whatcom County Public Works (Douglas 2017) or the associated project QAPP.

Water samples (suspected and unknown) were prescreened to determine viability for molecular analysis. Prescreening consisted of meeting thresholds for fecal coliform count (method SM9222D) and DNA yield. If both criteria were met, samples were further analyzed. Samples that did not meet thresholds for further analysis were filtered and stored at -20^{0}C in case follow up analysis was indicated.

In general, samples with less than 200 CFU/100mL fecal coliform (corresponding to WA State Water Quality Standards; WAC 173-201A) were not run for 16S analysis. A small number of samples with low

fecal coliform concentrations, but that had observed or unique putative sources, were selected for analysis. Samples analyzed for 16S analysis ranged in fecal coliform count from 6 to >6000 CFU/100mL, with most samples above 200 CFU/100mL. In total, 159 samples were collected and 83 samples were run for 16S DNA analysis (Table 2).

Table 2. Water samples (n = 83) run for 16S analysis by sample type

Sample Type	# samples run for 16S analysis	Range/Average fecal coliform count (CFU/100 mL)	Relevant water quality standards*
Freshwater- suspected source	38	6-6000 / 979	200 CFU/100 mL
Freshwater- unknown source	24**	8-7000 / 936	
Marine- unknown source	8	8-48 / 21	43 CFU/100 mL
Spiked QC sample	13	4-181,818 / 24,000	NA

*Estimated 90th percentile for freshwaters and marine waters based on WA State Water Quality Standards.
**Includes 2 field duplicates

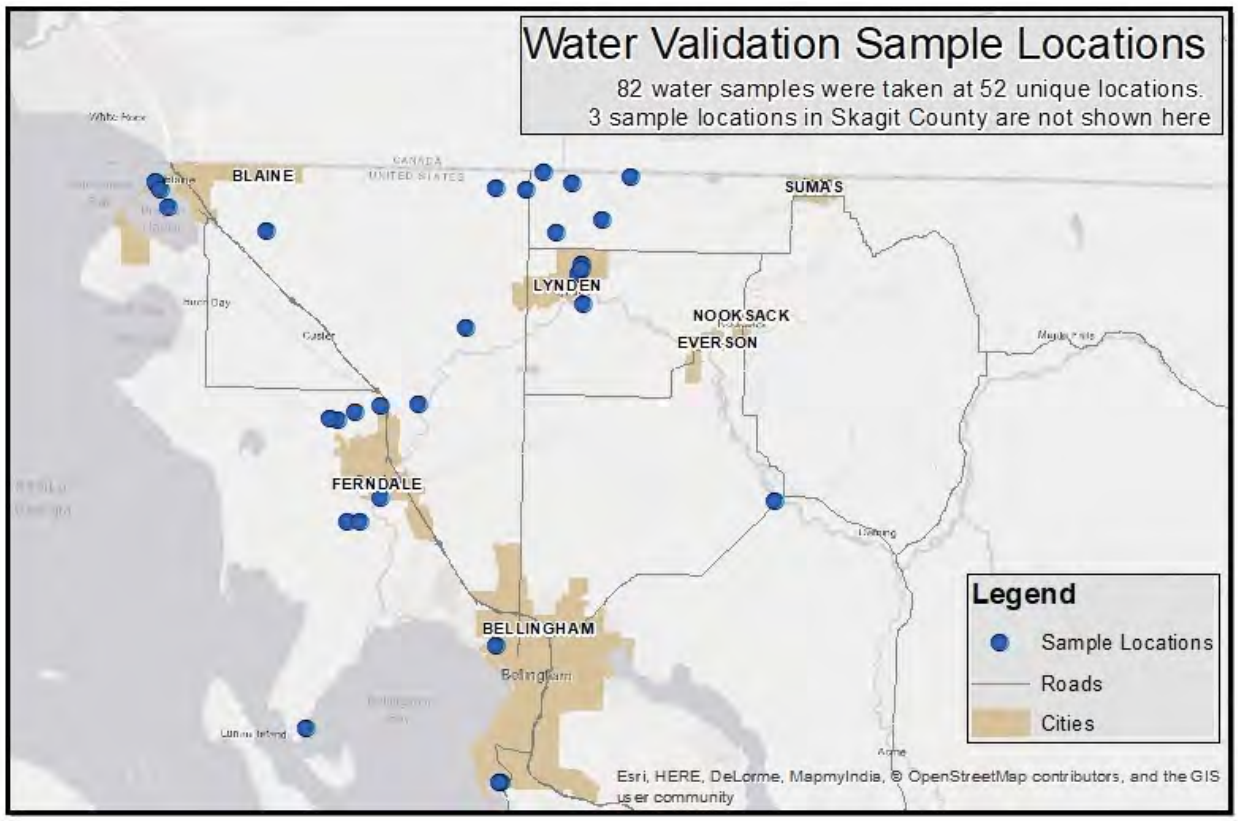

Figure 1. Map of distribution of water sample collection sites (blue dots) within Whatcom County, WA.

4.4.2. Quality Assurance Samples

Duplicate and Blank Samples

As outlined in the project QAPP, field duplicates (n=2) and field transfer blanks (n=3) were used to increase confidence in field collection methods. Field duplicates were collected and analyzed for fecal coliform (SM9222D) and DNA via 16S analysis. Field transfer blanks were analyzed solely for fecal coliform (SM9222D) to ensure samples had been collected without contamination.

Spike Samples
"Spiked" water samples (n=13) were created by adding known types and quantities of fecal matter to clean water samples as a positive quality control to monitor and validate the recovery and identification capability of the methods used in the project. Fecal spikes included fecal matter from individual species, some of which were also collected for, and included in, the fecal source reference catalog, and some that were not. Species used for water spikes included: alpaca, boat pumpout, coyote, dairy lagoon (2), dairy solids, deer, dog, duck, muskrat, roofbird, sheep, and swan. The spiked water samples were prepared by WCD using a ratio of 2L of tap water to 10 mg of fecal matter. Water samples were shaken vigorously for one minute to mix the water and fecal matter. Samples were then left to sit for two hours at room temperature. After two hours, any large pieces of debris were poured off the top of the sample. The samples were shaken vigorously for one minute and 100 ml of water was poured off for fecal coliform analysis. Tap water was added to the sample container to bring the volume back to 2L for subsequent filtering and 16S DNA sequencing by the lab.

4.4.3. Sample Custody and Documentation

Water sample custody and documentation was completed following the project QAPP, including sample naming structure, documentation of field notes and photographs, and data management. A field collection form was used to note sample location, sample data and time, and any field observations. Sample locations were mapped using ArcGIS (

Figure **1**). Laboratory reports and sample chain-of-custody forms were archived as outlined in the project QAPP.

4.4.4. Sample Receipt and Storage

At time of receipt, water samples were given a unique laboratory identifier which followed the sample through the entirety of the study. Samples were stored at 2-4°C for a maximum of 24 hours or immediately filtered. Temperatures of storage refrigerators were monitored and recorded twice daily. No water samples were frozen prior to processing for analysis.

4.4.5. Water Filtration

All labware and tubing used for water handling and filtration were sterilized with 10% bleach or flame, followed by rinsing with sterile ultrapure water until no residual chlorine could be detected. Water was pre-filtered by vacuum through 180 µm, 11 µm, and 2.5 µm filters to remove large diameter material. Pre-filtered water was first passed through 0.45 µm filters by vacuum, then through a 0.22 µm Sterivex™ filter via peristaltic pump. Turbid water samples that would not pass though pre-filtration were processed by centrifugation. The 0.45 µm and 0.22 µm filters were frozen at -20^{0}C until DNA extraction. During each filtration event, at least one blank was collected by aliquoting 1L of sterile ultrapure water in a sampling container and processing through pre-filtration filters, 0.45 µm filter, and a 0.22 µm Sterivex™ filter. The 0.45 µm filters and 0.22 µm Sterivex™ filters from each blank were logged, stored, DNA extracted, and analyzed in the same manner as the experimental water samples.

4.4.6. Water Sample Analysis Methods

DNA Extraction
DNA was extracted from filters (or centrifugation pellets) using the DNeasy PowerWater Sterivex™ Kit (Qiagen) with slight modifications to the manufacturer's protocol. Extracted DNA was quantified using the Qubit 3.0 fluorometer. All DNA extracted from filters was diluted to <5 ng/µL prior to PCR amplification.

16S DNA Sequencing and Data Analysis

PCR was performed with 16S rRNA gene primers targeting the variable V3 and V4 region of the gene producing approximately a 550-bp amplicon. Preparation, quantification and DNA sequencing was performed as previously described in *Deliverable 1, Sample Analysis Methods, 16S DNA sequencing protocol*.

All water sample sequence data were processed using the same workflow described in *Deliverable 1, 16S Sequence Data Analysis*. The SourceTracker software was used to estimate the proportion of shared microbial communities between sources and waters; all samples were rarefied to 10,000 reads. A custom script was used to create heatmaps for results. An alpha rarefaction curve was built using Qiime to determine if depth of DNA sequencing per sample was sufficient (data not shown).

5. Results and Discussion

5.1. DELIVERABLE 1 - Fecal Source Reference Catalog Development

Obtain fecal samples from a variety of sources and comprehensively characterize the microbial community of each fecal sample via whole sample DNA sequencing (WSS) utilizing a next generation DNA sequencing (NGS) platform.

5.1.1. Fecal Microbial Population Analysis

The variation in the microbial phyla of the fecal sources sampled was compared for each reference fecal source type (Figure 2). Actinobacteria, Firmicutes, Proteobacteria and Bacteroidetes were ubiquitously observed in varying abundances in all fecal sources analyzed. Higher abundances of Proteobacteria were observed in avian (excluding Chicken Comp) and both Nooksack Upper North fork and Nooksack Upper Mainstem, followed by Beef Old, Human Boat Pumpout, Human Septic, all Dairy (Lagoon, Solids, and Fresh), Goat Comp, and Deer Comp. A higher overall abundance for Firmicutes was seen in mammals followed by avian and human samples with lower abundances in both Nooksack River samples. Higher abundances of Bacteroidetes were observed in samples taken directly from mammals, except for Dog Comp, as well as in both Nooksack River samples. Highest abundance of Actinobacteria was observed in Goose Comp. Waterfowl (i.e. Duck Comp, Goose Comp, and Swan Comp) showed a higher presence of Cyanobacteria as compared to other fecal sources. Phyla represented in the data were approximately as expected for these fecal sources.

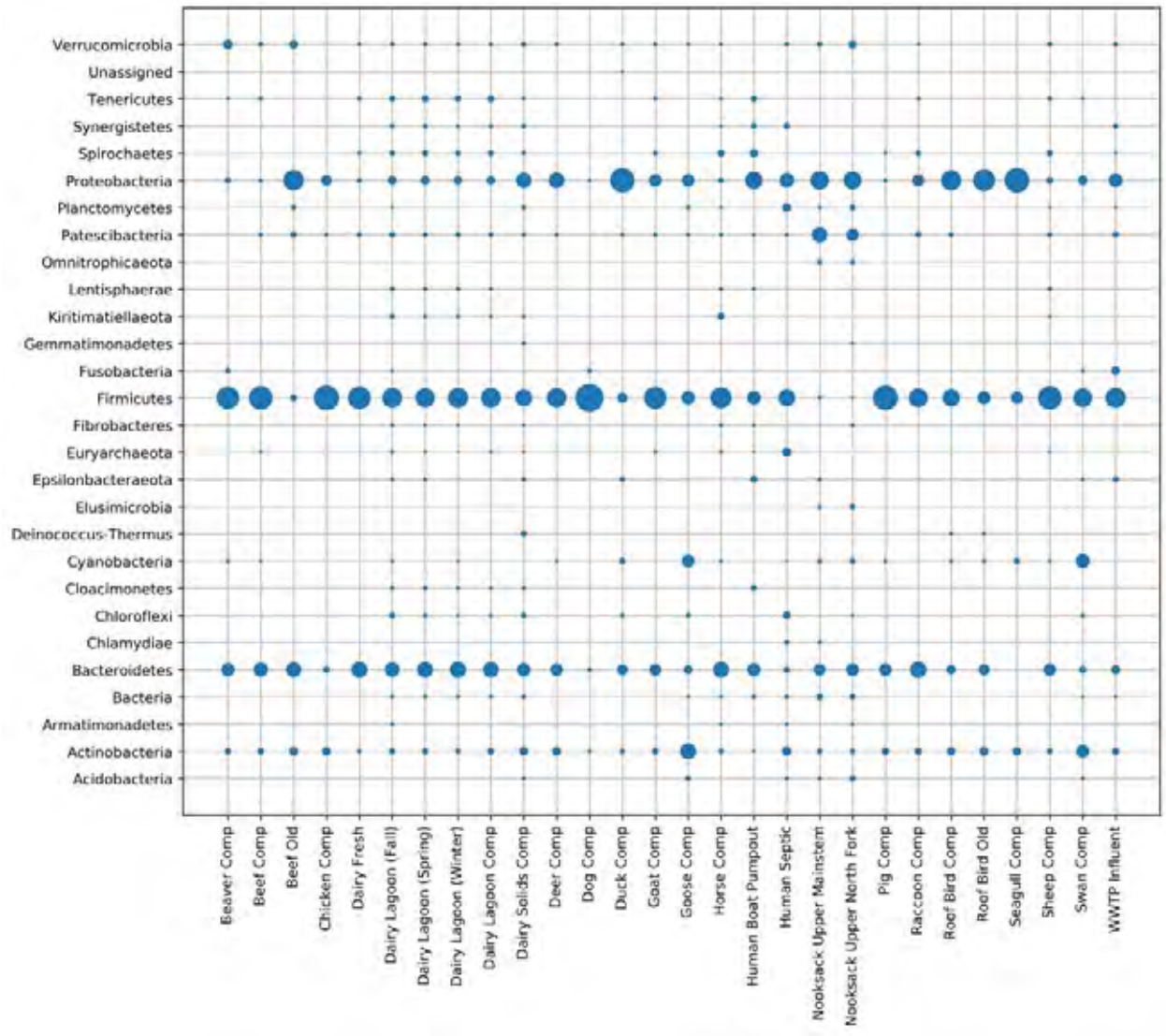

Figure 2. A bubble chart representing 26 phyla as well as Bacteria and Unassigned (y-axis) present in DNA sequence data from 28 fecal source types (x-axis) (Comp = composite sample). Each blue dot indicates the presence of a phylum in the fecal source where size of the dot represents the relative abundance of each phyla. Intercepts without dots represent no bacteria of that phyla were detected.

While each fecal type had a unique combination of phyla, four primary phyla were found to be present in all fecal source samples, Firmicutes, Proteobacteria, Bacteroidetes (one exception of Seagull), and Actinobacteria. This was expected as they are ubiquitous in the gut microbiome of animals (Lee et al. 2011). The abundance of different phyla types varied greatly between fecal sources.

Firmicutes was found in high abundance in all animal types except for Nooksack River samples, Beef Old, and a number of avian sources including Roof Bird, Seagull and Duck, which showed Firmicutes in low abundance. The latter observation is consistent with previous research that has shown that avian/waterfowl feces contain high abundance of Proteobacteria, and low abundances in other phyla (Unno et al. 2010).

Proteobacteria is a group of organisms that contain common fecal coliforms as well as many known pathogens, including bacteria in the genera Escherichia and Heliobacter (Shin et al. 2015). Some coliform

bacteria within this group are non-fecal derived coliforms such as Klebsiella, Enterobacter and Citrobacter. Although Proteobacteria is ubiquitous across all fecal source types sampled, the low abundance in some key fecal sources (i.e., Pig, Dog, Dairy Fresh, Beef, Horse, Sheep, Dairy Lagoon) is consistent with our hypothesis that a comprehensive analysis of the entire microbial community is inherently more discriminatory.

Bacteroidetes are found in almost every environment and are known to be inhabitants of fecal material. This group contains the commonly used Bacteroides species-specific biomarkers utilized for qPCR molecular source tracking methods. Bacteroidetes contains many host-specific bacteria and have been used to discern source types (Jeong et al. 2011).

These results demonstrate that each fecal source has a taxonomic microbial signature, and the similarities/differences between them can be used to compare sources to each other, as well as to water microbial community signatures. It is also important to note that, although these four main phyla play an important role in discerning fecal sources in water samples, there are many other phyla (n=22) that also contribute to the fecal source reference catalog.

5.1.2. Fecal Source Comparison

Principal coordinate analysis (PCoA) was used to explore the similarities and differences of fecal source microbial communities (Figure 3). PCoA was able to preserve a total of 48% of the variance between sources within the first two components, 34% in principal component 1 (PC1) and 14% in principal component 2 (PC2). Broad source groups were tested for significance using a PERMANOVA test. P-values for Dairy Lagoon (orange), Dairy Solids (green), Sheep (brown dashed), Horse (brown solid), Avian (red), Human/WWTP Influent (purple), and Nooksack (light blue) were all <0.01 indicating that their microbial signatures are distinct from each other. A seasonal signal was observed for Dairy Lagoon microbial community signatures, where Fall, Spring, and Winter seasons were found to be statistically different from each other. Significant differences were observed for Dairy Lagoon Comp samples between fall and spring (p = 0.008) and fall and winter (p = 0.001), showing that the fall season is distinguishable. Dairy Lagoon spring and winter composites were less differentiated from each other (p = 0.043). Dairy Fresh and Beef Comp were similar to the Sheep Comp, whereas Beef Old showed more similarity to Dairy Solids. Two Dairy Solids samples grouped more closely with Dairy Lagoon (Fall). One Horse sample was an outlier from the rest of the Horse samples. Chicken Comp and Swan Comp were outliers from the other avian sources. Other ungrouped animals (i.e., Beaver Comp, Deer Comp, Raccoon Comp, and Dog Comp (gray)) were dispersed, showing moderate similarity to various broader groups.

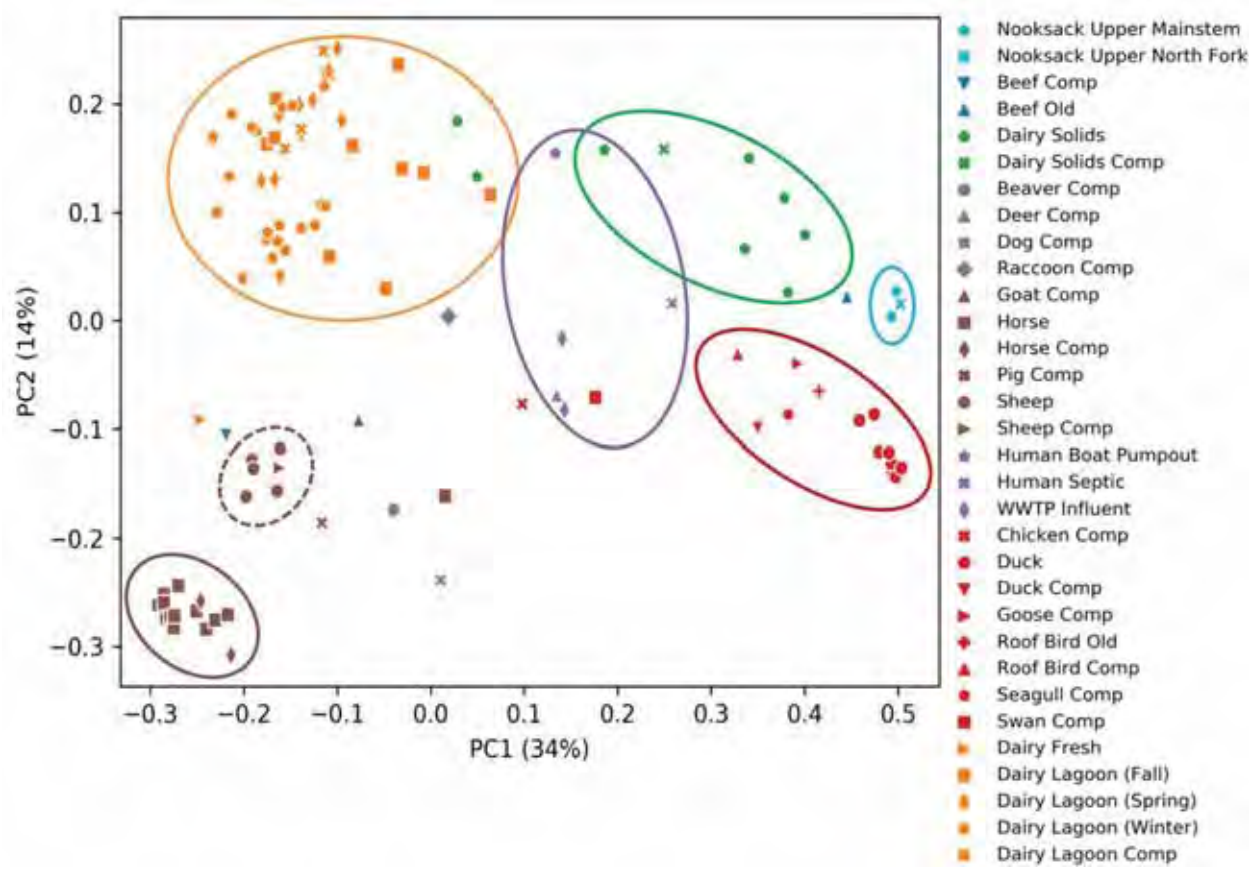

Figure 3. Principal coordinate analysis (PCoA) of weighted-unifrac distance for all fecal sources. X-axis is principal component 1 (PC1) and y-axis is principal component 2 (PC2). Circles indicate broad groups: Dairy Lagoon (orange), Dairy Solids (green), Sheep (brown dashed), Horse (brown solid), Avian (red), Nooksack (light blue). The PCoA is a visual depiction of the microbial community signature of each source relative to all other sources.

Results from the principal coordinate analysis (PCoA) (Figure 3) showed that many of the microbial communities from fecal sources were statistically distinct from each other (circled groups in Figure 3). This result indicates the utility of microbial community signatures for discriminating between fecal sources.

Beef Comp and Dairy Fresh grouped with the non-cattle ruminant livestock (Sheep) indicating the bacterial communities showed more similarity to fresh fecal material, rather than Dairy Lagoon or Dairy Solids, which have undergone biological and mechanical treatment and/or storage. This can also be seen with the Beef Old sample, which grouped more closely to the Dairy Solids suggesting that as Beef fecal samples age their microbial community becomes more similar to Dairy Solids and less like the fresh ruminant fecal microbial communities. In aggregate, these data indicate that microbial community signatures not only identify fecal source by species, but also highlights the effect that natural, chemical and/or biological process(es) can have on fecal material microbial communities after deposition.

Seasonal variation of microbial communities was investigated within the Dairy Lagoon fecal source type. Dairy Lagoon was selected for this investigation because we suspected that season would influence the microbial community. Material in dairy lagoons is subject to ambient temperature gradients between above and below ground storage and seasonal temperature fluctuation. It also experiences differing concentrations by season based on water content, holding time and stratification, manure deposition and removal throughout the year. Other exogenous inputs can include runoff from barn roofs, concrete

or asphalt, and natural surfaces exposed to wildlife, pets, or other animals. As expected, there were differences between seasons for Dairy Lagoons, which can be seen to some degree in Figure 3, where the Fall and Winter/Spring samples are somewhat separated (Lovanh et al. 2009). This was supported by statistical analysis, which indicates that the Fall and the Winter/Spring samples are different. It was not surprising that there was less significance between Winter and Spring, as they were collected closer in time to each other (early March and late April, respectively). These results highlighted the importance of seasonal composite sample collection and/or seasonal fecal references for complex source types with strong seasonal influence such as dairy lagoons. It should also be noted that, although there was statistical support discriminating Dairy Lagoon from Dairy Solids, there is a slight overlap between the two source types, specifically with Dairy Lagoon (Fall). This is likely because not all dairy lagoons sampled practiced solid separation as part of on-farm manure management; 50% of farms sampled had all manure solids deposited into the lagoon. Deeper analysis of individual dairy lagoon samples revealed no obvious and consistent grouping of individual samples by management type (i.e., separated vs non-separated solids, lagoon agitation, bedding type, manure handling, cow breed type). Therefore, we can conclude that the composite dairy lagoon samples are representative of any individual dairy lagoon over a wide range of management types.

While Human microbial signatures had low sample numbers (WWTP=2, Septic=1, Boat Pumpout=1), this broader group showed statistical significance as being distinct from other non-human sources. This supports the need for further sampling and characterization of these sample types to allow for distinction between types of human fecal sources. Human Boat Pumpout (septic) grouped between the Dairy Lagoon and Dairy Solids, suggesting that these samples have commonality. Human Boat Pumpout grouped away from WWTP influent, indicating that the two source types, although they share the commonality of human fecal sources, have different microbial signatures due to factors such as processing or source complexity.

Microbial similarity, and thus grouping, of avian species fecal signatures was expected, as was their distinction from other fecal source types due to their differing physiology. Chicken Comp and Swan Comp grouped away from other avian sources. Chicken differs from other avian sources in that it is a domesticated bird species. Differences in chicken diet and habitat could be contributing to its separation for other avian species. The separation of swan from other avian spices is unknown, particularly because of the overlap observed between the Swan Comp and Goose Comp (Figure 4) Statistical testing could not be employed on these outliers (chicken, swan) due to small sample size. Additional fecal collection should be conducted to expand the composites and verify the difference from other avian species. In particular, due to the relatively small sample size for Chicken (two farms representing nine individuals), additional fecal sampling should be conducted to increase the number of farms and individuals in this sample to account for variability across influencing factors such as farm management, diet, and/or breed.

To determine if compositing of individual fecal samples by source was representative of the individuals, and to determine the amount of variability of individuals within a composite, a number of species (Horse, Duck, Dairy Lagoon, Sheep) were chosen. In most groups, the composites and individuals grouped closely with each other (Horse, Sheep, Dairy Lagoon), and in the case that statistical testing was possible (Dairy Lagoon), the composites were not statistically different from the individuals. Other composited samples that had paired individuals were unable to be statistically tested and therefore warrant further investigation into the compositing methods for different fecal types.

The composite method appears to preserve the discriminating power between the broader fecal source types. Future sampling and investigation of fecal sources is necessary to strengthen the distinction between sources types. There are clear differences between “fresh” fecal samples and samples which

undergo post-processing. This area requires more sampling and characterization for fecal types that may have prolonged ambient exposure prior to entry into a sampled waterbody or location. Overall, the fecal source reference catalog contains enough information using microbial community signatures to statistically discriminate between select sources, Human (WWTP, Septic and Boat Pumpout), Dairy Lagoon, Dairy Solids, Duck, and Horse.

5.1.3. Comparison of Fecal Source Microbial Communities

Differences between the shared microbial communities of fecal sources can be observed in Figure 4 highlighting the estimated percent overlap between sources using a Leave-one-out (LOO) analysis. This LOO analysis was used to determine the overlap of the microbial community of a source with all other sources, effectively assessing the "uniqueness" of a source group; for example, "if Beaver Comp is not in the fecal source reference catalog, what sources would beaver most look like?". Uniqueness of sources, shown by the Unknown column, ranged from 19% (Dog Comp) to 96% (Beaver Comp, Beef Old). Sources with >70% Unknown microbial community signatures include Beaver Comp, Beef Old, Horse Comp, Human Septic, Nooksack Upper Mainstem, Nooksack Upper North Fork, Pig Comp, Raccoon Comp, Sheep Comp, and WWTP influent. Dairy Lagoons of different seasons show overlaps ranging from 10-24%. Seasonal Dairy Lagoon samples overlap with Dairy Lagoon Comp (all seasons together) from 9-21%. Overall, no Dairy Lagoon overlapped more than 7% with any other source type. Avian sources share a higher overlap (>25%) with other avian sources including: Chicken Comp, Duck Comp, Goose Comp, Roof Bird Old, Roof Bird Comp, Seagull Comp, and Swan Comp. Sheep Comp and Goat Comp overlapped by 24%, with no other overlap >10% with any other fecal type. Overlaps of >10% across broader source groups include: Human Boat Pumpout with Dairy Lagoon Fall/Spring (12%, 20%), Dog Comp and Raccoon Comp (20%), Dog Comp and Horse Comp (42%), Dog Comp and WWTP Influent (11%).

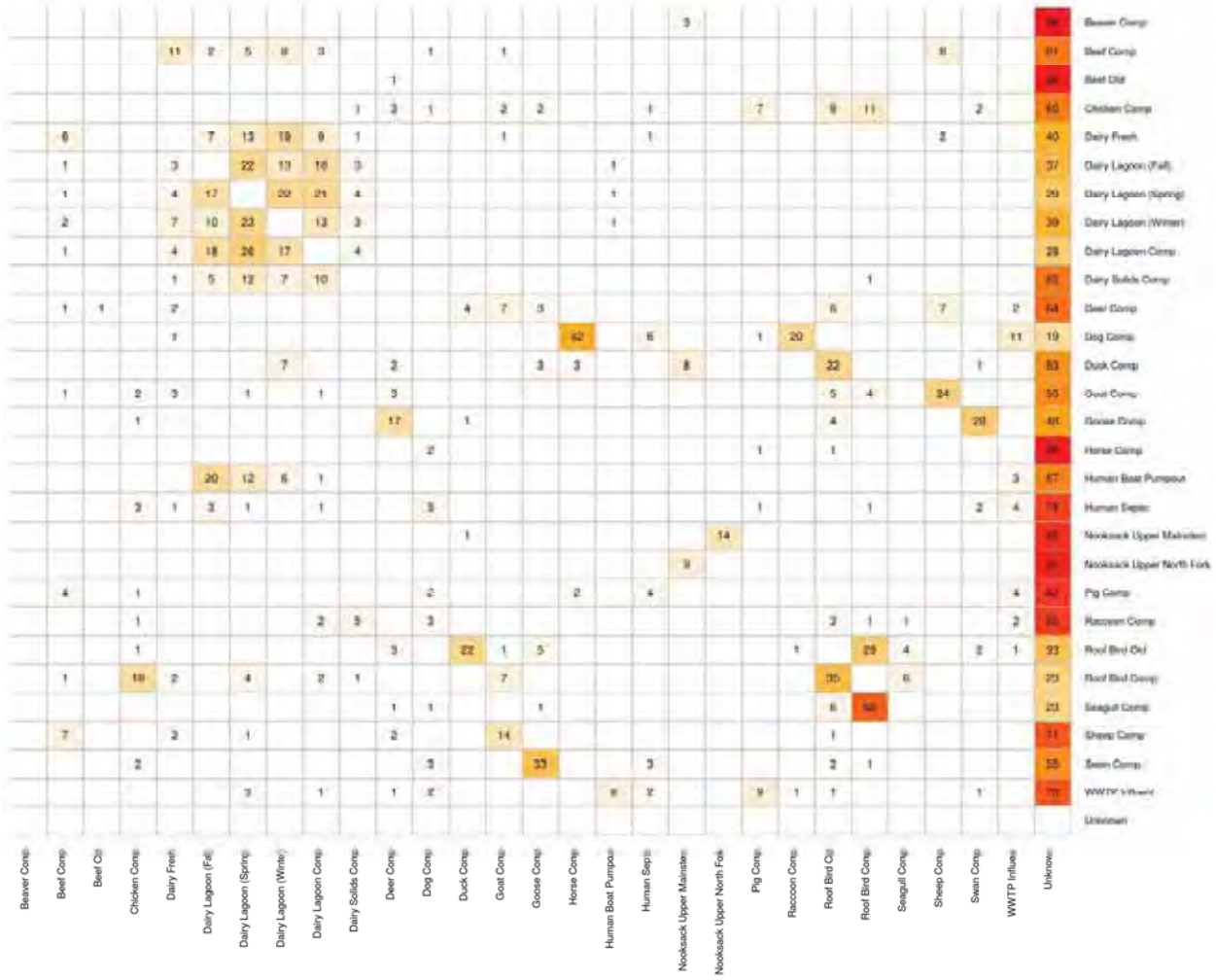

Figure 4. Leave-one-out (LOO) analysis comparison of sources. Intensity of color on a scale of yellow to red indicates the estimated percent overlap of microbial communities shared between sources. The value presented within each cell represents an estimated percentage of overlap between the fecal sources' microbial communities when compared to each other, i.e. the sources are not compared to themselves. An empty cell represents an overlap of <1%.

Establishing the extent of overlap between different fecal source microbial community signatures was crucial for drawing meaningful conclusions from comparisons of fecal sources to environmental water samples, as was the case of this study. Results from the LOO analysis (Figure 4) suggest that, in most cases, microbial community signatures from different environments do not overlap with each other. The uniqueness of a source within the fecal source reference catalog can be inferred from the percent of the microbial community classified as "Unknown". A higher Unknown proportion suggests that, when the fecal source is removed from the catalog, there are no other sources in the catalog which resemble the removed source. Samples which shared >10% of their microbial community signature with another source can be thought of as a "high" overlap for the purposes of this discussion. A high overlap suggests similarity of microbial communities between fecal sources, and thus low differentiation.

The Nooksack Upper Mainstem and Nooksack Upper North Fork river samples were taken to determine the underlying microbial community signature of water prior to introduction of potential human, pet, and livestock fecal sources; this does not include wildlife that may be impacting the watershed naturally.

Nooksack sources did not show greater than 1% overlap with other sources, which strongly supports that most of the microbial community in environmental water samples is distinct and contains no overlap with fecal sources of interest. This is an important result because if Nooksack sources showed overlap with animal sources, the confidence level of source identification in experimental water samples would be difficult to assess.

Dairy Lagoon sources showed high (>10%) overlap only with other Dairy Lagoon samples, supporting previously reviewed results (Figure 3) that this is a well resolved and distinct group. To a lesser extent (<7%), there was some overlap of Dairy Lagoon with other Dairy sources (Solids, Fresh), which is not unexpected as they are related in their fecal source type (dairy). Exceptional sources, with higher (>10%) overlap with other non-alike sources, were observed including Avian and Dog Comp. There was a slight overlap between Beef Comp and Human Boat Pumpout, suggesting that there is both an animal source influence (difference) as well as a post-processing (similarity) influence on the microbial community signatures (see discussion of PCoA).

While in most cases avian fecal sources overlapped, there were also overlaps with other fecal types such as Goat Comp and Deer Comp. This result underlines the importance of further investigation specifically within avian source groups. Also included in the avian group are the Roof Bird sources, which are somewhat of a catch-all for many bird species that might spend time on the roofs of buildings (i.e., starling, pigeon, black bird, etc.). Furthermore, the high overlap of many avian fecal source types is evidence that avian fecal sources are particularly difficult to differentiate from each other and may require additional sample numbers, and/or perhaps a different sample technique. The latter is particularly relevant to Roof Bird, as it was very difficult to obtain a “fresh” bird sample due to the environment in which it was collected (barns) and the inability to catch birds for samples. Additional research should be conducted on improved methodologies for compositing naturally mixed sample types such as Roof Bird.

Others have observed challenges discriminating dog fecal samples (Brown et al. 2017) and this result was supported in our study; Dog Comp had high overlap (>10%) with several non-canine sources (Horse, Raccoon, WWTP). This is concerning, as dog fecal matter is an emerging priority for water quality monitoring in Whatcom County and elsewhere. However, only one composite, not individuals, was sequenced which made it difficult to rigorously investigate this sample type’s variability between individuals. Additionally, dogs are known to ingest the fecal matter of other animal species, perhaps adding to the complexity of their own fecal microbial community signature. The need for a more in-depth investigation, including additional numbers, of dog fecal samples is indicated by these data and observations.

In Figure 3, it was shown that Human Boat Pumpout grouped close to Dairy Lagoon and Dairy Solids. This relationship was also seen in Figure 4 with Human Boat Pumpout overlapping (up to 20%) with Dairy Lagoon. This suggests that the microbial community of Dairy Lagoon and Human Boat Pumpout have a certain degree of similarity, which can have impacts on discernibility between the two source types.

Future efforts to resolve confounding results with sources such as Dog and Human Boat Pumpout, as well as other less sampled sources, should be undertaken in order to more confidently distinguish between these sources. Overall, the results from the Figure 3 add to the support that there are fecal sources which have distinct and discernable microbial community signatures and others which need further investigation before they can be confidently established as distinct fecal source reference catalog signatures.

5.1.4. Wildlife Tracker App

A Wildlife Tracker App was created using ArcGIS and Survey123 to help us locate wildlife species for fecal sample collection. This app enabled the public/landowners/partners to provide a location, date, species, number of animals, notes, and photo of animal(s) seen. This tool allowed us to locate wildlife for fecal and water sampling, as well as enable public viewing of patterns and locations of wildlife across the County for a better understanding of wildlife annual migration, concentration, and potential impacts to water quality. Local water quality partners have embraced the tool and partners in Skagit County and British Columbia are also using it to track wildlife.

- Wildlife Tracker App: https://arcg.is/14nWLj
- Wildlife Tracker Map Results: https://www.arcgis.com/apps/webappviewer/index.html?id=941f5a0847bd4d1493a6b2a7740e4df0

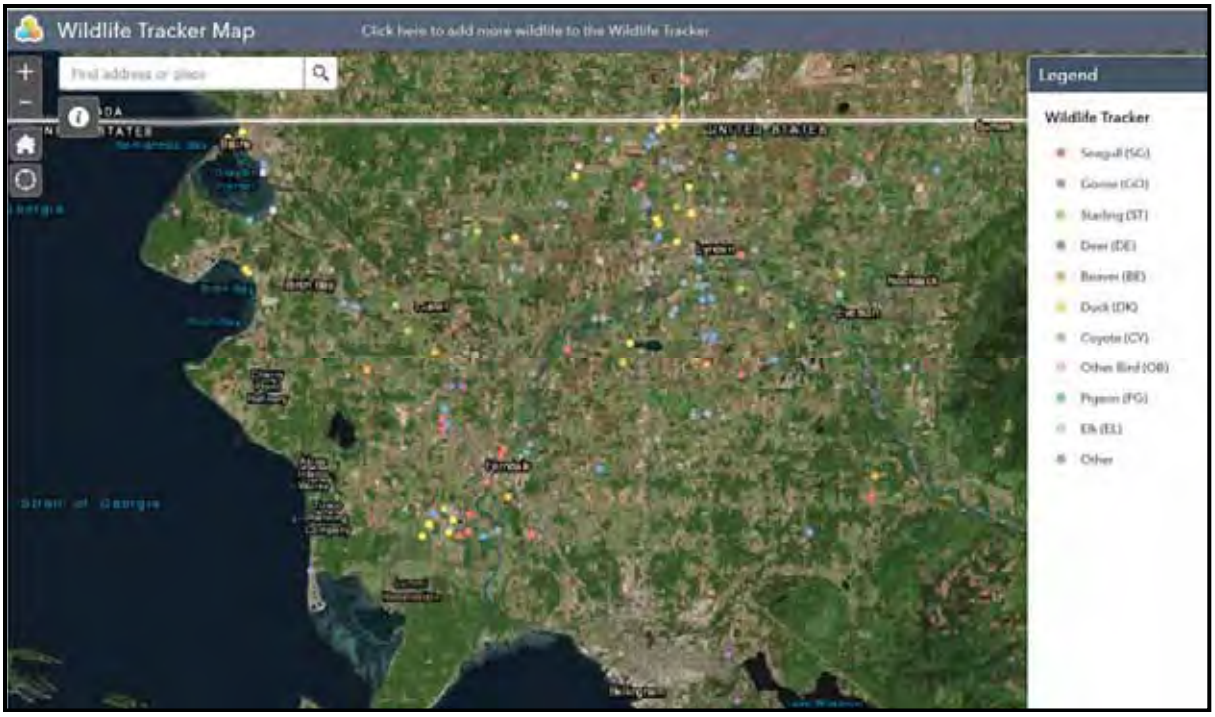

Figure 5. Screenshot of Wildlife Tracker App with identified sightings of wildlife (screenshot taken May 2019).

5.1.5. Deliverable 1 Conclusions and Future Work

The primary goal of this study was to build a fecal source reference catalog comprised of select fecal sources of interest for Whatcom County. These sources were chosen due to their regional specificity and their wide range of potential inclusion in local watersheds. Comparison of the microbial community signatures revealed that most source types are distinctive from one another, supporting the use of these signatures for discriminating fecal sources.

Overlaps observed in source microbial signatures indicate that minimally sampled source types require further collection efforts to increase discriminatory power. Understanding what drives these overlaps is critical to distinguishing sources. Fecal sources that showed considerable similarities to other sources were Dog, Roof Bird, Swan, Goose, and Human Boat Pumpout. These sources are of great importance to on-going monitoring efforts to mitigate impacted waters in Whatcom County. Increasing the sample

number for each of these sources will allow for more in-depth understanding of the shared and unique properties of their microbial community signatures. This will in turn lead to better methodologies for accurately monitoring their impacts in the environment.

A specific area of interest for future work is focused on critically assessing the interface of fecal sources with the environment. The Dairy Lagoon microbial community signature after application to a field was not explored in this study, but presents an available model to investigate this interface. Dairy lagoons are primarily anaerobic, and the microbial community reflects that. After field application, the microbial community would be expected to shift with exposure to aerobic conditions. If the microbial community signature of "applied manure" is distinct from sources within the current fecal source reference catalog (Dairy Lagoon, Dairy Solids, etc.), then identifying the origin of a runoff event from a field with applied manure could be challenging. It is unlikely that such an event would be missed completely due to the breadth of the current source types in the catalog; however, discrimination of sources benefits greatly from comprehensive inclusion of source-specific signatures. Establishing the effects that the time between field application and runoff events, dairy lagoon field application methods, and ambient exposure have on microbial community signatures is an active area of work.

This and other future projects are meant to increase coverage of under-sampled sources and extend our understanding of the microbial communities that are distributed across environments. In doing so, the sources in the fecal source reference catalog will be expanded and upgraded continually, enhancing its utility for water monitoring projects.

5.2. DELIVERABLE 2 - Fecal Source Reference Sample Collection SOP

Establish a standard operating procedure (SOP) for collection and handling of fecal source reference samples and share widely for better coordination and consistency of fecal source reference sample collection and characterization.

5.2.1. Fecal Collection Standard Operating Procedure (SOP)

Note: At the time this report was finalized the fecal collection standard operating procedure (SOP) document was still under review pending final study results. It will be made available once completed.

A fecal collection standard operating procedure (SOP) was created using the results of the project analysis, as well as field observation. The development of the fecal collection SOP encouraged a variety of methods assessment procedures including an exploration of sample hold time, homogenization, and fresh versus frozen sample comparison. Results of these "special studies" are included below.

One notable result from this study that affects the SOP was data supporting revision of the number of samples collected for each fecal type. The project QAPP established collection methods and sample number based on best available guidance. Data from this study showed that for fecal types with a large amount of variability, notably with diet, require a larger sample size than originally planned to account for the high degree of variability. Additionally, to obtain statistical confirmation of sample numbers, we should have allocated a larger portion of time and funding towards analysis of all individual samples within a composite. This would have provided more information on the range of variability on individuals and increased the confidence of a signature as unique from others. Additional work is needed in this area to improve the sample number collection section of the SOP.

5.2.2. Hold-time and Fresh/Frozen Study

Three fecal sample types, horse, duck and dairy lagoon, were assessed for stability of microbial community signature across varying sample hold times and conditions (Figure 6). Hold times of 0 days, 3 days and 30 days (frozen for 30 days, then thawed) showed no significant differences from each other using a pairwise PERMANOVA (n=3; $p = 0.914$). The Firmicutes:Bacteroidetes ratio was tested for statistical difference between fresh and frozen sample matrices. A paired T-test was performed for each unfrozen and frozen sample (n=6) and showed that there is no difference, $p < 0.01$, between the Firmicutes:Bacteroidetes ratio for fresh and frozen samples ($0.02 < p < 0.99$) for any of the source types.

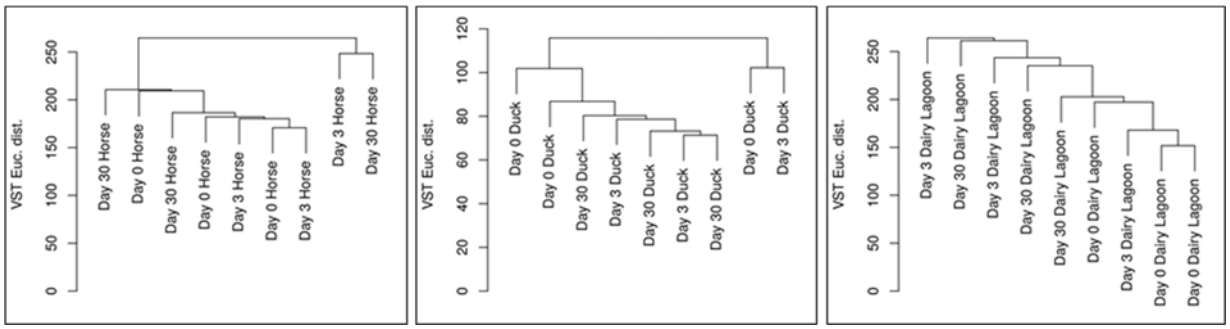

Figure 6. Hierarchical clustering analysis (HCA) dendrograms showing similarity for three fecal sources (Horse, Duck, Dairy Lagoon) across three hold-times, Day 0, 3, and 30. Day 0 and 3 were fresh, refrigerated samples. Day 30 was frozen for 30 days then thawed before analysis. Y-axis of each dendrogram equals Variance Stabilized Euclidean distance.

In order to establish a proper fecal sampling procedure, it was necessary to establish a hold-time protocol. It was unknown whether hold-time or freeze-thaw would significantly alter the microbial community signature of source types. This is crucial due to the nature of field collections and the lag-time that can occur during sample processing. Duck, Horse, and Dairy Lagoon samples were selected for this experiment due to the differences between the source animal physiology (post animal processing for Dairy Lagoon) as well as the distinctness of their microbial community signatures (Figure 3). For the three selected source types, there was no apparent effect on quality or stability of the microbial community signature between hold times of 0, 3, or 30 days. This study also showed there was no difference between hold times of day 0 and 3 when refrigerated, and day 30 which underwent a single freeze-thaw cycle. Therefore, collected fecal samples can be held for a up to 3 days in the refrigerator at 4°C, and/or can undergo a single freeze-thaw cycle after 30 day storage at approximately –20 °C, with no significant impact to the overall microbial community. A previous study had indicated that there could be a shift in the Firmicutes:Bacteroidetes ratio after a single freeze-thaw cycle (Bahl et al. 2012), however this was not observed in the current study with any of the source types. In this study, we deliberately did not use samples that had undergone multiple freeze-thaw cycles. However, since multiple freeze-thaw cycles may be unavoidable for experimental or practical reasons, the impact of repeat freeze-thaw cycles should be investigated.

5.2.3. Homogenization Study

Sample homogenization method was assessed for its ability to sufficiently mix and sub-sample composited source samples (data not shown). Kruskal-wallis pairwise testing showed no statistical difference ($p < 0.01$) between individual and composited (homogenized) sources for the two tested sources, Liquid Manure ($p = 0.46$) and Solid Manure ($p = 0.56$).

The results from the fecal sample homogenization method assessment showed that this laboratory method sufficiently mixed liquid/slurry and solid fecal composited samples to allow for representative

sub-sampling. This technique was applied to all composited source samples in this study. In the future, this test can be used to assess compositing methods for other fecal source sample types that may have different physical properties than the ones tested here.

5.2.4. Deliverable 2 Conclusions and Future Work

The creation of a standard operating procedure (SOP) for the collection, handling, storage and homogenization of fecal material was a valuable part of this study. This fecal source collection SOP will be available for future agencies and other parties to use for similar studies. The SOP will help guide field collection in a way that encourages cross-study comparisons and collaborations between researchers looking to improve on the quality of our environment.

A major question that came out of the development of the SOP was the effect of freeze/thaw cycles on the microbial community signature. Fecal samples were collected opportunistically (especially fecal samples of wildlife or other hard to collect sample types), and storage was required until enough could be collected to build a composite sample. Once they were provided to the laboratory for compositing, only a single freeze-thaw cycle occurred, although this may not always be the case. In the event of collaboration with other researchers, or if samples should be sent to another laboratory for analysis, the potential effects of multiple freeze-thaw cycles on the microbial community signature should be determined.

5.3. DELIVERABLE 3 - Fecal Coliform Plate DNA Characterization

Characterize the microbial community of colonies grown on standard fecal coliform membrane filtration culture plates, and compare to the fecal source reference catalog.

Assessment of the phyla present and relative abundance was performed on fecal sources (Alpaca, Dairy Lagoon (1 and 2), Dog, Duck, and Horse) sequenced from both a spiked water sample and colonies cultured on fecal coliform plates (“mFC plate”) (Figure 7). Three Phyla were shared between spike waters and their respective mFC plates: Proteobacteria, Firmicutes, and Actinobacteria. Proteobacteria made up the majority of identifications for the spike mFCs, whereas the spike waters had a more diverse set of phyla represented.

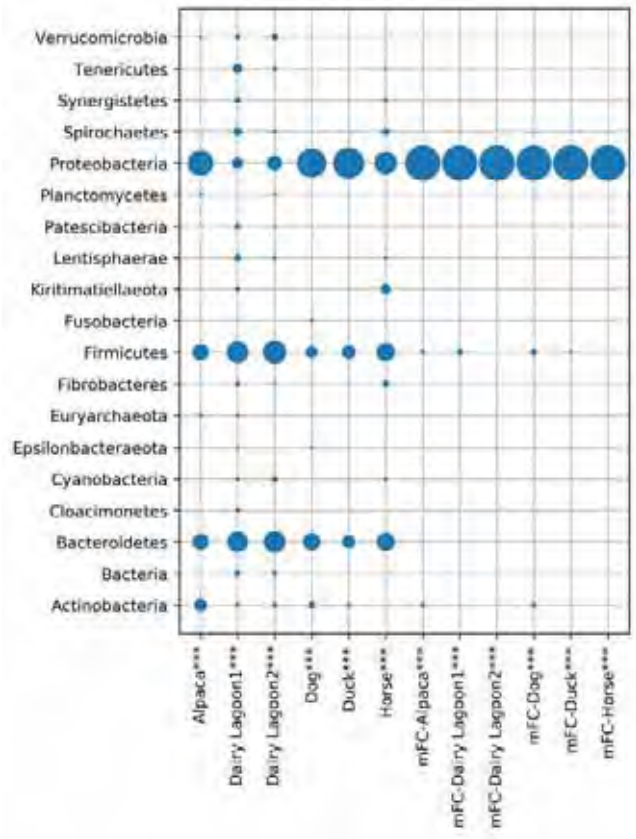

Figure 7. Bubble chart representing the abundance of phyla (y-axis) of spiked water samples compared to their respective fecal coliform plates, denoted by the prefix "mFC-" (x-axis). Each blue dot indicates the presence of a phylum and the size of the dot represents relative abundance. Intercepts without dots indicate that no bacteria from that phylum were detected.

Hierarchical clustering analysis (HCA) was performed for only the Proteobacterial taxa found in mFC plates and water samples (spiked, suspected, and unknown) analyzed (Figure 8). HCA showed observable groupings separating most mFCs plates from the different types of water samples. All mFC plates grouped together regardless of source (Figure 8 - Group 1). Spike waters and environmental waters grouped more closely with alike sample types, with the exception of Water-Spike-Dog, Water-Spike-Coyote, and both CA14ab-November samples (Figure 8 - Group 2). Nooksack Upper Mainstem water samples grouped farthest from mFCs and away from most of the environmental waters, showing a distinction from the water spikes as well.

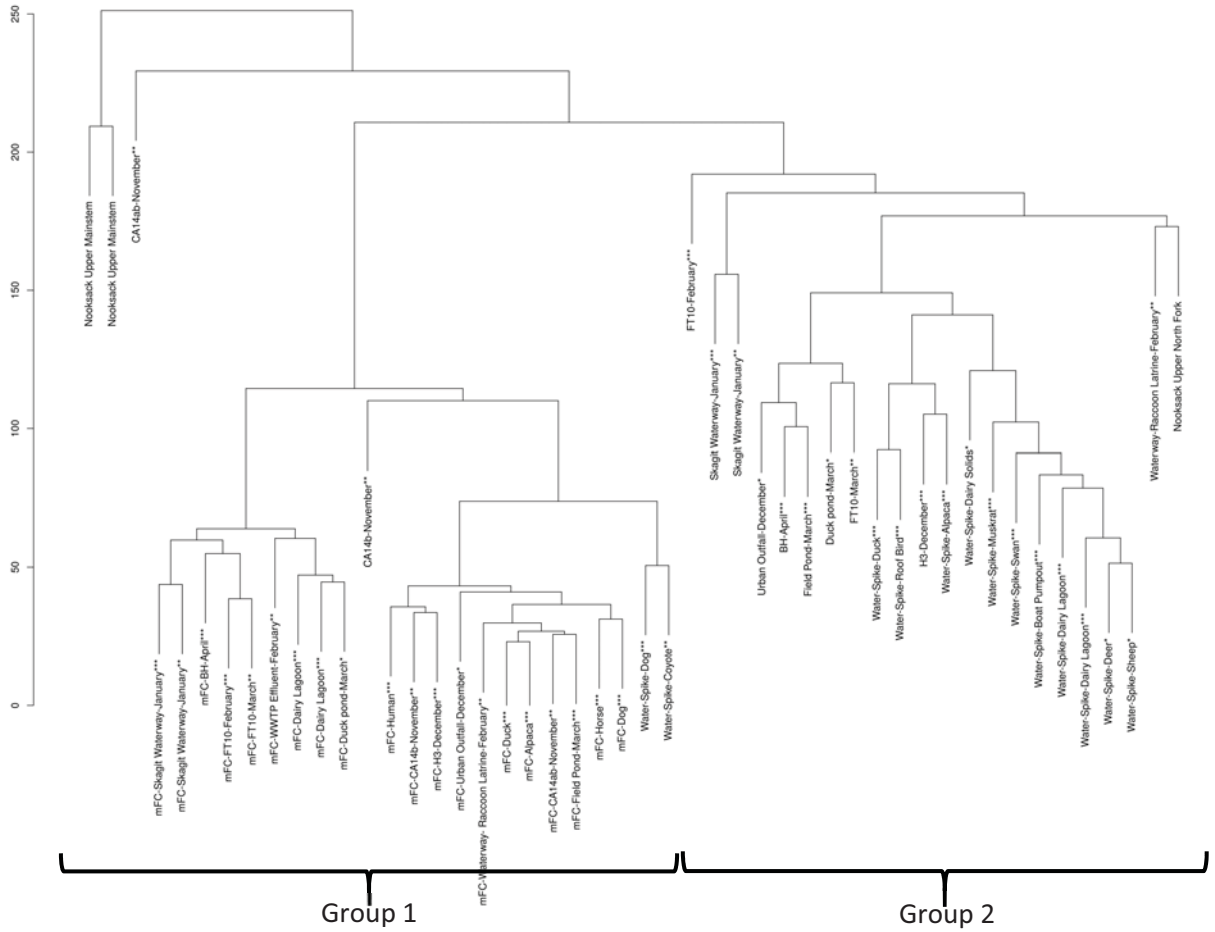

Figure 8. Hierarchical clustering analysis (HCA) dendrogram showing results for only the Proteobacterial taxa for fecal coliform (mFC) plates (Dairy Lagoon, Human, environmental water, and spike water) against environmental water, and spike water. Brackets show groupings (1 and 2) that are of interest. Distance metric used was variance stabilized Euclidean distance (y-axis). The * after each sample ID indicates the membrane filtration fecal coliform (mFC) count (CFU/100mL) for the water sample using the following range: "*" (0-200), "**" (200-1000), and "***" (>1000).

Results from a SourceTracker analysis (data not shown) of mFC spike waters against sources showed mFCs had no overlap with associated sources, with the exception of a >5% overlap for mFC-Dairy Lagoon with Dairy Lagoon (Winter) and a >1% overlap of mFC-WWTP Influent-December with WWTP Influent. The mFC-Dog showed >1% overlap with Roof Bird Comp. All plates, except for mFC-WWTP Effluent-February showed >1% overlap with Dairy Solids and all mFCs overlapped >5% with Deer Comp. HCA showed no discernable patterns for samples with respect to mFC spikes and distinct environmental water samples (data not shown).

Membrane filtration fecal coliform (mFC) plate counts are one of the standard methods used to indicate if waters (especially freshwaters) are impacted by bacterial contaminants. Currently, fecal coliforms are used as an indicator of fecal contamination and potential presence of human pathogens. Escherichia coli is another common bacterial indicator, serving much the same purpose as the broader fecal coliform group. Traditionally, the genera that are expected to grow on fecal coliform plates are: Escherichia (fecal-associated), Enterobacter, Klebsiella, and Citrobacter (non-fecal associated), all belonging to the

phylum Proteobacteria. While including organisms known to inhabit warm-blooded animal gastrointestinal tracts, the fecal coliform group excludes a wide range of fecal-associated microbes that may not be as easily cultured or would require different media or temperature for culturing (e.g. Firmicutes, Bacteroidetes, and Actinobacteria). Additionally, mFC plate media are not exclusive to fecal microbes and therefore may culture organisms other than those originating from fecal sources. To explore what other types of bacteria will grow on a mFC plate, and how these cultured representatives compare to a sample's total microbial community, both environmental waters and fecal sources were cultured and investigated. This study also evaluated whether the mFC method was able to correctly identify fecal sources in spiked water samples or if the inherent culturing bias of the mFC method would lead to an inability to discriminate between sources.

As expected, the majority of bacteria that grew on the mFC plates were in the Phylum Proteobacteria, with trace Firmicutes and Actinobacteria growth. This supports that the mFC plates select for a subset of the total microbial community. In contrast, the paired water samples contained several other phyla in addition to Proteobacteria, including high abundances of Bacteroidetes and Firmicutes (Figure 7).

When comparing fecal reference sources to mFC spiked water samples using SourceTracker, there was very little overlap between mFC spike water samples and their corresponding sources (data not shown). The few overlaps that did occur, with dairy lagoon and WWTP, would need to be investigated further to determine the validity of the overlap. Additionally, all spike water samples, including dairy lagoon and WWTP spikes showed a majority (>25%) overlap with Deer Comp and Dairy solids. This overwhelming overlap with non-distinct sources indicates that the culture-based mFC method greatly reduces the microbial community creating an inability to discriminate between sources. Figure 7 shows this reduction of bacterial community and clearly displays the lack of overlap between mFC plate microbial communities of water samples and the microbial communities of whole water samples via 16S.

Utilizing a Hierarchical clustering analysis (HCA) Dendrogram (Figure 8), two distinct groups were observed when comparing mFCs from water samples, and corresponding waters. The clear distinction between Groups 1 and 2 is not surprising due to the severe reduction in represented bacteria. This finding also agrees with results from a recent study that questioned the future usefulness of DNA characterization of mFC plates as a tool to detect fecal source inclusion in water samples when compared to the fecal source reference catalog entries built for this study (Leight et al. 2018).

A full assessment of bacteria present on mFC plates analyzed was also performed using a bubble chart analysis to show the presence and relative abundance of each genera cultured (Figure 9). There were four genera that were shared by more than 10 samples: Pseudomonas (n=21), Dickeya (n=24), Comamonas (n=14), and Acinetobacter (n=20). All of these were found in waters and at least one fecal source. All other genera were exclusive to either sources or waters, with the exception of Staphylococcus (found in low abundances in Alpaca and Drayton 8 –March) and Novosphingobium (found in low abundance in Alpaca and Field Pond – March).

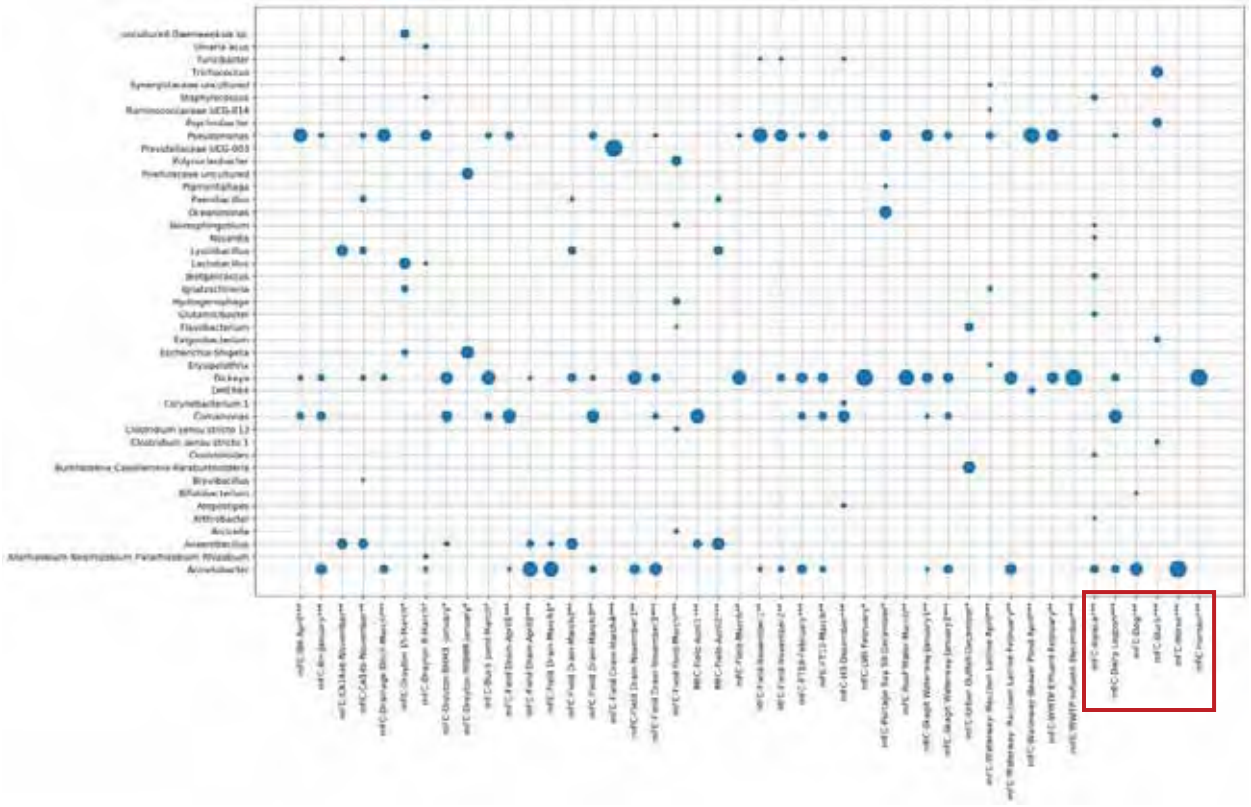

Figure 9. Bubble chart representing abundance of genera (y-axis) cultured on fecal coliform plates (mFCs) from environmental water samples and spiked water samples (x-axis). Red box indicates spiked water sample mFCs. Each blue dot indicates the presence of a genera and the size of the dot represents relative abundance. Intercepts without dots indicate that no bacteria from that genera were detected. The * after each sample ID indicates the membrane filtration fecal coliform (mFC) count (CFU/100mL) for the water sample using the following range: "*" (0-200), "**" (200-1000), and "***" (>1000).

The four shared genera (Pseudomonas, Dickeya, Comamonas, and Acinetobacter) found in the majority of water samples analyzed via mFC, varied in abundances (Figure 9). Acinetobacter is a ubiquitous bacterial genus, common to many environments including: soil, water, horse, dog, cattle, and human (Costa et al. 2012, van der Kolk et al. 2019). Comamonas, abundant in the Dairy Lagoon mFC plate, has been observed in swine lagoons, cow manure anaerobic digestion systems, as well as in soil environments (Xin et al. 2014, Macauley et al. 2007, Ntougias et al. 2004). Pseudomonas is a diverse genus of bacteria found in many environments such as soil, water, animal/plant hosts (Dogan et al. 2003). Dickeya is best known as a genus containing phytopathogens for potatoes and other agriculturally important crops (Toth et al. 2015, Glasner et al. 2011). Although, unrelated to the specific goals of this study, these results provide a possible utility for fecal plates in phytopathogen detection for agricultural environmental samples and warrants future study. For example, Dickeya species are known to infect crops; therefore, testing for phytopathogenic Dickeya species in soil or irrigation water where manure is used may help to prevent crop infection or pathogen transmission on plants (Sledz et al. 2017). It should be noted, that none of these genera are distinctive to a specific environment in this study and therefore are not considered to be useful for discriminating sources.

5.3.1. Deliverable 3 Conclusions and Future Work

Due to their wide usage for water quality management, mFC agar plates were included in this study. Initial results showed that in most cases they lacked the ability to discriminate the spike samples when

compared to fecal sources. However, there were several results that prompt further inquiry. It was observed that both sources and waters had a mix of shared and unique microbial taxa present on their mFC plates. This begs the question of whether any of these taxa could be used as indicators for sources or perhaps for events such as flushing during high precipitation. In order to determine the viability of these identified taxa as indicators, replicates from sources and from waters must be collected and analyzed. By carefully collecting multiple samples from site locations across multiple seasons, the degree of fidelity for taxa growing on plates can be estimated. It is possible that with enough data, plates can serve as the initial "scouting assay" when looking for possible contamination events.

5.4. DELIVERABLE 4 - Water Quality Characterization Validation/Demonstration

Collect water samples from high priority areas in Whatcom County and analyze via WSS to compare DNA sequence data to those produced from the fecal source reference catalog.

5.4.1. Spiked Water Samples

A set of water samples were "spiked" with known fecal sources and compared against the fecal source reference catalog for overlap (Figure 10). Three of the spiked samples (Alpaca, Coyote, Muskrat) were included although a corresponding fecal source catalog reference did not exist in the catalog. An overlap of >25% for all spiked water samples with Unknown source was observed. Twelve out of thirteen spike samples showed an overlap with Nooksack Upper Mainstem with six showing >1% overlap. Duck-spike showed overlap with Duck Comp and Roof Bird Old. Sheep-spike showed only trace (<1%) overlap with sources. Deer-spike and Muskrat-spike showed overlap >1% with Roof Bird Old and Roof Bird Comp. Overlap (>1%) between sources and Alpaca-spike included Roof Bird Comp, Raccoon Comp, Goat Comp and Deer Comp. Dog-spike showed overlap with Roof Bird Comp, Raccoon Comp, Dog Comp and Deer Comp; overlap with Deer Comp was >25%. Dairy Lagoon1-spike showed overlap >1% with Dairy Lagoon (Fall) and >5% with all other Dairy Lagoon source types (i.e., Spring, Winter, Comp). Dairy Lagoon2-spike had overlap >1% with Beef Comp, Dairy Fresh, Dairy Solids Comp, Goat Comp, and Roof Bird Comp, as well as a >5% Dairy Lagoon (Winter). Roof Bird-spike showed >1% overlap with Nooksack Upper Mainstem, >5% with Duck Comp, and >25% with Roof Bird Old. Swan-spike showed >1% overlap with Roof Bird Old and Roof Bird Comp, and >5% overlap with Duck Comp and Goose Comp, as well as a trace overlap with Swan Comp. Boat Pumpout-spike sample showed a >1% with Dairy Lagoon (Fall) and Dairy Lagoon (Winter) and a >5% with Human Boat Pumpout. Considering a threshold of >5%, 6 out of 10 spikes with representation within the source catalog showed overlap with their spiked-in source.

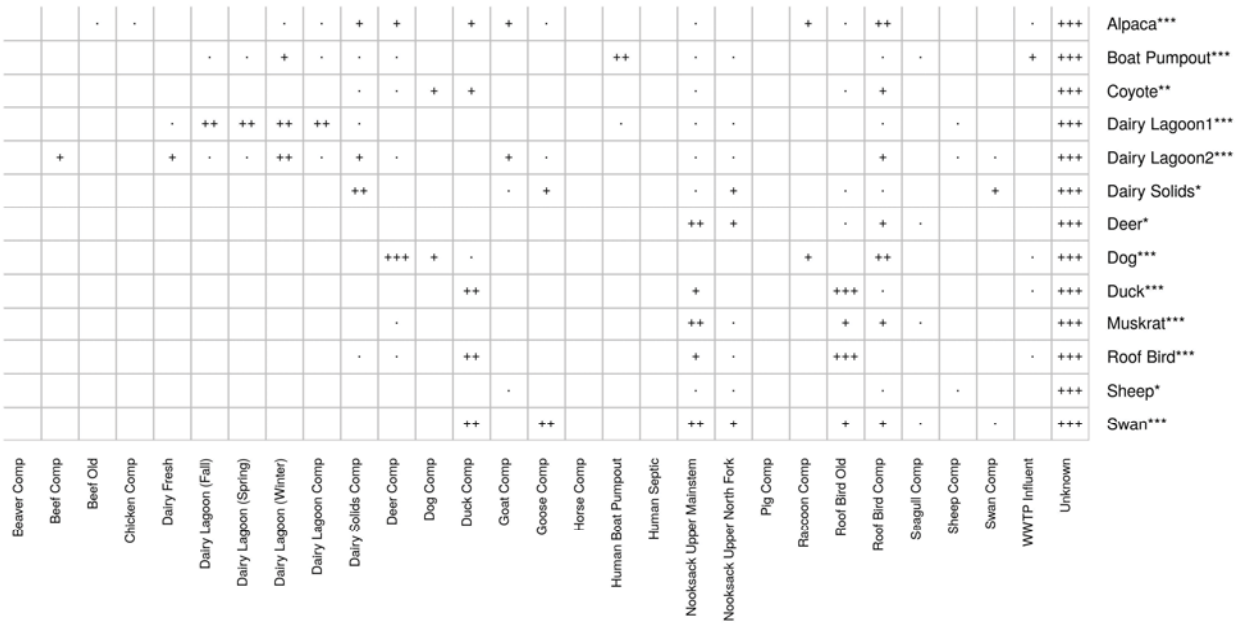

Figure 10. Estimated percent overlap of all analyzed fecal source microbial communities (x-axis) to 13 fecal "spiked" water sample microbial community (y-axis). Labels indicating range of estimated proportions: "•" (0.1% < 1%), "+" (1% < 5%), "++" (5% < 25%), and "+++" (>25%). The * after each sample ID indicates the fecal coliform via membrane filtration (mFC) count (CFU/100mL) for the spike water sample using the following range: "*" (0-200), "**" (200-1000), and "***" (>1000).

The water spike samples represented the positive controls for this study, a "best case scenario" for identifying sources in water samples. Overall, more than half (6/10) of the spikes showed more than trace (>1%) presence of their respective source (Figure 10). Considering the overlap that some of the sources have with each other, Dog/Horse, Human Boat Pumpout/Dairy Lagoon, and avian species (Figure 4), and the limited sample size of some sources, these results are encouraging. The overlap of the Dairy Lagoon and Human Boat Pumpout observed in the LOO analysis was reflected in the spike results as well. However, even though Human Boat Pumpout had a maximum overlap of 20% with Dairy Lagoon sources, the larger contribution in the Human Boat Pumpout spike was correctly identified as its respective source. This is also seen for the Dairy Lagoon spikes where Dairy Lagoon sources were the greater contributors. This highlights two notable observations: 1) the analysis is more robust to overlaps in source microbial community signatures and 2) the importance of further sampling efforts to better resolve source groups with shared microbial communities. By these results, a minimum overlap of >5% might be suggestive of the presence of a source in a water sample and therefore could serve as a threshold for determining importance of sources in an environment. Importantly, this putative threshold will need to be calibrated by further studies before it can be established as the absolute lower bound, particularly in the proposed use/case where waters (known to have lower abundances of microbes) are compared to sources.

Avian spike results (Swan, Duck, Roof Bird) reflected the convoluted relationships of the fecal source microbial community signatures for bird species. It is notable that, except for trace amounts, Swan, Duck and Roof Bird spikes overlapped with other avian reference fecal sources. This is further evidence that avian sources may be better suited as either a combined source or that a larger sample size is needed for the sources to be better discriminated from one another.

The Sheep spike did not show more than a trace (<1%) overlap with any sources, however the trace overlaps included both Sheep and Goat. Both Sheep and Deer spikes had very low fecal counts (<50

CFUs), which could be due to prior freezing of spike sample material, the freshness of the sample when it was collected, or because not enough feces were used to make the spike and thus provide enough bacteria for an adequate result. The hold study showed that some fecal types were unaffected by both longer hold time and a single freeze-thaw, but not all fecal types may be as robust.

There were three spike waters which had no representation in the source catalog: Coyote, Muskrat and Alpaca. The Coyote overlapped with Dog Comp suggesting that perhaps there is a signal with related animals. However, with the small sample size and confounding overlap for the Dog Comp, it is unclear if this is a real signal. Muskrat's largest overlap was seen with the two Roof Bird sources, both of which seem to be highly overlapping with other avian sources (Figure 4). It is possible there is more overlap between wildlife sources than was initially anticipated due to their varied diets, transient or migratory nature, and/or their interaction with natural environments (e.g. microbes in soil and aquatic environments.) Alpaca also showed overlap with some sources that have similar eating habits, namely those from Dairy Lagoon, Dairy Solids, Deer, and Goat. Similarities were also observed in the PCoA, where fresh Sheep, Goat, Horse, and Cattle sources grouped (Figure 3). It is possible that diet has an influence on microbial community signature similarities, and while it alone is not a proper grouping methodology, it is an important consideration when addressing similarity of fecal sources and groupings (McKenzie et al. 2017). Additionally, it may indicate that species with similar diets and/or digestive systems, need additional sample numbers, as well as analysis of individuals within a composite to confidently differentiate fecal source references.

Nooksack and Roof Bird sources showed overlap with most spike water samples, at least at the trace category (<1%). Six of the 13 spikes overlapped with one or both Nooksack River sources at >1% overlap. With the exception of Dairy Solids, these six overlapping spikes were made from wildlife fecal sources, which may share bacteria with the natural environment (aquatic or terrestrial) in which they live. Alternatively, shifts in the microbial community of feces after being deposited in the environment have been observed in other wildlife (Menke et al. 2015) and could be contributing to this shared overlap. Nine of the 13 spikes overlapped with either Roof Bird Old or Roof Bird Comp: 4 avian spikes (including Roof Bird spike), three non-avian wildlife sources (Coyote, Muskrat, Deer), Alpaca, and Dairy Lagoon. It is possible that Roof Bird sources share ubiquitous environmental bacteria with other wildlife sources. It was hypothesized that Roof Bird would be found- to some extent- in Dairy Lagoon samples, as the Roof Bird samples were collected from the roof gutter downspouts of dairy barns, the runoff of which is often diverted to the lagoon. Dairy roofs are heavily populated residences for starling, pigeon, and other roof birds, thus providing a mixed sample source of avian species.

The Unknown column was shown to overlap with all spikes at >25% suggesting two important things: 1) there remain portions of the microbial communities of the source types that require further characterization and 2) assessment of within-source variation in microbial community membership and abundance is important in establishing a comprehensive reference for each source type. It would be beneficial to explore the organisms that are found within the Unknown category for each source. It is possible that there are characteristics shared between all the microbes that are placed within this group. Perhaps they are extremely rare members of the microbial community or, conversely, they may be ubiquitous and are unable to provide specificity between sources. Further analysis of this dataset is recommended to answer these questions.

5.4.2. Suspected Source Water Samples

A set of water samples with "suspected" fecal pollution sources were obtained to test the fecal source reference catalog (Figure 11). An overlap of >25% for all suspected water samples (n=24) with Unknown source was observed, indicating that in part, water samples have a microbial community distinct from

any measured fecal sources. Nooksack Upper Mainstem was estimated to overlap >1% for most water samples (n=18), followed by Duck Comp (n=9), Roof Bird Old (n=4), WWTP Influent (n=2), and Dairy Lagoon (Winter) (n=1). Skagit Waterway-January1 & 2 show similar patterns of estimated overlap for Duck and Roof Bird Old though Skagit Waterway-January2 showed an overlap >5% for Duck Comp. Skagit Waterway-March was the only overlap with Dairy Lagoon (Winter) of >1%. Roof Water-March showed overlap >5% with Duck Comp and Roof Bird Old. Field Drain-March1 & 2 showed overlap >1% with Duck Comp, as did all three Field-April samples. Field Drain-April showed overlap >1% with Roof Bird Old. Both WWTP Effluent-April and WWTP Influent-December showed overlap >1% with WWTP Influent.

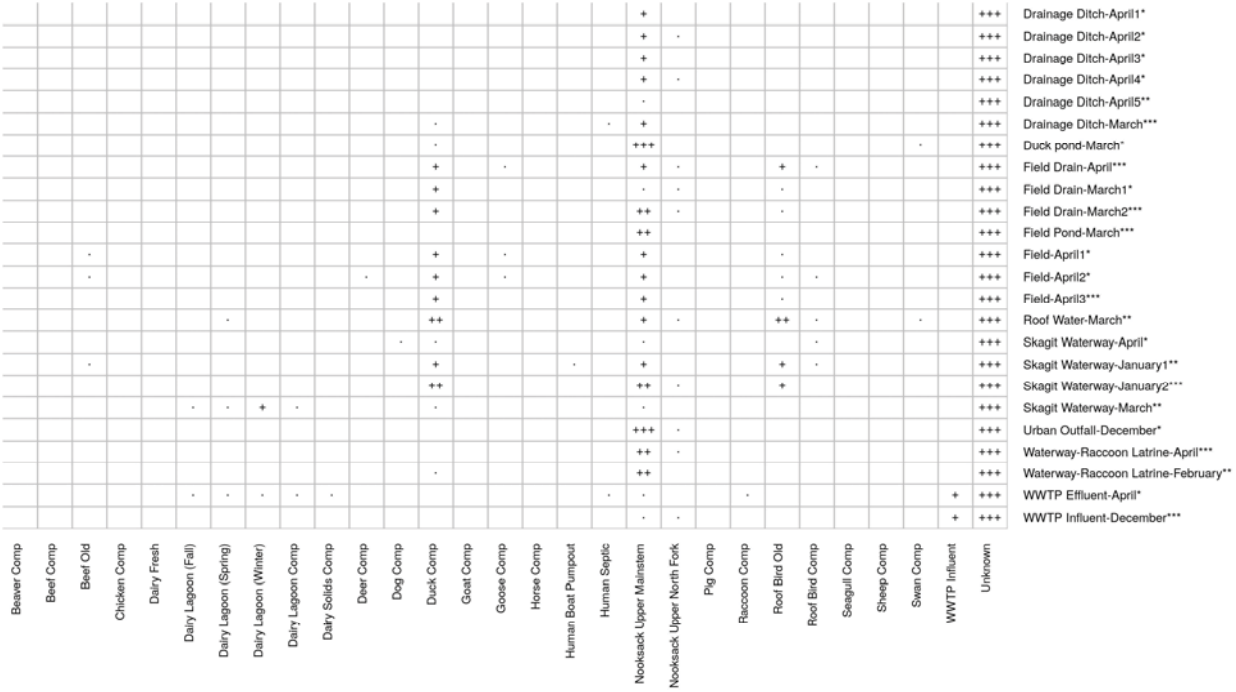

Figure 11. Estimated percent overlap of microbial community shared between fecal source (x-axis) and 24 surface water samples (freshwater) with suspected fecal impacts (y-axis). Water samples are named by their monitoring site type and month. Labels indicate range of estimated overlap of microbial community between sources and samples: "•" (0.1% < 1%), "+" (1% < 5%), "++" (5% < 25%), and "+++" (>25%). The Unknown source column represents the percent proportion of the microbial community in the waters that did not match any of the fecal sources represented. The * after each sample ID indicates the fecal coliform via membrane filtration (mFC) count (CFU/100mL) for the water sample using the following range: "*" (0-200), "**" (200-1000), and "***" (>1000).

The locations and/or dates of water sample collection were chosen to represent sites with "suspected" fecal sources and high fecal counts (>200 CFU). These suspected samples (Figure 11) were meant to be an extension of the spike water samples (Figure 10), further testing the fecal source catalog on increasingly more complex sample types, but with a suspected rather than known fecal source types. As may be expected in environmental samples, a large percentage of the microbial population of the water samples were attributed to the Unknown source category. This indicates that the suspected water samples had a microbial community that is largely distinct from the cataloged fecal sources. The most commonly observed signature, except for Unknown source, was that of the Nooksack Upper Mainstem. This supports that certain microbes are shared throughout freshwater aquatic environments, even in different systems or waterways. Though all suspected waters showed overlap with at least one of the

Nooksack River Mainstem or North Fork sources, none of the suspected water samples were taken downstream of these locations.

Duck and Roof Bird were the most common avian fecal sources found in the suspected water samples. The Field Drain and Field samples, which overlapped with these avian sources, were taken in areas known to have waterfowl and migratory birds present during some parts of the year, or from barn roof gutter downspouts. Migratory and endemic avian species are numerous in Whatcom and Skagit Counties and can be found in populations scattered throughout the area. Avian sources are a particularly challenging fecal source due to birds' ability to travel uninhibited over large areas and propensity to gather in field ponds and near waterways. The limited resolution of these sources as references, due to higher overlap of their microbial community signatures, also introduces complexity in distinguishing the different avian source types. It may be necessary to group avian sources into a single reference and/or combine source identification methods with field observations in order to draw stronger conclusions about which avian species may be contributing fecal bacteria to a system (i.e. duck, goose, swan, or roof birds).

The wastewater treatment plant influent and effluent overlapped with the WWTP Influent source. WWTP Effluent showed trace overlap with Dairy Lagoon and Dairy Solids. WWTP Influent source also showed a small overlap with some dairy sources, as well as with Human Boat Pumpout (which shared overlap with Dairy sources). Water comparisons may be improved with increased sampling for human fecal source references to further resolve the overlaps between these sources.

Waterway-Raccoon Latrine samples (April and December) showed no overlap with the Raccoon Comp source. Raccoon are known to be opportunistic omnivores and their dietary habits are greatly dependent on the area in which they live (WDFW 2019). As previously described in this report, and as other research has suggested, diet can play a large role in the microbial community signature of animals (McKenzie et al. 2017). Raccoon, as with dog, may have more variable microbial communities and therefore require a greater number of individual samples within a composite to account for the variation. For this project, we collected fecal samples from 5 sites representing up to 15 individual animals. A larger sampling size for animals with highly variable diets may help improve the ability of the fecal reference catalog sample to be discriminated within an environmental water sample.

5.4.3. Unknown Source Water Samples

Environmental, "unknown", water samples were taken from locations around Whatcom County that were known to have occasional or frequent high fecal counts, but do not have known fecal sources, to investigate the potential of the fecal source reference catalog to identify fecal sources in ambient waters (Figure 12). An overlap of >25% was observed for all environmental water samples (n=38) with Unknown source, indicating that in part water samples have a microbial community distinct from any measured fecal sources. Nooksack Upper Mainstem was estimated to overlap >1% for most water samples (n=34), followed by Nooksack Upper North Fork (n=7), Dairy Solids (n=7), Duck (n=7), Dairy Lagoon (Fall (n=3), Spring (n=1), Winter (n=1), Comp (n=1)), Roof Bird Old (n=2), WWTP Influent (n=2), Beef Comp (n=1), and Goat Comp (n=1). B1-April and CA14b-November showed overlap >1% with Dairy Lagoons. BH-January, CA14c-November, and FT10-February showed overlap >1% with Duck Comp. DD6-November1 & 2 showed overlap >1% with Dairy Solids. H3-December showed an overlap >1% with multiple seasons of Dairy Lagoon, Dairy Solids, Duck, Goat Comp, Roof Bird Old (>5%), and Roof Bird Comp. LYNCS-December shared overlap >5% with Dairy Solids and >1% with WWTP Influent. LYNCS-November showed overlap >1% with Dairy Lagoon (Fall), Dairy Solids, and WWTP Influent.

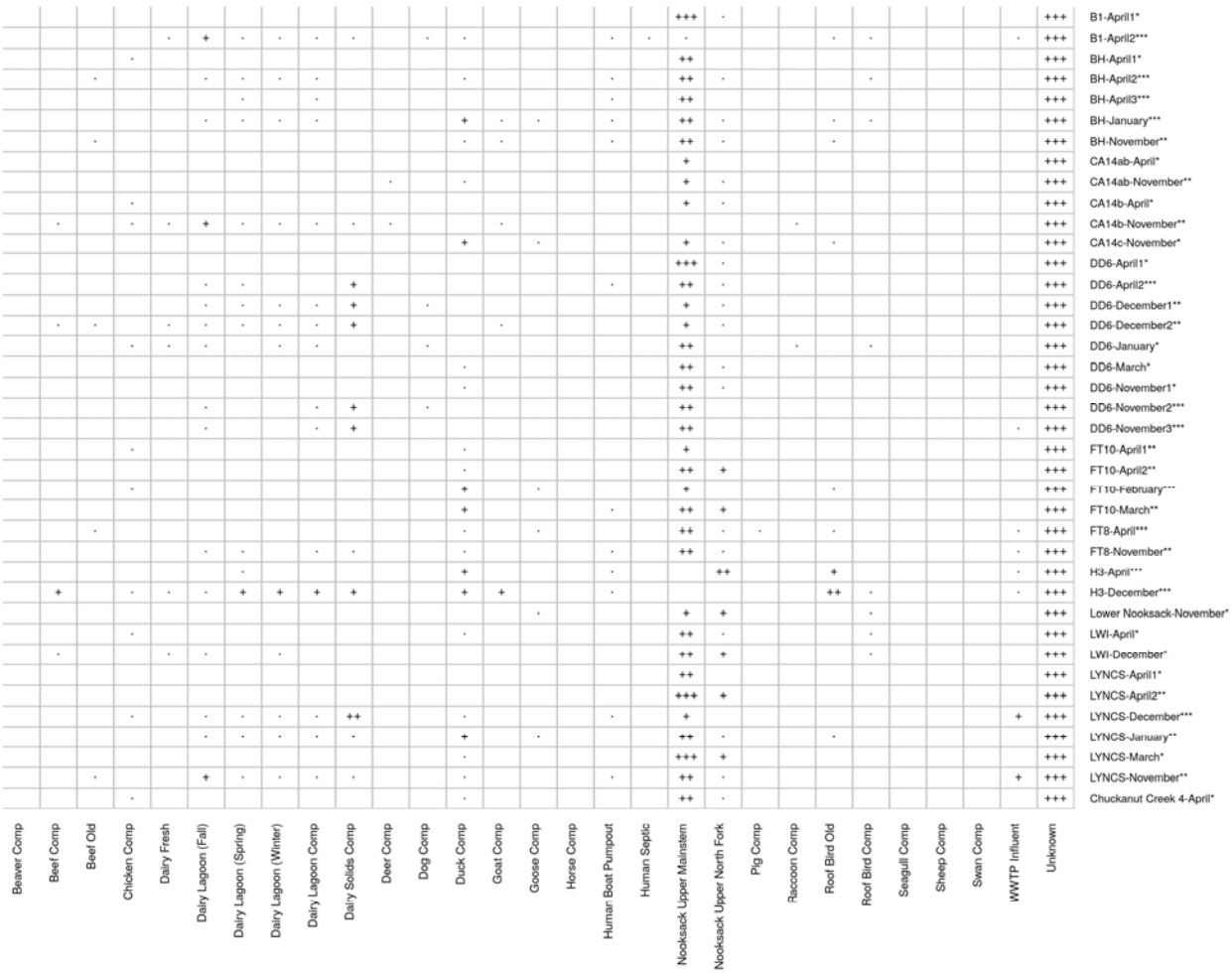

Figure 12. Estimated percent overlap of microbial community shared between fecal source (x-axis) and surface waters (freshwater) within Whatcom County with unknown or unidentified fecal impacts (y-axis). Water samples are named by their monitoring site location and month. Labels indicate range of estimated overlap of microbial community between sources and samples: "•" (0.1% < 1%), "+" (1% < 5%), "++" (5% < 25%), and "+++" (>25%). The Unknown source column represents the percent proportion of the microbial community in the waters that did not match to any sources. The * after each sample ID indicates the fecal coliform via membrane filtration (mFC) count (CFU/100mL) for the water sample using the following range: "*" (0-200), "**" (200-1000), and "***" (>1000).

The sampling of unknown waters (Figure 12) was done as a supplement to the experimentally controlled spiked water samples (Figure 10) and the suspected source waters (Figure 11). These samples not only served the purpose of determining the potential, and desired, utility of the fecal source reference catalog, they also showed the variation in sample location sites when repeated sampling took place over time. The fecal coliform counts of water samples ranged from <4 CFU to >6000 CFU/100ml; 26 of 44 samples had a fecal coliform count >200 CFU/100ml, which was the sample obtainment goal. Ten sites were sampled twice, and three sites were sampled five or more times to capture seasonal and event variables; BH (n=5), DD6 (n=9), and LYNCS (n=6).

The site BH was sampled five times between November 2018 and April 2019. BH is located on Bertrand Creek near the Canadian border. A pattern emerges when looking at the source inclusions at this site. Four out of the five samples had Roof Bird and Human Boat Pumpout, and three out of the five had inclusion of Dairy Lagoon Comp and Duck Comp. Although most of the inclusions were trace (<1%),

except for Duck Comp in BH-January, the BH sample sites showed overlap with the same sources suggesting a locational microbial "signature". This is at least the case within the months of collection; further year-round sampling would give stronger evidence to support this hypothesis. It is likely that this site has dairy and avian fecal contributions. It is unlikely that boat pumpouts are influencing this site, but further work to investigate septic system contributions and RV park practices upstream of this site are needed.

The DD6 and LYNCS sample locations, on Double Ditch at the Canadian border and Fishtrap Creek within the city of Lynden, respectively, show shared patterns of microbial signatures when compared with fecal sources. Both DD6 and LYNCS did not overlap with Roof Bird, except for one sample from each site. The sites also had an overlap with Dairy Solids >5%, and trace inclusions of Dairy Lagoon (all seasons). The overlap of sources that occurred in these sites suggests that there is value in repeated sampling over time or seasons. With repeat sampling, observations of source impacts to waterways could be collected and utilized for monitoring and mitigation efforts.

5.4.4. Marine Water Samples

Marine waters were collected to determine if fecal sources could be identified in the marine water system (Figure 13). The majority of overlaps between marine samples (n=8) and fecal source microbial community signatures were <1%, with the exception of the Unknown, which indicates microbial community signatures that are distinct from available sources. Drayton 4-January showed an overlap of >1% with Duck. Portage Bay 50-December showed a >1% overlap with Nooksack Upper Mainstem. Fecal sources that overlapped at >0.1% with marine waters included: Chicken Comp, Dairy Lagoon Comp, Deer Comp, Dog Comp, Duck Comp, Goose Comp, Horse Comp, Nooksack Upper Mainstem, Nooksack Upper North Fork, Roof Bird Old, Roof Bird Comp, and WWTP Influent.

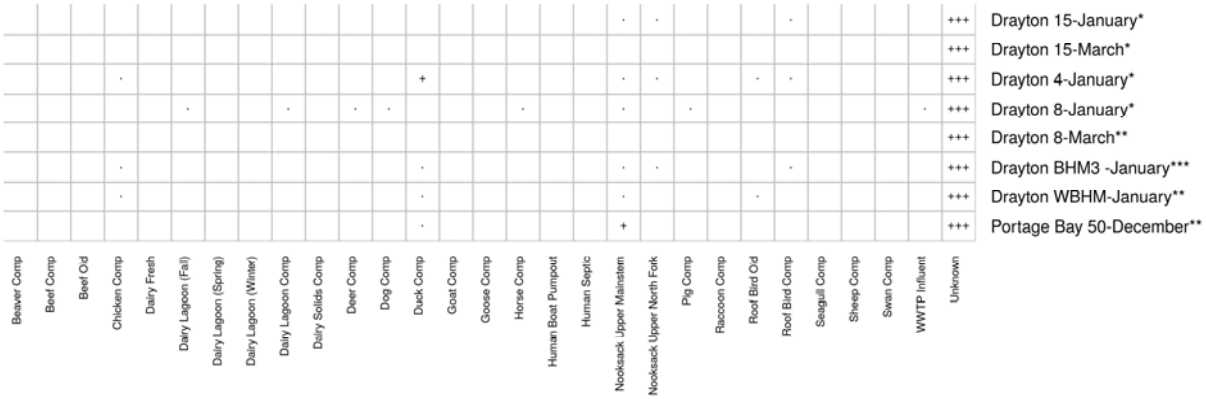

Figure 13. Estimated percent overlap of microbial community shared between fecal source (x-axis) and marine waters in Whatcom County (y-axis). Water samples are named by their monitoring site location and month. Labels indicate range of estimated overlap of microbial community between sources and samples: "•" (0.1% < 1%), "+" (1% < 5%), "++" (5% < 25%), and "+++" (>25%). The Unknown source column represents the percent overlap of the microbial community in the waters that did not match to any sources. The * after each sample ID indicates the fecal coliform via membrane filtration (mFC) count (CFU/100mL) for the water sample using the following range: "*" (0-14), "**" (15-43), and "***" (>43).

Six unique marine sites were sampled for this project within Portage Bay and Drayton Harbor (Figure 13). The locations chosen were expected to have high fecal counts (>43 CFU/100ml) at the time of sampling. The fecal counts for the samples collected ranged from 8-48 CFU/100ml, with most marine water fecal counts lower than 20 CFU/100mL. The ambient conditions and fecal counts for samples

collected were representative of routine marine water samples collected throughout the year in these locations; however due to low fecal counts their application for testing the fecal source reference catalog may have been limited.

Unlike freshwater samples, the marine water samples shared much less overlap with Nooksack sources (<1%), which suggests that marine waters have a unique microbial community composition. One marine water overlapped >5% with Nooksack Upper Mainstem, Portage Bay 50-December, which is located in Portage Bay at the mouth of the Nooksack River. (Drayton Harbor is not influenced by the Nooksack River). Ability to detect the microbial community signature of a freshwater source can be detected in its receiving marine water (Portage), and was not detected in non-receiving waters (Drayton), supports the utility of this study.

Repeat sampling can give insight into the temporal variability of the microbial community in marine water, as previously demonstrated in freshwater portion of this study. Similarly, comparing the microbial communities of Drayton sites 8 and 15 from January to March show different dominant phyla (data not shown) and different fecal sources (Figure 13). For example, eight additional phyla were seen at Drayton 8 in January, including Firmicutes, which were not present at the same site in March. These microbial community differences may have been a contributing factor to the identification of more trace (<1%) overlaps with sources in January than in March. These overlaps of fecal sources to Drayton 8 in January consisted of sources that had a high relative abundance of Firmicutes. Further work with repeated sampling of marine sites would further elucidate this pattern; this type of sampling was part of the initial study design, but low fecal counts in the marine water samples collected precluded analysis of the samples.

Marine water samples demonstrated overlap with several avian sources including Duck Comp, Chicken Comp and Roof Bird Old and Comp. Drayton 4 from January had the only overlap >1% with a source (>5% overlap with Duck). Both Drayton Harbor and Portage Bay have a large population of marine waterfowl, which may explain this overlap; however, ducks are not the most likely avian source at these sites. Drayton Harbor, specifically, has many seagulls and is home to nesting cormorants on the marina breakwater most of the year. Interestingly, there was no overlap in any of the marine waters with Seagull. It is possible that seagulls were not present in enough abundance to significantly affect the water microbial community of the collection locations- or perhaps more extensive sampling of seagull feces needs to occur (a low sample size was collected for Seagull of one location representing approximately 10 individuals). Other shorebirds, marine waterfowl and other marine animals that may have an impact on marine waters should be included in the fecal reference catalog as it continues to be developed.

5.4.5. Deliverable 4 Conclusions and Future Work

Spike water samples were created to test the ability of the catalog to identify source inclusions in water samples. More than half of spikes were identified as their respective spiked source, showcasing the feasibility of using this method to recover sources from water samples. Along with the spike waters, environmental waters (suspected and unknown) showed similarities between their microbial communities providing preliminary data that describes connectivity between these sample sites within the water system. A subset of microbial community members appears to be ubiquitous in freshwater sample environments. Importantly, the marine sample site (Portage Bay) that is closest to the mouth of the Nooksack River showed overlap with the Nooksack sources, offering evidence that it may be possible to monitor for freshwater signals in local marine waters.

These encouraging preliminary results from water comparisons have led to several next steps for improving and increasing the utility of microbial signature-based water quality assessment and

monitoring. A more robust spike recovery study will need to be performed to precisely assess the detection limits of sources in water samples. As part of this work, a water volume study is a needed to address questions of how much water is required for a reasonable assessment of an impacted site. Establishing detection limits and sample volume requirements will facilitate large scale microbial metagenome monitoring within the impacted waters. Longitudinal water sampling will also need to be included in future monitoring efforts to build a robust baseline for water systems of concern. This would encompass historically impacted fresh and marine waters over every season and across similar geographic regions. Our goal is that this study will contribute to future water quality monitoring research for Whatcom County as well as to help foster lasting collaborations within the greater Washington State network of local, state, and federal conservation groups.

6. Outreach and Partner Connection

Outreach to stakeholders is a critical part of a successful project. Project staff and partners conducted a number of meetings, presentations, and outreach discussion throughout the project period, with additional outreach planned after the project concludes. Additionally, project staff enrolled local partners in the selection and/or collection of environmental water and fecal samples, as well as input into the Wildlife Tracker App that was created for the project. This enabled us to increase the interest, relevance, and capacity of the project locally.

To garner local support and inform the community of the project, numerous presentations were made at local Whatcom Clean Water Project (WCWP), Watershed Improvement District (WID), Whatcom Family Farmers, and Pollution Identification and Correction (PIC) Program meetings, as well as local Farmer Speaker Series events and the annual Manure Nutrient Management Training (January 2019). All communication of the project has been met with positive feedback and appreciation for the goals of the project for local water quality characterization. Additional presentations are planned at the conclusion of the project to share final results with local partners and at State and/or national conferences over the next year.

Due to the natural timeline of the project, with fecal and water sample collection occurring over a timespan to accommodate season, opportunity, and/or desired levels of contamination in water, followed by sample processing and analysis, and lastly data analysis, results were not made available until the very end of the project. Therefore, creation and submission of any peer reviewed publication or journal articles will occur after the completion of the project.

The successful completion of the project provides local partners and Tribes with improved water quality DNA/source characterization techniques to integrate into their programs for better pollution source identification and correction. The long-term outcome of this outreach effort will contribute to improved water quality and reopening/reclassification of shellfish production areas.

7. QAPP Development and Effectiveness

A Quality Assurance Project Plan (QAPP) was created for this project and approved by the WSCC in September 2018 prior to commencement of work. The QAPP was consulted throughout the project to ensure we were meeting project goals and deliverables. The QAPP offered a helpful framework and expectation for multiple staff and agencies to work within. However, as we progressed through the project, we did encounter some required modifications and adaptations of the QAPP that had to be addressed. Some of the prominent benefits and challenges to the QAPP are outline below.

- The sample methods outlined in the QAPP worked well.
- The sample nomenclature outlined was clear and the QAPP was instrumental as a reference document for consistent nomenclature and fecal reference sample design.
- Project communication and coordination procedure outlined in the QAPP were helpful, especially clarification of roles and responsibilities, when working with multiple project partners.
- Due to the unpredictable and uncontrollable nature of the weather, we were not able to collect all water samples on an exact schedule. This was noted in the QAPP to allow flexibility, but should be accommodated in future projects of this type.
- While some of the fecal samples were easy to collect (livestock, human), wildlife samples were more unpredictable and had to be discovered. Because of this, we didn't collect as many replicates for composites of some wildlife species as we had anticipated in the QAPP.
- Because of the unpredictability of precipitation, runoff, and/or pollution events, there was difficulty collecting water samples with high fecal coliform counts - especially marine waters - as was outlined in the QAPP.
- Spike samples were not available until the end of the sample collection period, which limited opportunities to optimize methods.

8. Lessons Learned

The following represent the "lessons learned" - good and bad - throughout this project.

- To achieve a higher degree of confidence and statistical support for fecal reference samples, all individuals within a composite should be analyzed to account for variability and similarity. The Dairy Liquid samples taught us that there is more discriminatory power with more replicates. Unfortunately, we only allocated time and funding to do this for a few of the fecal source sample types. This lesson supports the case to build a bigger, more robust catalog as we progress. It should be noted that the results showed that we are able to discriminate samples that lack sufficient replicates. However, statistical support is needed to determine if separations between specific source types are valid.
- The collection of upstream Nooksack water samples provided microbial signatures prior to the introduction of certain fecal source types. This showed that having a baseline microbial signature for an environment of concern is important for being able to distinguish potential contamination events in water samples.
- The creation of spiked water samples as positive controls was necessary in testing the identification of sources included (or not included) in the fecal source reference catalog. The use of spike samples early in the study to determine the effectiveness of the methodology is very important and should be used in similar future studies.
- The volume of water (2L) utilized in this study is challenging to obtain and transport. Future work needs to be done to determine the necessary sample volume for detection of fecal sources in water samples.
- The addition of the fecal hold time experiment proved to be valuable. It demonstrated that samples could be held for up to 3 days at 4°C, or frozen for 30 days without compromising the microbial community signature. This is useful for both the collection agency/partner and the laboratory conducting the analysis as it provides options for holding time prior to analysis to accommodate for time delays such as late day drop-offs, holidays, weekends, etc.
- A longer timeline for the study would have been helpful to provide more time for results assessment. Due to the nature of the project, fecal and water samples were collected over a long period of time to capture seasonal and opportunistic sampling events. This meant that data

analysis, which takes time, could not occur until all samples were obtained, processed, and analyzed. In some cases, the results of the analysis provided us information that was used to revise or improve methods. Therefore, we found ourselves doing much of the assessment of data at the very end of the project when it was finally available. More time for this process would have allowed deeper analysis and perhaps, additional conclusions of value.

9. Recommendations

Due to the novelty nature of this project, we discovered many areas of follow up work, improvement, and method modification we would recommend for future projects of this type. The following are recommendations for similar projects in future.

- Fecal hold times and freeze-thaw cycles – Determine how many freeze-thaw cycles can occur without significantly degrading the microbial community signature of fecal samples and establish the limit of refrigeration time (5, 7, 10 days, etc.) before freezing is required.
- Sample fecal source reference material in greater numbers and/or replication to allow statistical tests for comparisons between reference sources and water samples. More than one data point is required ($n \geq 3$) to be able to determine statistical differences between closely related sources.
- Expand on fecal source reference catalog development and/or representation of species. Maintain a list of fecal reference samples that are desired additions to the catalog and pursue these additions if future funding is obtained. This list includes: 1) fecal references which were identified but not prioritized for collection due to budget constraints, 2) seasonally or opportunistic samples that were not able to be collected (e.g. spring cormorant nesting season), or 3) new fecal sources that become important on the landscape with changing land use.
- Water volume sampling – Determine how much water is needed to capture the microbial community signature if samples have low, medium, or high fecal counts. A small study was conducted to begin answering this question, however the results are not yet available at the time of reporting.
- Correlate comprehensive metadata from water samples (turbidity, precipitation at/just before time of sampling, water chemistry, fecal count, location) with DNA extraction concentrations that are <0.05 ng/µL. The goal of this assessment would be to determine if DNA extraction success can be improved over a wider range of water sample conditions.
- Determine whether 0.45 µm filters alone are enough to compare water sample metagenomes to the fecal source reference catalog. In principle, this experiment would require several spiked water samples to be processed through 0.2 µm filters and 0.45 µm filters in parallel instead of in series.
- Create semi-quantitative spiked water samples to begin to establish thresholds for discriminating between sources and/or examine how complexity within a sample impacts the limit of detection for individual fecal sources within complex mixtures.
- The lack of specificity of the Dog signature makes using it as a reference challenging at this point, since it overlaps with other reference sources. (This is despite a composite of 14 individual dogs, which followed the project QAPP.) The Dog data includes numerous similarities with data from other fecal sources, which may imply that dog microbiomes are variable or that they may “carriers” of other microbes within an environment due to their omnivorous diet and proclivity to eat soil, vegetation - or even feces. Further work to define the Dog fecal source reference signature is recommended given its potential impact on the landscape and importance in bacteria source reduction efforts.

10. References

Ahmed, W., et al. "Toolbox approaches using molecular markers and 16S rRNA gene amplicon data sets for identification of fecal pollution in surface water." Appl. Environ. Microbiol. 81.20 (2015): 7067-7077.

Anderson, Marti J. "A new method for non-parametric multivariate analysis of variance." Austral ecology 26.1 (2001): 32-46.

[APHA] American Public Health Association. 2005 Standard method for examination of water and wastewater, 21st edn. APHA, AWWA, WPCF, Washington.

Bahl, Martin Iain, Anders Bergström, and Tine Rask Licht. "Freezing fecal samples prior to DNA extraction affects the Firmicutes to Bacteroidetes ratio determined by downstream quantitative PCR analysis." FEMS microbiology letters 329.2 (2012): 193-197.

[Battelle]. "Quality Assurance Project Plan: EPA Targeted Watershed Grants Program Dungeness River and Estuary: Task 1 – Microbial Source Tracking." Battelle Marine Sciences Laboratory. (2005) EPA Grant WS-97098101-0. Prepared by: Battelle Marine Sciences Laboratory.

Bernhard, Anne E., and Katharine G. Field. "Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes." Appl. Environ. Microbiol. 66.4 (2000): 1587-1594.

Bokulich, Nicholas A., et al. "Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin." Microbiome 6.1 (2018): 90.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. "Trimmomatic: a flexible trimmer for Illumina sequence data." Bioinformatics 30.15 (2014): 2114-2120.

Bolyen, Evan, et al. QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. No. e27295v1. PeerJ Preprints, 2018.

Brown, Clairessa M., et al. "A high-throughput DNA-sequencing approach for determining sources of fecal bacteria in a Lake Superior estuary." Environmental science & technology 51.15 (2017): 8263-8271.

Brown, Clairessa M., et al. "Influence of Library Composition on SourceTracker Predictions for Community-Based Microbial Source Tracking." Environmental science & technology 53.1 (2018): 60-68.

Callahan, Benjamin J., et al. "DADA2: high-resolution sample inference from Illumina amplicon data." Nature methods 13.7 (2016): 581.

Costa MC, Arroyo LG, Allen-Vercoe E, Stampfli HR, Kim PT, Sturgeon A and Weese JS. "Comparison of the fecal microbiota of healthy horses and horses with colitis by high throughput sequencing of the V3-V5 region of the 16S rRNA gene." PloS one 7.7 (2012): e41484.

Dogan, Belgin, and Kathryn J. Boor. "Genetic diversity and spoilage potentials among Pseudomonas spp. isolated from fluid milk products and dairy processing plants." Appl. Environ. Microbiol. 69.1 (2003): 130-138.

Douglas, Erica. "Whatcom County Water Quality Program Bacteria Monitoring QAPP." Whatcom County Public Works. (2017) http://www.whatcomcounty.us/DocumentCenter/View/25505

[Ecology]. Washington Department of Ecology. 2011. Review and Critique of Current Microbial Source Tracking (MST) Techniques. Washington State Department of Ecology. Pub No. 11-03-038.

[EPA]. U.S. Environmental Protection Agency. 2005. Microbial Source Tracking Guide Document. EPA/600-R-05-064. Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, OH.

[EPA]. U.S. Environmental Protection Agency. 2011. Using Microbial Source Tracking to Support TMDL Development and Implementation. Prepared for U.S. Environmental Protection Agency, Region 10, Watersheds Unit by Tetra Tech Inc. and Herrera Environmental Consultants, Seattle, WA.

Eddy, Sean R. "Accelerated profile HMM searches." PLoS computational biology 7.10 (2011): e1002195.

Ervin, Jared S., et al. "Characterization of fecal concentrations in human and other animal sources by physical, culture-based, and quantitative real-time PCR methods." Water research 47.18 (2013): 6873-6882.

Faith, Daniel P. "Conservation evaluation and phylogenetic diversity." Biological conservation 61.1 (1992): 1-10.

Glasner, Jeremy D., et al. "Genome sequence of the plant-pathogenic bacterium Dickeya dadantii 3937." (2011): 2076-2077.

Hagedorn, Charles, et al. "Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci." Appl. Environ. Microbiol. 65.12 (1999): 5522-5531.

Harwood, Valerie J., John Whitlock, and Victoria Withington. "Classification of antibiotic resistance patterns of indicator bacteria by discriminant analysis: use in predicting the source of fecal contamination in subtropical waters." Appl. Environ. Microbiol. 66.9 (2000): 3698-3704.

Hyatt, Doug, et al. "Prodigal: prokaryotic gene recognition and translation initiation site identification." BMC bioinformatics 11.1 (2010): 119.

Jeong, Ju-Yong, et al. "Microbial community analysis and identification of alternative host-specific fecal indicators in fecal and river water samples using pyrosequencing." The Journal of Microbiology 49.4 (2011): 585.

Katoh, Kazutaka, et al. "MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform." Nucleic acids research 30.14 (2002): 3059-3066.

Klindworth, Anna, et al. "Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies." Nucleic acids research 41.1 (2013): e1-e1.

Knights, Dan, et al. "Bayesian community-wide culture-independent microbial source tracking." Nature methods 8.9 (2011): 761.

Lee, Jung Eun, et al. "Analysis of human and animal fecal microbiota for microbial source tracking." The ISME journal 5.2 (2011): 362.

Leight, Andrew K., Byron C. Crump, and Raleigh R. Hood. "Assessment of fecal indicator bacteria and potential pathogen co-occurrence at a shellfish growing area." Frontiers in microbiology 9 (2018): 384.

Li, Dinghua, et al. "MEGAHIT v1. 0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices." Methods 102 (2016): 3-11.

Lovanh, Nanh, et al. "The effect of stratification and seasonal variability on the profile of an anaerobic swine waste treatment lagoon." Bioresource technology 100.15 (2009): 3706-3712.

Love, Michael I., Wolfgang Huber, and Simon Anders. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome biology 15.12 (2014): 550.

Lozupone, Catherine A., et al. "Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities." Appl. Environ. Microbiol. 73.5 (2007): 1576-1585.

Lozupone, Catherine, and Rob Knight. "UniFrac: a new phylogenetic method for comparing microbial communities." Appl. Environ. Microbiol. 71.12 (2005): 8228-8235.

Macauley, John J., Craig D. Adams, and Melanie R. Mormile. "Diversity of tet resistance genes in tetracycline-resistant bacteria isolated from a swine lagoon with low antibiotic impact." Canadian journal of microbiology 53.12 (2007): 1307-1315.

McKenzie, Valerie J., et al. "The effects of captivity on the mammalian gut microbiome." Integrative and Comparative Biology 57.4 (2017): 690-704.

Menke, Sebastian, Matthias Meier, and Simone Sommer. "Shifts in the gut microbiome observed in wildlife faecal samples exposed to natural weather conditions: Lessons from time-series analyses using next-generation sequencing for application in field studies." Methods in Ecology and Evolution 6.9 (2015): 1080-1087.

Ntougias, Spyridon, et al. "Bacterial diversity in spent mushroom compost assessed by amplified rDNA restriction analysis and sequencing of cultivated isolates." Systematic and applied microbiology 27.6 (2004): 746-754.

Odagiri, Mitsunori, et al. "Validation of Bacteroidales quantitative PCR assays targeting human and animal fecal contamination in the public and domestic domains in India." Science of the Total Environment 502 (2015): 462-470.

Parveen, Salina, et al. "Association of multiple-antibiotic-resistance profiles with point and nonpoint sources of Escherichia coli in Apalachicola Bay." Appl. Environ. Microbiol. 63.7 (1997): 2607-2612.

Price, Morgan N., Paramvir S. Dehal, and Adam P. Arkin. "FastTree 2–approximately maximum-likelihood trees for large alignments." PloS one 5.3 (2010): e9490.

Quast, Christian, et al. "The SILVA ribosomal RNA gene database project: improved data processing and web-based tools." Nucleic acids research 41.D1 (2012): D590-D596.

[WDFW]. “Species and Habitats, Species in Washington - Raccoons.” Washington State Fish and Wildlife. (2019) https://wdfw.wa.gov/species-habitats/species/procyon-lotor#living

Sargeant, Debby, William R. Kammin, and Scott Collyard. Review and critique of current microbial source tracking (MST) techniques. Environmental Assessment Program, Washington State Department of Ecology, 2011.

Segata, Nicola, et al. "Metagenomic microbial community profiling using unique clade-specific marker genes." Nature methods 9.8 (2012): 811.

Shin, Na-Ri, Tae Woong Whon, and Jin-Woo Bae. "Proteobacteria: microbial signature of dysbiosis in gut microbiota." Trends in biotechnology 33.9 (2015): 496-503.

Sledz, Wojciech, et al. "Growth of bacterial phytopathogens in animal manures." Acta Biochimica Polonica 64.1 (2017): 151-159.

[SM9222D]. “Membrane Filter Technique for Members of the Coliform Group: 9222D. Fecal Coliform Membrane Filter Procedure.” Standard Methods Online – Standard Methods for the Examination of Water and Wastewater. . MF Partition Procedures. (2019) http://standardmethods.org

Tatusov, Roman L., et al. "The COG database: a tool for genome-scale analysis of protein functions and evolution." Nucleic acids research 28.1 (2000): 33-36.

Toth, Ian, et al. "Why genomics research on Pectobacterium and Dickeya makes a difference." American Journal of Potato Research 92.2 (2015): 218-222.

Unno, Tatsuya, et al. "Use of barcoded pyrosequencing and shared OTUs to determine sources of fecal bacteria in watersheds." Environmental science & technology 44.20 (2010): 7777-7782.

van der Kolk, J. H., et al. "Acinetobacter in veterinary medicine, with an emphasis on Acinetobacter baumannii." Journal of Global Antimicrobial Resistance 16 (2019): 59-71.

[WDFW] Washington Department of Fish & Wildlife. 2019. Raccoons, Species and Habitats, Species in Washington. https://wdfw.wa.gov/species-habitats/species/procyon-lotor#living. Accessed 8 July 2019.

Wiggins, Bruce A. "Discriminant analysis of antibiotic resistance patterns in fecal streptococci, a method to differentiate human and animal sources of fecal pollution in natural waters." Appl. Environ. Microbiol. 62.11 (1996): 3997-4002.

Xin, K. E., et al. "Effects of oxytetracycline on methane production and the microbial communities during anaerobic digestion of cow manure." Journal of Integrative Agriculture 13.6 (2014): 1373-1381.