AN ABSTRACT OF THE THESIS OF

Carolyn A. Petersen for the degree of Master of Science in Microbiology presented on September 5, 2006. Title: Analysis of *Bacteroidales* 16S rRNA Gene Sequences from a Geographically Isolated Human Population.

Abstract approved

Katharine G. Field

Standard methods of measuring fecal pollution in water do not distinguish between human and non-human sources. Molecular technology enabled the development of host-specific markers that distinguish fecal sources. Human specific PCR primers, HF183F and HF134F, were designed based on phylogenetic analyses of partial 16S rRNA gene sequences from the *Bacteroidales* group of fecal anaerobes. Both primers amplify human fecal DNAs in the U.S., Europe, New Zealand, and Japan. However, they did not amplify human fecal DNAs from a geographically isolated population in Alaska, although amplification was possible with general *Bacteroidales* primers. We undertook phylogenetic analysis to compare *Bacteroidales* 16S rRNA genes from the isolated population to a non-isolated population. Our ultimate objective was to create new *Bacteroidales* human-specific primers from the full-length 16S rRNA gene to amplify fecal DNAs from both isolated and non-isolated geographic areas. Sequence libraries from the isolated Alaskan population and from Oregon were

created by amplifying the full-length *Bacteroidales* 16S rRNA gene. Fragments were cloned and sequenced and 96 colonies from each geographic location were screened for inserts. Phylogenetic analysis and primer design used ARB software. Humans in the Alaskan isolated population did not have sequences from certain common human *Bacteroidales* groups, and also contained unique clades. The trees constructed showed that none of the Alaskan sequences grouped with sequences from which the current human primers were developed, explaining why fecal samples from this population did not amplify with those primers. A novel clade only contained human sequences from the Alaskan study and was one focus for primer design. Primers were designed from the sequences found in these two clades and from other human-specific clades. A promising primer in the latter half of the 16S rRNA gene targeted human-specific fecal bacteria in both Oregon and Alaskan populations. Humans in the isolated Alaskan population may differ in their fecal bacteria for reasons of geographic isolation, population history, or diet.

©Copyright by Carolyn A. Petersen
September 5, 2006
All Rights Reserved

Analysis of *Bacteroidales* 16S rRNA Gene Sequences from a Geographically Isolated Human Population

By

Carolyn A. Petersen

A THESIS

Submitted to

Oregon State University

In partial fulfillment of

the requirements for the

degree of

Master of Science

Presented September 5, 2006

Commencement June 2007

Master of Science thesis of Carolyn A. Petersen presented on Septmeber 5, 2006.

APPROVED:

Major Professor, representing Microbiology

Chair of the Department of Microbiology

Dean of the Graduate School

I understand that my thesis will become part of a permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Carolyn A. Petersen, Author

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. Katherine Field, who provided me with advice and guidance as I pursued my degree. I would also like to thank my committee members, Drs. Peter Bottomley, Stephen Giovannoni, and Nancy Kerkvliet, who provided teaching and advice. I thank Linda Dick, who provided me with valuable advice and tutelage. I thank my lab mate Sarah Walters for providing me with advice and acting as sounding board when I needed it.

I would also like to thank my friends and family for supporting me through this process. Without having people to encourage me, make me laugh, and keep things into perspective I would have never made it.

CONTRIBUTION OF AUTHORS

All experiments were performed in the laboratory of Dr. Katherine Field, who was involved in the preparation and editing of this manuscript. Anne Bernhard, Linda Dick, and Sarah Walters provided host-specific *Bacteroides*-specific 16S rRNA clone library sequences for analysis.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

LIST OF FIGURES

LIST OF TABLES

Chapter 1. INTRODUCTION

Water is a vital resource, with safe drinking and recreational water an essential component of any community. In 1972 the Clean Water Act established water quality standards that all water in the United States must meet in order to be deemed "fishable and swimmable" (68). Monitoring is done continually to guarantee that water meets these standards, not just for drinking and recreational water, but to ensure environmental quality. Although standards and monitoring water quality are in place, problems still exist. For example, in 2005 there was a 5% increase from 2004 in pollution-related beach closings, for a total of 20,000 days across the country of closed beaches at the oceans, bays, and Great Lakes (15).

One area of the U. S. that has continually struggled with clean drinking water is Alaska, particularly parts of rural Alaska where there are no piped water or sewer systems. One such rural Alaskan village inhabited by native Inuits was the focus of this study. This unnamed village underwent extensive surveying by researchers from the University of Alaska, to develop a greater understanding of water source use and storage, and to research ways in which drinking water sanitation could be improved. In these surveys, the researchers found coliforms and *Escherichia coli* to be almost ubiquitous along the main road and boardwalk and wherever pools of water existed within the community (72).

Since the researchers established the presence of fecal pollution, it was the task of our lab to determine the host source responsible for the contamination. The likely candidates were dog and human because they are present in the greatest numbers. Almost everyone in the community owns a dog that is kept outside at all times, and they often roam the village unfettered, so contamination can be quite easy. Fecal contamination from humans is also likely, considering the manner in which human waste is disposed of in the community. Households share outdoor privies in which waste is collected in "honeybuckets", which are emptied into shared collection hoppers that are in turn emptied into a sewage lagoon just outside of town. The collection hoppers are transported by local teens or adults by snowmobile or all-terrain vehicle and spills are not unheard of while transporting to the lagoon.

The overall goal of this study was to identify the host source responsible for fecal contamination. To accomplish that goal, this study had four main objectives. The first objective was to analyze water samples and fecal samples with host-specific *Bacteroidales* 16S rRNA primers. The second objective was to construct and analyze 16S rRNA gene libraries from Oregonian and Alaskan fecal samples. This led directly to the third objective, which was to design and test novel host-specific *Bacteroidales* 16S rRNA primers for fecal source identification. Finally, the fourth objective was to analyze Alaskan water samples with the newly designed primers.

Chapter 2. LITERATURE REVIEW

2.1. Water Quality

Water quality can be impaired by a variety of pollutants including thermal, chemical, pathogen, metal, and a host of other contaminants. Once introduced, entire communities can be exposed to these contaminants in surface, recreational, and ground water. One of the greatest concerns is the health risk associated with fecally contaminated water and the economic losses due to closing beaches and rivers to recreational use.

Remediation of contaminated water is complicated when there are conflicts between user groups and their different demands on water resources. For example, Tillamook Bay, OR, is frequently contaminated with fecal pollution but there is debate as to its source. Likely sources include leaky sewer systems from the towns, or runoff from the dairy farms surrounding the bay (67). Knowing the source of contamination is vital to then solving the problem. However, it is notoriously difficult to identify and quantify the source of fecal pollution, causing opposing groups to erroneously identify the sources of contamination without any scientific basis (58).

Indicator organisms are used to identify fecal pollution. There are several key characteristics of an ideal indicator: dense enough to detect, easy to enumerate, associated with pathogens, unable to proliferate in the environment, and released into the environment solely by warm-blooded animals (64). The

indicator organisms specified by the U.S. Environmental Protection Agency are *Escherichia coli* for fresh waters and enterococci for marine waters (1). While they are easy to enumerate, detect, and are associated with pathogens, there are several reasons why they fall short in accurately assessing water quality.

First of all, *E. coli* and enterococci fail to identify the source of contamination. A variety of warm-blooded, and even some cold-blooded, animals shed *E. coli* and enterococci in their feces (37). Additionally, health risks associated with fecal pollution have been traditionally centered on human fecal pollution. Domestic, agricultural and wildlife fecal pollution can spread zoonotic pathogens including *Salmonella*, *E. coli*, *Campylobacter jejuni*, *Giardia* spp., *Cryptosporidium* spp., *Leptospirosis* and hepatitis E. Indicator counts from different sources of fecal contamination are combined, although the associated pathogens vary depending on which sources are present (23).

Another shortcoming of *E. coli* and enterococci is that the estimation of the contribution of *E. coli* and other indicator organisms is difficult to determine, because different species contain different numbers and different relative proportions of *E. coli* and enterococci in their feces (2, 23, 26, 73). Perhaps the biggest disadvantage of *E. coli* and enterococci is their ability to survive and persist in the environment, making it difficult to determine whether identification and quantification is the result of recent fecal contamination or persistence from an earlier contamination (75, 4).

2.2. Microbial Source Tracking

Microbial source tracking allows investigators to identify the source host(s) responsible for fecal contamination in waters and/or to quantify the amount present. Knowing what host(s) caused contamination may also suggest the types of associated pathogens. Currently, there are a number of methods, both microbiological and non-microbiological, employed in microbial source tracking using *E. coli*, enterococci, and other species, and each having unique advantages and disadvantages.

2.3. Non-Microbiological Methods

2.3.1. Caffeine

Detecting the presence of caffeine has been used to indicate human fecal pollution. Several methods have been used to detect the presence of caffeine such as liquid and gas chromatography (8, 74), capillary electrophoresis (45) or liquid chromatography coupled to mass spectrometry (7, 47). While caffeine detection has been successfully used to identify human fecal pollution, that is also one of its biggest limitations—the method only detects human fecal pollution. Nonpoint sources, like agricultural runoff, or wildlife, cannot be determined by the detection of caffeine. Another limitation is that there is no a priori reason why caffeine would be expected to correlate with pathogens.

2.3.2. Sterols

Fecal steroids, most notably coprostanol (5-cholestan-3-ol), have been used to detect fecal contamination (39, 60, 5, 62). Often the detection of fecal steroids is used to complement fecal coliform counting methods (60). Coprostanol makes a good indicator because it can be quantitatively related to the amount of sewage-derived organic matter (39). Due to their poor water solubility, fecal steroids are often collected from sediments, which can limit the use of this method in deep bodies of water such as lakes and oceans. Fecal sterols are found in the feces of higher animals, making the determination of the host source impossible from this method alone. Additionally, pathogens would not necessarily be correlated with these chemicals.

2.3.3. Other Chemicals

Another method used in the past to detect traces of human fecal contamination is by identifying wastewater indicator compounds such as ethylene diaminetetraacetic acid (EDTA), nitrilotriacetic acid (NTA), a naphthalene dicarboxylate (NDC) isomer, alkylophenol polyethoxy carboxylates (APECs), and select haloacetic acids (HAAs) (20). These compounds found in sewage effluent can be identified and quantified by gas chromatography/mass spectrometry (69). While this method has proven effective at identifying human fecal contamination from wastewater, it does not detect other nonpoint sources of contamination, making its application of limited value.

2.4. Microbiological Methods

2.4.1. Phage

Phages, viruses that infect bacteria, have been used to track the source of fecal contamination. F-specific coliphages, which primarily infect gram-negative bacteria, have been used to help discriminate between human and non-human sources of fecal pollution in water samples (40, 41). Another study found that the identification of phages that infect *Bacteroides fragilis* increases the probability of also finding pathogens (59). Long and colleagues (54) found that F-specific coliphages were detected more often in wastewater and slurry lagoon samples and no one subtype of F-specific coliphage could be associated with human or animal waste. These studies suggest that phages could be a potential indicator of fecal contamination, but would not differentiate between host sources, making the use of the method limited.

2.4.2. Fecal Coliform/Fecal Streptococci Ratios

Geldreich and colleagues (28, 29) found that human feces have a ratio of fecal coliform (FC) to fecal streptococci (FS) greater than or equal to 4.0, whereas ratios below 0.7 are associated with animal feces. However, another study showed that the FC/FS ratio could not discriminate between human and domesticated animal fecal samples (46). This method has also been criticized because of the difference in fecal enterococci densities found in individuals with different diets and the environmental factors affecting the survival of coliforms and streptococci bacteria (68). Because of these criticisms the method's use has

dramatically decreased or is complemented with other source tracking methods to bolster the reliability of classification.

2.5. Library-Dependent Methods

Microbial source tracking methods can be divided into culture-dependent and culture-independent methods, with some requiring a library. A library is a set of bacterial isolates from fecal samples of known origin. Most library methods require growing environmental isolates from water samples and as such are culture-dependent and time-consuming (23). Culture-dependent methods are inherently biased because of the inability to culture fastidious microbes. Fecal bacteria fall into this category, as they are difficult to culture because most are anaerobic, have unknown nutrient requirements, and may even require the synergistic benefits of other bacteria. For this reason, library-dependent methods all utilize the easily grown fecal indicator bacteria *E. coli* and enterococci. In addition, library-dependent methods are not geographically stable, and therefore most require constructing of a new library for each new area studied (23).

The statistic most often used when using library-dependent methods is the average rate of correct classification (ARCC), more recently called internal accuracy (12). The ARCC and percent of misclassification was derived by using a classification table produced by statistical methods (76). The table is a source-by-source matrix in which the numbers and percentages of correctly classified isolates are found on the diagonal (77). The ARCC for a given combination of

antibiotics was computed by averaging the percentages along the diagonal (76). In a study that examined compared the library size the researchers found that small libraries had higher rates of correct classification, but were less able to correctly classify nonlibrary isolates (78).

However, this statistic can be high if the library is small and does not represent the true diversity in the system. Harwood and colleagues (38) found that the ARCC (a.k.a. internal accuracy) did not correlate with the accuracy of source prediction in water samples. The ARCC should not be solely relied on to predict classification and any study that uses the ARCC should address the issues associated with this statistic before relying on the data provided. The validity of this statistic is not only limited by the size of the library, but also whether it is representative of the diversity being studied. Additionally, the statistic is not reliable across geographic areas, and is time-consuming to derive.

2.5.1. Antibiotic Resistance Analysis

Antibiotic resistance analysis (ARA) is a phenotypic-based method based on patterns of antibiotic resistance of bacteria from human and animal sources (reviewed in (58, 68, 70). The assumption for ARA is that various animal populations are exposed to and have developed resistance to different arrays of antibiotics; therefore, antibiotic resistance patterns can be used to differentiate fecal bacteria from different animal or human sources. One advantage of ARA is that hundreds of isolates can be processed in a week. Because of the decreased correct classification when examining unknown isolates, ARA is a better method

for small-scale studies where the possible sources of contamination are known as in simple watersheds. However, in two studies comparing microbial source tracking methods ARA proved to be virtually useless in differentiating host sources (69).

2.5.2. Carbon Utilization Profile

Another phenotypic-based method is based on the diet requirements of the various fecal bacteria. The rationale behind this method is that the diets of the animal hosts have shaped the evolution of the various gut microbes to utilize different carbon and nitrogen sources for energy and growth (70). The BIOLOG system allows for rapid performance, scoring, and tabulation of 96 carbon source tests per isolate (35). Hagedorn and colleagues (35) used carbon utilization profiles (CUP) to identify sources of fecal pollution in water. However, environmental factors can affect bacterial nutrient requirements, making the BIOLOG system unreliable for field determination (68). In comparative microbial source tracking studies CUP performed poorly in correctly identifying host sources of fecal contamination (70).

2.5.3. Ribotyping

Ribotyping generates a molecular fingerprint based on genomic 16S rRNA restriction fragment length polymorphisms (70). This method involves matching the "fingerprint" patterns from known sources of feces in the library to patterns from water isolates. In this case, the size of the library is very important as is the

statistical analysis used to develop the fingerprint. Samadpour and coworkers (66) demonstrated that choosing restriction enzymes to develop a fingerprint is critical; in fact, double enzyme analyses should always be performed, as they are more accurate than a single enzyme digestion. Ribotyping has been successfully used for epidemiology of food outbreaks, and while microbial source tracking is considerably more complex, it has proven a reliable method (14, 36, 61).

2.5.4. Pulse-Field Gel Electrophoresis

Pulse-field gel electrophoresis (PFGE) is similar to ribotyping in that banding patterns are analyzed after restriction digest, except that instead of just 16 rDNA, the whole DNA genome is restricted (48). After digestion the DNA is imbedded into specialized electrophoresis gels and electrophoresed for an extended period of time with alternating currents from different directions using specialized equipment (48). Tynkkynen and colleagues (71) compared ribotyping and PFGE for typing two strains of *Lactobacillus* and found that PFGE was the most discriminatory method. In a comparative study of microbial source tracking methods only PFGE correctly classified all isolates into the correct species-level category (58). As with ribotyping adding more enzymes increases accuracy of classification; however, PFGE is time-consuming and the number of isolates that can be analyzed simultaneously is limited.

2.5.5. Denaturing-Gradient Gel Electrophoresis

Denaturing gradient gel electrophoresis (DGGE) is another electrophoretic technique that separates PCR products of similar size that differ in base sequence based on changes in electrophoretic mobility, influenced by the melting properties of the DNA fragments (58). Farnleitner and colleagues (22) showed that DGGE could detect and differentiate *E. coli* populations from freshwater samples polluted with fecal matter, but did not use DGGE to differentiate between sources. However, Buchan and coworkers (13) used DGGE to differentiate environmental *E. coli* isolates from three host sources (bovine, human, and poultry) but were unable to identify the source of contamination in the watershed studied. High levels of genetic diversity in environmental isolates make DGGE unreliable for use in the field, although it is useful for differentiating the strain level of bacteria.

2.5.6. Repetitive PCR

Repetitive extragenic palindromic polymerase chain reaction (REP-PCR) uses conserved sequences in bacterial repetitive elements as PCR primers to distinguish among different strains of the same bacterial species (reviewed in (68). REP-PCR has been used to examine fecal bacteria strains isolated from different sources of fecal pollution (23). REP-PCR has not performed as well as ribotyping and PFGE in microbial source tracking studies and has the same limitations since it is library-dependent (23).

2.5.6. Amplified Fragment Length Polymorphism

Amplified fragment length polymorphism (AFLP) uses RFLP and PCR amplification to generate between 50 and 100 DNA fragments that are commonly analyzed by DNA sequencers containing fluorescence-based detectors (reviewed in (68)). Using DNA sequencers allows for automation, so that over one hundred strains can be analyzed per day. Based on ARCC, Guan and colleagues (33) found AFLP to be the most effective of three methods tested—ARA, host-specific *E. coli* 16S rRNA PCR, and ALFP—when discriminating among *E. coli* isolates from animal and human sources. However, the ARCC statistic does not predict a method's ability to classify isolates from outside the library, but instead is inflated for small libraries. This method requires a large isolate library to rely on the accuracy of the results; otherwise, the high accuracy of discriminating host sources could be due to a small library that is not diverse.

2.6. Library-Independent Methods

Library-independent methods for microbial source tracking have a distinct advantage over library-dependent methods, as they do not rely on culturing bacteria. Problems with viable culture methods include maintaining the viability of bacteria between the time of collection and enumeration, and lack of growth of viable but nonculturable bacteria. These methods rely on molecular biological techniques and only require small amounts of DNA to detect fecal bacteria in fecal and water samples. In several studies comparing the abilities of

microbial source tracking methods, library-independent methods outperform library-dependent methods (24), (32). In addition to being more reliable, library-independent methods provide results in hours, not days as with library-dependent methods. These methods are usually cheaper, less labor-intensive, and often host-specific. A number of methods have taken advantage of the host-specificity of fecal bacteria genes and use host-specific primers designed from these genes.

2.6.1. Coliform-Specific PCR

The *lacZ* and *lamB* genes found in coliforms, which include members of the genera *Escherichia, Enterobacter, Citrobacter,* and *Klebsiella*, have been used to detect fecal contamination by PCR and gene probes (6). While coliforms are relatively easy to culture, culture-independent methods of detection can theoretically detect more, if not all, the coliforms present. Bej and coworkers (6) found that PCR amplification of *lacZ* and *lamB* was capable of detecting as little as 1 to 10 fg of genomic *E. coli* DNA in 100 ml of water, and that amplification of *lamB* permitted detection of *E. coli* and enteric pathogens such as *Salmonella* and *Shigella* spp. This method, while useful, does not discriminate between host sources of fecal contamination, which limits it use.

2.6.2. *E. coli* toxin gene-specific PCR

One method for detecting the presence of cow fecal contamination is based on PCR primers that detect a portion of the heat labile toxin IIA (*LTIIa*)

gene from enterotoxigenic (ETEC) *E. coli* (49). This cow biomarker method proved highly specific for cattle fecal contamination in water and was able to detect pollution between 0-3 weeks in age in various water types (41). Another toxin gene-specific PCR, amplifying a portion of the *STIb* gene, has been used to detect human fecal contamination; however a comparative study found *STIb* in 1 of 4 gull fecal samples and 1 of 12 dog fecal samples (24). In this study the *STIb* assay correctly identified the presence of human fecal contamination in all unknown samples containing sewage effluent, but it only detected 2 of 4 unknown samples containing human feces (24). This suggests that this method is better suited to detecting contamination from sewage and not from single human fecal sources. Using *E. coli* toxin gene-specific PCR to detect sources of fecal contamination has proved successful for these 2 toxin genes; however, the number of hosts that can be detected is limited and cross-reactivity is a constant worry and must always be tested for when using this method. Also, it is important to note that this assay was not culture-independent, as the occurrence of the gene is so rare that samples must be enriched for *E. coli* before the gene can be detected.

2.6.3. Community Terminal Restriction Fragment Length Polymorphism

Terminal restriction fragment length polymorphism (T-RFLP) separates fluorescently labeled 16S rDNA PCR products for host specific genetic markers based on length differences using an automated DNA sequencer (9, 10). In a comparative study, the 16S rRNA gene was amplified by PCR with the forward

primer being fluorescently labeled (20). The PCR products were then purified and digested with *HhaI* and the fluorescently labeled terminal restriction fragments (TRFs) were separated and fragment lengths determined by capillary electrophoresis (20). In this comparative study, community T-RFLP correctly identified human contamination and was the only method capable of detecting all sources; however it produced many false positives (20).

With T-RFLP it is important to use multiple restriction enzymes as banding patterns tend to be underestimated when using just one restriction enzyme, especially when restricting such a highly conserved gene like the 16S rDNA gene (10). Also, the amount of DNA yield may influence the results and limit the effectiveness of this method. In the comparative study the DNA yield varied greatly between samples and most yielded insufficient DNA for community analysis (< 5 ng/µl) (20).

2.6.4. Length Heterogeneity PCR

In length heterogeneity PCR (LH-PCR) a fluorescently labeled primer is used to determine the relative amounts of amplified sequences originating from different microorganisms (65). Labeled fragments are separated by gel electrophoresis and detected by laser-induced fluorescence with an automated gene sequencer (65). Bernhard and Field (10) designed primers from *Bacteroides* or *Bifidobacterium* 16S rRNA genes that were specific to cows and humans, respectively. In this study they were able to identify the host source responsible for fecal contamination in Tillamook Bay, OR (10).

2.6.5. *Bacteroides*-specific PCR

Bacteroides-specific PCR takes advantage of the host-specificity of the 16S rRNA gene and the ubiquity of *Bacteroides* in the guts of warm-blooded animals. Several host-specific primers have been developed that allow for the identification of host source fecal contamination in water (19). These primers detect the rRNA gene sequences specific to anaerobic fecal bacteria, specifically *Bacteroidetes* and *Bifidobacterium*. *Bacteroidetes* and *Bifidobacterium* are present in higher densities in animal feces than conventional indicator species and are thought not to proliferate in the environment (9, 10).

Bacteroides-specific PCR is an ideal method for microbial source tracking for two main reasons. One, it is a library-independent method, which in several comparative studies out-performed library-dependent methods (21, 28, 47). Secondly, in the same comparative studies, host-specific PCR out-performed other library-independent methods (21, 27, 48). Ribotyping and PFGE also performed well, but PCR provides results more rapidly. Also, ribotyping and PFGE rely on restriction digest, which is more problematic since different enzymes and different numbers of enzymes produce different results. *Bacteroides* makes a better indicator organism because these bacteria are more abundant in feces of warm-blooded animals than *E. coli* and are unlikely to survive outside the intestinal tract (3, 25). Combining the superior indicator fecal bacteria, *Bacteroides*, with the preferred microbial source tracking method, host-specific

PCR, guarantees an ideal method for detecting and identifying the host source of fecal contamination in water.

2.7. *Bacteroidetes* Fecal Anaerobes

Fecal members of the phylum and class *Bacteroidetes*, order *Bacteroidales*, include the genera *Bacteroides* and *Prevotella* (27) . They are anaerobic, saccharolytic, Gram-negative, nonsporulating rods found predominantly in the large intestine, where they transform bile acids, degrade proteins, and ferment polysaccharides (27). Members of the *Bacteroidales* order are among the most numerous intestinal microflora of warm-blooded animals (18, 25).

Because of their need for anaerobic conditions, members of the *Bacteroides-Prevotella* group do not survive long in waters, and as such their presence indicates recent fecal contamination (50). Members of the *Bacteroides-Prevotella* group are notoriously difficult to culture, but molecular methods have bypassed the need for cultivation. A significant diversity of uncultured representatives from *Bacteroides-Prevotella* 16S rDNA sequences are present in animal feces (10, 16, 42). The ecological and phenotypic diversity of the *Bacteroides-Prevotella* group could be due to the ability of the group to adapt to a particular host or environment (31). Several studies have also shown that some species of fecal *Bacteroides* have host-specific distributions (50, 10, 18).

The *Bacteroides-Prevotella* group has several qualities of a good fecal indicator microorganism. They are dense enough to detect; in fact, members of the *Bacteroides fragilis* group are present in 1,000-fold higher numbers per gram of

human feces than fecal coliform bacteria (25). Since the presence of the *Bacteroides-Prevotella* group indicates fecal pollution there is also an associated potential risk of human pathogens such as hepatitis A and/or pathogenic *E. coli* (30).

Members of this group are unlikely to proliferate in the environment as well as traditional aerobic indicator organisms like *E. coli* and enterococci. For example, it was found that aquatic sediments provided ideal nutrient conditions for *E. coli* and enterococci, providing protection from sunlight inactivation and protozoan grazing (75, 17, 4). These conditions led to high counts of indicator bacteria at all depths in sand. Additionally, laboratory cultures of *E. coli* were better able to donate plasmids in the dark (as in sediments) than those in the light (75). However, temperature and predation influenced the persistence of PCR-detectable DNA from *Bacteroides disasonis* (51). At lower temperatures, when predators and degradative processes are less active, the PCR target persisted for an extended time, indicating that seasonal variations must be considered (51).

Molecular methods also have an advantage over culture-dependent methods in being much more rapid, providing a nearly real-time answer to the question of fecal contamination. Traditional methods relying on cultivation of indicator organisms can take 24-72 hours, which can have a greater impact on the economic loss due to beach closures.

2.8. The Role and History of Commensal Microbes

Humans have evolved in a world dominated by microbes (43). An example of this is that the human genome encodes 223 proteins with significant homology to bacterial but not eukaryotic proteins, suggesting that they were acquired through horizontal transfer of bacterial genes (43). Traditionally, these commensals have long been thought to simply provide more nutrients to their hosts, but the relationship is much more complex than that.

For example, there is evidence that commensals acquired during the early postnatal period are required for the development of tolerance not only to themselves but also to other luminal antigens (44). Hooper and Gordon (43) also found that commensals modulate the expression of genes involved in several important intestinal functions, including nutrient absorption, mucosal barrier fortification, xenobiotic metabolism, angiogenesis, and postnatal intestinal maturation. Rakoff-Nahoum and coworkers (63) found that commensals are recognized by toll-like receptors (TLRs) under normal steady-state conditions, and this interaction plays a crucial role in the maintenance of intestinal epithelial homeostasis. This study also uncovered that activation of TLRs by commensals protects epithelial cells from injury and mortality (63).

Through studies of germ-free and conventional mice, Wolf (79) found that commensals are an important regulator of energy uptake and storage. In this study it was discovered that the commensals caused the suppression of a circulating inhibitor of lipoprotein lipase, resulting in increased lipoprotein

lipase activity and thus fat deposition (79). These effects resulted in a significantly increased body fat deposition in germ-free compared to conventional mice (79).

Additionally, Mazmanian and colleagues (57) found that during colonization of animals with *Bacteroides fragilis*, a bacterial polysaccharide directs the cellular and physical maturation of the developing immune system. During colonization, this bacterial polysaccharide corrects systemic T cell deficiencies and T_H1/T_H2 imbalances and directs lymphoid organogenesis (57).

Molecular methods, taking advantage of the highly conserved 16S rDNA gene sequence, are now used to characterize and monitor the microbial communities in humans. Applying molecular phylogenetic methods to study the complex microbial ecosystem of the human gut has revealed unexpected evolutionary lineages and in the case of *Helicobacter pylori* has helped trace our own lineage. Studies of *H. pylori* found all modern strains can be traced to 5 ancient populations, and genetic variations in *H. pylori* can be used to trace human settlement and migration patterns over the past 60,000 years (11, 21). There are few studies that have investigated the evolutionary history of fecal commensals, as they have been difficult to study; instead pathogenic microbes have been the focus of research. Also, it was assumed that the composition and diversity of fecal commensals is similar no matter the population under question. However, the results of this study show that there are differences in commensal

population, which may be due to geographic isolation, population history, or diet.

2.9. Alaska

Contamination of drinking water stored in home water containers has been an ongoing problem in rural Alaskan villages, but there is little understanding where the contamination is coming from and how it is spread (72). The rural Alaskan village in this study lacks a piped water and sewer system. Eighty-three percent of the population regularly use untreated water sources such as ice and rainwater either as their main source or to supplement treated water from the washeteria (72). The washeteria is a facility where residents can go and pay 25 cents for a gallon of treated water. Most use the facility for washing and cleaning, not for drinking as cost, taste, and appearance of the water factors into the decision (72). Also, rainwater can be collected from roof catchments providing ready access (72).

In the home, water is stored in a variety of containers, with most households using a dedicated non-food grade 35-gallon trash can (dipbucket) and dipper (60). Also, most families do not treat water collected from traditional sources, although occasionally it may be boiled or filtered (60). Sixty-two percent of households report that they do not sterilize the container between refills and water is often stored for a week or more in the dipbuckets, underscoring the high risk of contamination (60). Solid wastes are collected in honeybuckets from

outdoor privies and hand-carried several hundred yards to a collection hopper. When full, the collection hoppers are transported by ATV or snowmobile to a honeybucket lagoon. There are several opportunities for contamination into the village and home through the emptying of the honeybuckets and collection hoppers.

Broad sweep sampling was done in this village to detect the total coliforms and *E. coli* through Colilert presence/absence tests and follow-up Colilert Most Probable Number counts of positive locations (IDEXX Laboratories, Inc., Westbrook, ME). Swabs were taken from the boardwalk, the bottoms of shoes, dog paws, dog fur, and ATV tires returning from the dump and ATV tires around town to test for the presence/absence of *E. coli* (60). From this sampling, total coliform was ubiquitous, except in the drier areas tested like the roads (60). *E. coli* presence was less uniform but still prevalent in many areas (60). Dog paws tended to be positive for *E. coli* while their fur was less likely to be positive (60). ATV tires were positive more often returning from the dump than around town, while all boardwalk surfaces were negative, even though shoes carried *E. coli* 50% of the time (60).

Based on the results of these Colilert presence/absence tests and follow-up Colilert Most Probable Number counts, there was a high incidence of *E. coli* contamination, well above background levels, in the surrounding river, lakes, and tundra ponds. Contamination points corresponded for the most part to

locations within the community, or between the village site, honeybucket lagoon dump and dump.

While the presence of coliforms and *E. coli* highlighted the ubiquitous nature of fecal contamination it did not identify the source of the fecal contamination. Researchers at the University of Alaska thought that the two most likely sources were humans and dogs since both were present in the greatest numbers. In particular, the manner in which solid waste is disposed of in the community could readily explain how human fecal waste could contaminate the environment. Our lab was provided with fecal samples from dogs and humans, as well as water samples, and charged with identifying the host source contaminating the water samples, if possible.

Chapter 3. MATERIALS AND METHODS

Fecal and Water Sample Analysis

3.1 Sample Collection

All fecal and water samples from Alaska were kindly donated by Malcolm Ford of the University of Alaska.

Human fecal samples: Human fecal samples were collected from 2 geographic regions—Oregon and Alaska. The Alaskan samples were from a geographically isolated Alaskan village, inhabited by native Inuits. The human fecal samples were household samples taken from buckets shared by all members of 10 different households. The samples were kept on ice for transport to the lab, and stored at -80°C until extraction. The Oregon samples were collected from 9 individual healthy adult volunteers from Corvallis, Oregon and stored at –80°C until extraction.

Animal fecal samples: Dog fecal samples were collected from 10 individual dogs from a native Inuit village in southwest Alaska. The samples were kept on ice for transport to the lab, and stored at -80°C until extraction. Cat, cow, horse, elk, sheep, pig, chicken, and gull fecal samples were collected and extracted as previously described by Bernhard and colleagues and Dick and colleagues (9, 18).

Water samples: Alaskan samples included 11 water samples from coliform and *E. coli* positive puddles within the village. Water samples were filtered by

syringe through 0.2 µm Supor-200 filters. Filters were placed in sterile 50 ml centrifuge tubes with 5 ml of guanidine isothiocyanate buffer (5M guanidine isothiocyanate, 100 mM EDTA [pH 8.0], and 0.5% Sarkosyl), stored on ice, and transported to the lab. Samples were stored at -80°C until DNA extraction was completed.

3.2. DNA Extraction

Fecal samples: Bacterial DNA was extracted from 300 mg of each fecal sample using the Bio 101 Systems kit (Qbiogene, Carlsbad, CA) with the following modifications: 300 mg of protein and cell debris were pelleted for 10 minutes at 14,000 x g, and 3 washes were done with SEWS-M buffer.

Water samples: Bacterial DNA from water samples was extracted using the Qiagen high-throughput Dneasy 96 Tissue kit (Qiagen, Valencia, CA) with the following modifications: seven hundred microliters of buffer AL/E was added to the filters preserved in GITC and vortexed for 1 minute, five hundred microliters of AW1 buffer was added, followed by the addition of 500 µL of AW2 buffer, and after a 15 minute incubation at 70°C the DNA was eluted with 100 µL pre-heated AE elution buffer. All centrifugation was done at 4,700 rpm for 15 minutes. The eluted DNA was stored at -20°C.

3.3. Polymerase Chain Reaction

Table 3.1. shows the sequences of PCR primers used in this study (53). All PCR reactions were carried out in a 25 µL volume with appropriate template and

targets. PCR mix #1 was used for amplification of all DNA with the general *Bacteroidales* primer set *Bac*32F and *Bac*708R and contained 2.5X *Taq* polymerase buffer (TaKaRa, Shiga, Japan), 0.2 μM for each primer, 200 μM for each dNTP, 0.125 units of *Taq* polymerase (TaKaRa), 0.4% BSA and 1.5 mM $MgCl_2$. PCR reactions done with this primer set were carried out for 30 cycles of 94°C for 1 min, 53°C for 30 s, and 72°C for 1 min.

PCR mix #2 was used for amplification of fecal and water samples, and contained 2.5X *Taq* polymerase buffer, 0.2 μM for each primer, 200 μM for each dNTP, 0.125 units of *Taq* polymerase, 0.4% BSA, and 2.0 mM $MgCl_2$. PCR mix #2 was used for 3 different primer sets: HF134F and HF183F both coupled with *Bac*708R, and DF475F coupled with *Bac*708R. The PCR reactions used to amplify the Alaskan water samples with the human-specific primers HF134F and HF183 coupled with *Bac*708R were carried out for 35 cycles of 94°C for 1 min, 63°C for 30 s, and 72° for 1 min. The PCR reactions used to amplify the Alaskan water samples with the dog-specific primer set DF475F coupled with *Bac*708R were carried out for 35 cycles of 94°C for 1 min, 62°C for 30 s, and 72°C for 1 min.

PCR reactions used to amplify the Alaskan human, Oregonian human, and dog fecal samples were carried out under the same conditions using their respective primer sets except that the PCR reactions were carried out for 30 cycles as opposed to 35 cycles. All PCR products were separated using 1.5% agarose gel electrophoresis stained with ethidium bromide. The gels were run for 30 min at 100 volts, and band sizes were estimated with a 100 bp DNA ladder

(Fermentas, Amerherst, NY). Bands were visualized and recorded with a UVP gel imager (UVP, Upland, CA).

Table 3.1: Primers used in this study

Primer[a]	Sequence (5'-3')	Target	Reference
Bac 32F	AACGCTAGCTACAGGCTT	*Bacteroides-Prevotella*	9
Bac 708R	CAATCGGAGTTCTTCGTG	*Bacteroides-Prevotella*	9
HF134F	GCCGTCTACTCTTGGCC	Human *Bacteroides-Prevotella*	9
HF183F	ATCATGAGTTCACATGTCCG	Human *Bacteroides-Prevotella*	9
DF475F	CGCTTGTATGTACCGGTACG	Dog *Bacteroides-Prevotella*	18
Bac 1492R	GGTTACCTTGTTACGACTT	Universal	52
HF185F	GGCATGGTGGAACTATTA	Human *Bacteroides-Prevotella*	This study
HF1001R	CTGTCCGAAGAAAGAACC	Human *Bacteroides-Prevotella*	This study

[a]The numbers correspond to numbers in the *E. coli* 16S rRNA gene

Construction and Analysis of 16S rRNA Gene Libraries

3.4. Gene Library Construction

Two sequence libraries were constructed to investigate the 16S rRNA gene sequences of Alaskan and Oregon fecal samples (52). The 9 Oregonian human and 10 Alaskan human fecal samples were amplified with the general *Bacteroidales* 16S rRNA primer *Bac*32F and the *Bacteroidales* 16S rRNA primer *Bac*1492R using PCR mix #2 to generate a 1,460 bp fragment of the 16S rDNA gene. PCR reactions were carried for 30 cycles of 94°C for 1 min, 53°C for 30s, and 72°C for 1.5 min. The PCR products were separated and visualized as above. DNA concentrations were estimated by comparing the band intensity to a low molecular weight DNA mass ladder (Fermentas). After determining the concentration of the amplified DNA the 10 Alaskan PCR products were pooled in equal concentrations to a final volume of 50 µL, with 10 µL set aside in order to determine any yield loss. The 9 Oregon PCR products were also pooled in equal concentrations to a final volume of 50 µL, with 10 µL set aside. Forty µL of the 2 pools of PCR amplified fecal DNA was separated on a 1.5% agarose gel, the band excised and gel-purified using the QIAquick PCR purification kit (Qiagen). To compare the yield of the gel-purified fecal DNA to the original pool of fecal DNA, the 10 µL initially set aside was run on a 1.5% agarose gel with 10 µL of the gel-purified fecal DNA and quantified by comparing the band intensity to a low molecular weight DNA mass ladder (Fermentas).

To ligate the DNA insert into a plasmid vector, 4 μL of the gel-purified fecal DNA was mixed with 1 μL of TOPO TA vector (Invitrogen, Carlsbad, CA) and incubated for 30 minutes at room temperature. Transformation was done as per the manufacturer's instructions with some modifications. Two tubes of competent *E. coli* cells were thawed, mixed and aliquoted in equal volumes into 3 different 2 ml screw-cap tubes. To one tube of competent *E.coli* cells 2 μL of the plasmid vector was added. The remaining two tubes were designated as controls, one with 2 μL sterile water and the other with 2 μL of pUC plasmid. The *E. coli* cell solution was allowed to incubate on ice for 5 minutes and then heat-shocked at 42°C for 1 minute to transform the plasmid vector into the competent *E. coli* cells.

To each of the 3 tubes of transformed *E. coli* cells, 250 μL of S. O. C. medium was added and shaken at 200 rpm at 37° for 1 hour. Three different volumes—50 μL, 100 μL, and the remaining volume (~120 μL)—of the transformed *E. coli* cells grown in SOC media were plated on 3 separate Luria-Bertani (LB) plates containing 100 mg/ml ampicillin with 40 μL X-Gal spread. Fifty μL of cells were plated for the negative and positive control of *E. coli* cells. The plates were then incubated overnight at 37°C and stored at 4°C. Up to 96 colonies were picked from the 3 plates and placed into separate wells of a 96-well plate containing 100 μL of LB-Amp media. The 96-well plate was then incubated overnight at 37°C. Sequence libraries were constructed in this manner for both Oregon and Alaskan fecal DNA. Each sequence library was screened for the

desired 1,460 base pair DNA insert by amplifying each picked clone with the primers *Bac*32F and *Bac*1492R.

3.5. Restriction Digest

48 clones, 22 from the Oregon library and 26 from the Alaskan library, were chosen at random for restriction with the enzyme *Hae*III. From each clone chosen, 20 ng of DNA was digested with 6 units of enzyme. The restriction digest was allowed to continue overnight at 37°C to achieve complete digestion of the sample. Banding patterns were visualized on 3% agarose gels dyed with SYBR Safe (Molecular Probes, Eugene, OR). Banding patterns of distinct RFLP types were determined by measuring the migration distance of the size standard, from which a standard curve was calibrated. Restriction patterns were also mapped using *in silico* digests (www.restrictionmapper.org).

3.6. Gene Library Analysis

Representative samples of each RFLP banding pattern were sequenced at Central Services Laboratory, OSU on a ABI Prism 3730 Genetic Analyzer (Applied Biosystems, Foster City, CA) in both directions by amplification with T-7PF and M13R, whose complementary sequences flank the 1,460 bp insert on the plasmid. After this initial round of sequencing, 48 more clones were chosen at random between the two sequence libraries for sequencing on a ABI Prism 3730 Genetic Analyzer at the University of California, Davis in the same manner. Contigs were constructed from the resulting sequences using the program

CodonCodeAligner (version 1.3.4, Dedham, MA). All contigs were checked for validity with BLAST searches (National Center for Biotechnology Information website) and representative *Bacteroides* and *Prevotella* sequences were downloaded for comparison of phylogenetic relationship. Sequences were aligned and phylogenetic trees were constructed using ARB software (55).

Trees were inferred from 1,450 sequence positions using three tree-building programs: maximum likelihood, neighbor-joining with a Kimura-2 parameter correction, and maximum parsimony. Trees were constructed using described *Bacteroides* and *Prevotella* sequences downloaded from GenBank, the Alaska and Oregon sequences, and half-length sequences of the 16S rRNA gene from previous studies (19) (10). Bootstrap analysis was conducted to validate the branching patterns using 1,000 replicates.

Design and Testing of Novel Host-Specific Bacteroidales 16S rRNA Primers

3.7. Primer Design

Clades containing Alaska and/or Oregon sequences were mined for primers using the Probe Design function of the ARB software. The primers found were then tested *in silico* by using the ProbeMatch function in ARB and chosen based on whether they amplified Alaskan and Oregon sequences, but not other host species. Potential primers were then coupled with the primer *Bac*32F or *Bac*708R for use with *in silico* amplification, which was performed in MacVector. Primer pairs were tested *in silico* against Alaskan human, Oregonian human, cat, dog, cow, pig, horse, sheep, elk, chicken, and gull sequences to

determine if they would generate a PCR product with the different hosts. A search for matches to the new primers was conducted by utilizing the RDP database (56). Primers that generated a PCR product with Alaskan or Oregonian sequences *in silico*, but none of the other hosts, were ordered from Invitrogen and tested with *in vitro* amplification by PCR.

3.8. Fecal Pool Constructs

Individual fecal pools from 9 animals—cat, dog, cow, horse, elk, sheep, pig, chickens, and gull—were made by adding 3 ng/μL of fecal DNA extracted in previous studies (9, 18). Ten samples were pooled to constitute the cat pool, 12 samples were combined for the dog pool, 18 for the cow pool, 6 for the horse pool, 8 for the elk pool, 12 for the sheep pool, 5 for the pig pool, 7 for the chicken pool, and 5 for the gull pool. These fecal pools were constructed to use when testing novel primers.

All fecal pools were tested with the general *Bacteroidales* primer pair *Bac*32F and *Bac*708R to ensure that they could be successfully amplified. PCR mix #1 was used for this PCR reaction and carried out for 30 cycles of 94°C for 1 min, 53°C for 30 s, and 72°C for 1 min. All PCR products were separated and visualized as described earlier.

3.9. Primer Testing

Annealing temperature for each primer pair was optimized using a temperature gradient PCR reaction with plasmid DNA as the template. Using

the optimized PCR conditions for each primer pair, the primers were tested for specificity using the fecal pools. The detection limit for the new host-specific primer pairs was determined by generating dilutions of known plasmid template concentrations. The detection limit was determined using PCR mix #2. Using the new primer pair HF185F and *Bac*708R, PCR reactions were carried out for 30 cycles of 94°C for 1 min, 61°C for 30 s, and 72°C for 1.5 min. PCR reactions using the new primer pair *Bac*32F and HF1001R were carried out for 30 cycles of 94°C for 1 min, 63°C for 30 s, and 72°C for 1.5 min.

3.10. Alaskan Water Sample Analysis

The final step to determine the host source responsible for fecal contamination was to test the 2 new primer pairs, HF185F coupled with *Bac*708R and *Bac*32F coupled with HF1001R, on the contaminated Alaskan water samples. PCR mix #2 was used with both PCR reactions. With the primer pair HF185F and *Bac*708R, the PCR reactions were carried out for 35 cycles of 94°C for 1 min, 61°C for 30 s, and 72°C for 1.5 min. The PCR reactions with the primer pair *Bac*32F and HF1001R were carried out for 35 cycles of 94°C for 1 min, 63°C for 30 s, and 72°C for 1.5 min. All PCR products were separated and visualized as described earlier.

Chapter 4. RESULTS

Fecal and Water Sample Analysis

4.1. PCR Analysis

*Fecal samples***:** Each Alaskan household human fecal sample was successfully amplified using PCR with the general *Bacteroidales* primer pair *Bac*32F and *Bac*708R (Figure 4.1.). However, none of the Alaskan human fecal samples amplified with the human-specific primer pair HF183F and *Bac*708R; and only one of ten amplified with the other human-specific primer pair HF134F and *Bac*708R (Figure 4.2.1. & 4.2.2.). The Alaskan dog fecal samples also amplified successfully with the general primer pair *Bac*32F and *Bac*708R (data not shown). Nine out of the 10 Alaskan dog fecal samples were successfully amplified with the dog-specific primer set DF475F and *Bac*708R (data not shown).

Water samples: The 10 contaminated Alaskan water samples were successfully amplified with the primer set *Bac*32F and *Bac*708R, meaning that *Bacteroidales* bacteria was present and confirming fecal contamination (data not shown). In addition, the water samples were tested with both human-specific primer sets, HF183F and HF134F coupled with Bac708R, and the dog-specific primer set DF475F and *Bac*708R with no positive amplification observed (data not shown).

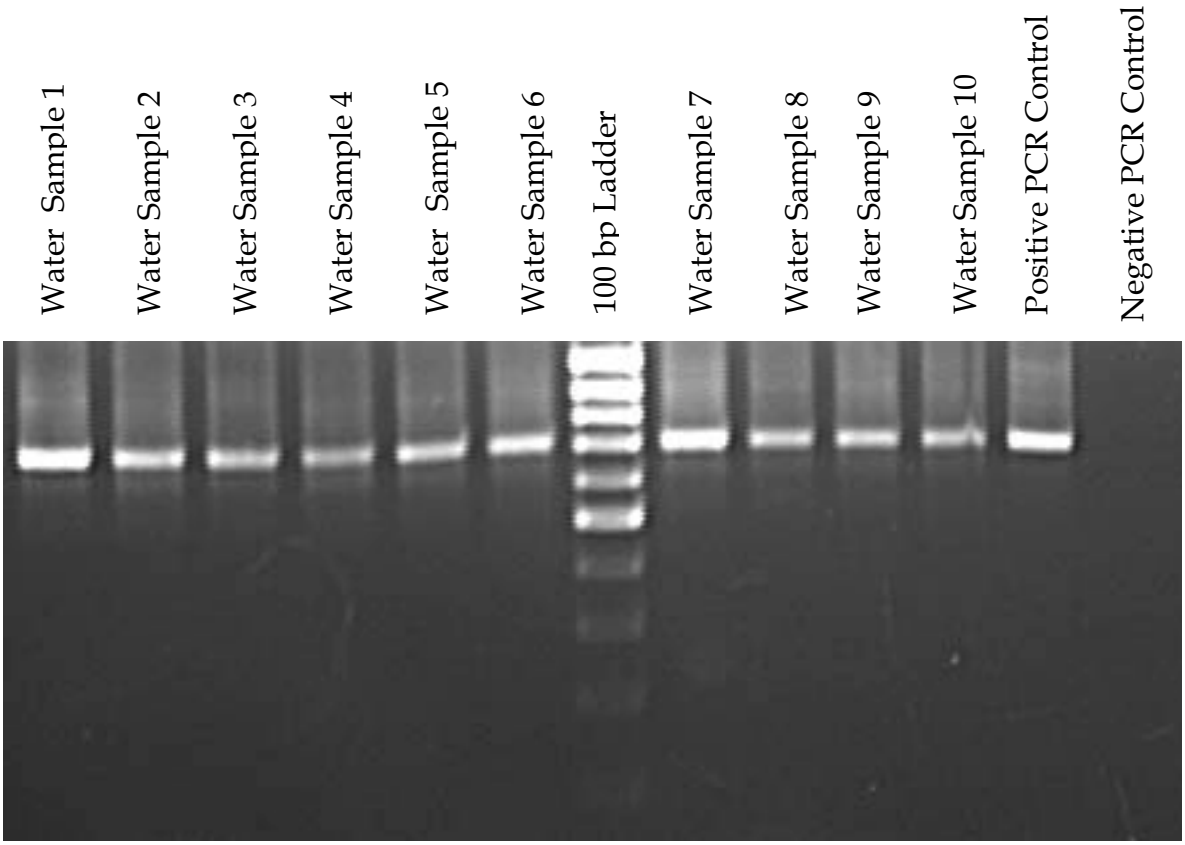


Figure 4.1: Alaska water samples amplified with *Bac*32F and *Bac*708R.

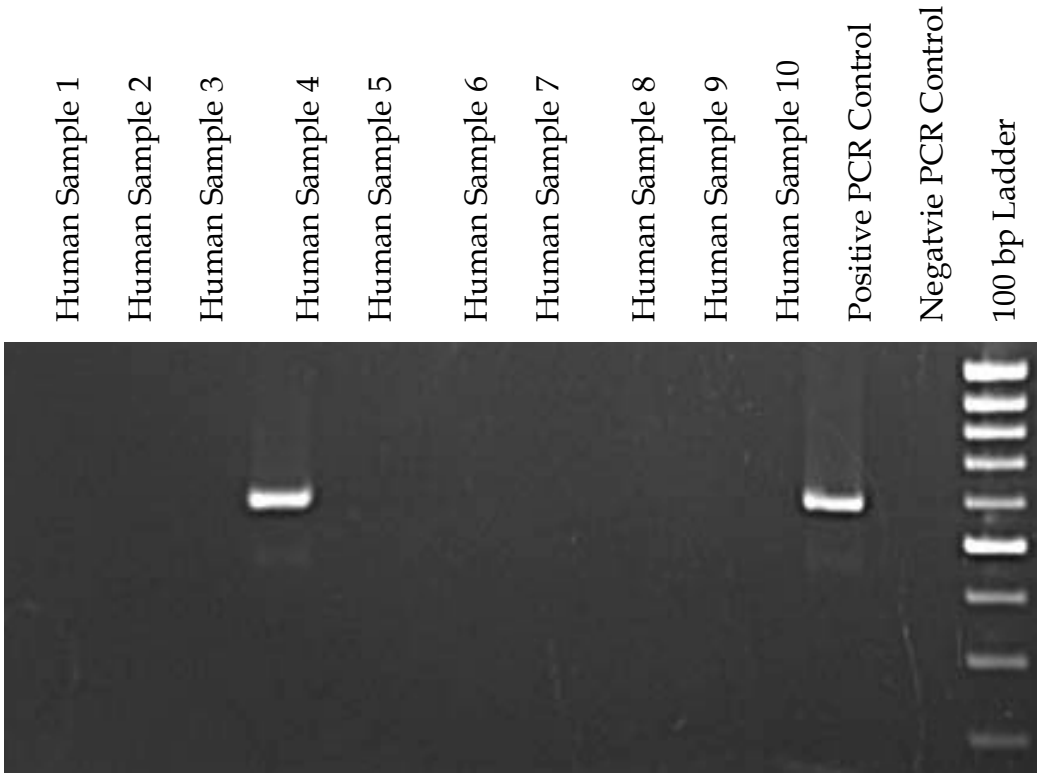


Figure 4.2.1: Alaska human feces amplified with HF134 and *Bac*708R.

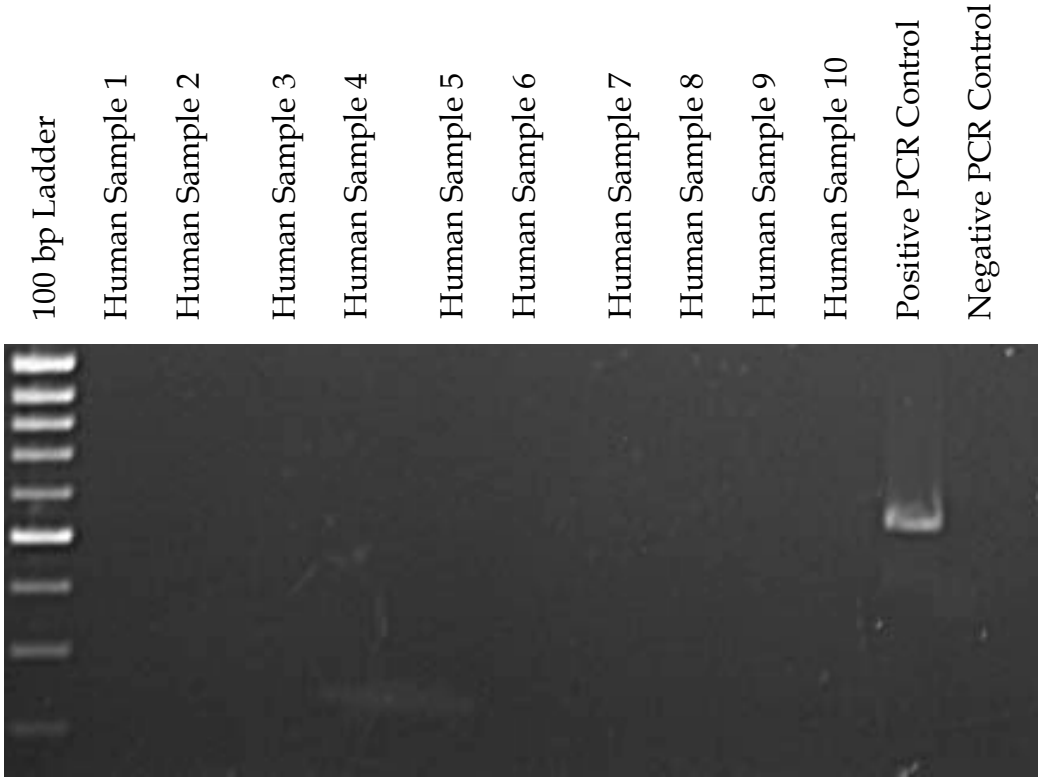


Figure 4.2.2: Alaska human feces amplified with HF183 and *Bac*708R.

Construction and Analysis of 16S rRNA Gene Libraries

4.2. Gene Library Analysis

Eighty-six colonies that contained a DNA insert were picked to comprise the Oregonian sequence library, and 96 colonies were picked to comprise the Alaskan sequence library. After screening with the primer set *Bac*32F and *Bac*1492R, it was found that 84 of the 86 Oregonian clones and 94 of the 96 Alaskan clones contained the desired full-length sequence of the 16S rRNA gene.

4.3. Restriction Digest Analysis

Thirty-two unique restriction patterns were distinguished between both sequence libraries using data obtained from *in vitro* restriction digests. However, after sequencing representative samples from the restriction patterns and analyzing sequences for the presence of restriction sites, the *in silico* restrictions were found to comprise only 8 unique restriction patterns. Because of the disparity in restriction patterns identified *in vitro* versus *in silico*, the next round of sequencing was done on clones chosen at random. These sequences that were restricted *in silico* showed a diverse representation of clones found in the sequence libraries, lending confidence to the diversity of the gene libraries constructed.

Of the 82 samples sequenced, 53 contigs were constructed—24 of the 41 Oregon sequences and 29 of 41 Alaskan sequences. Contigs could not be constructed from the 29 other sequences because the sequence was too degraded in either the forward or reverse direction or both. From the GenBank BLAST

searches performed at the NCBI website, most sequences shared 95-100% identity to cultured *Bacteroides* or *Prevotella* sequences, if they matched to known samples at all. Some sequences also matched to uncultured sequences isolated from humans or ruminants with varying degrees of identity.

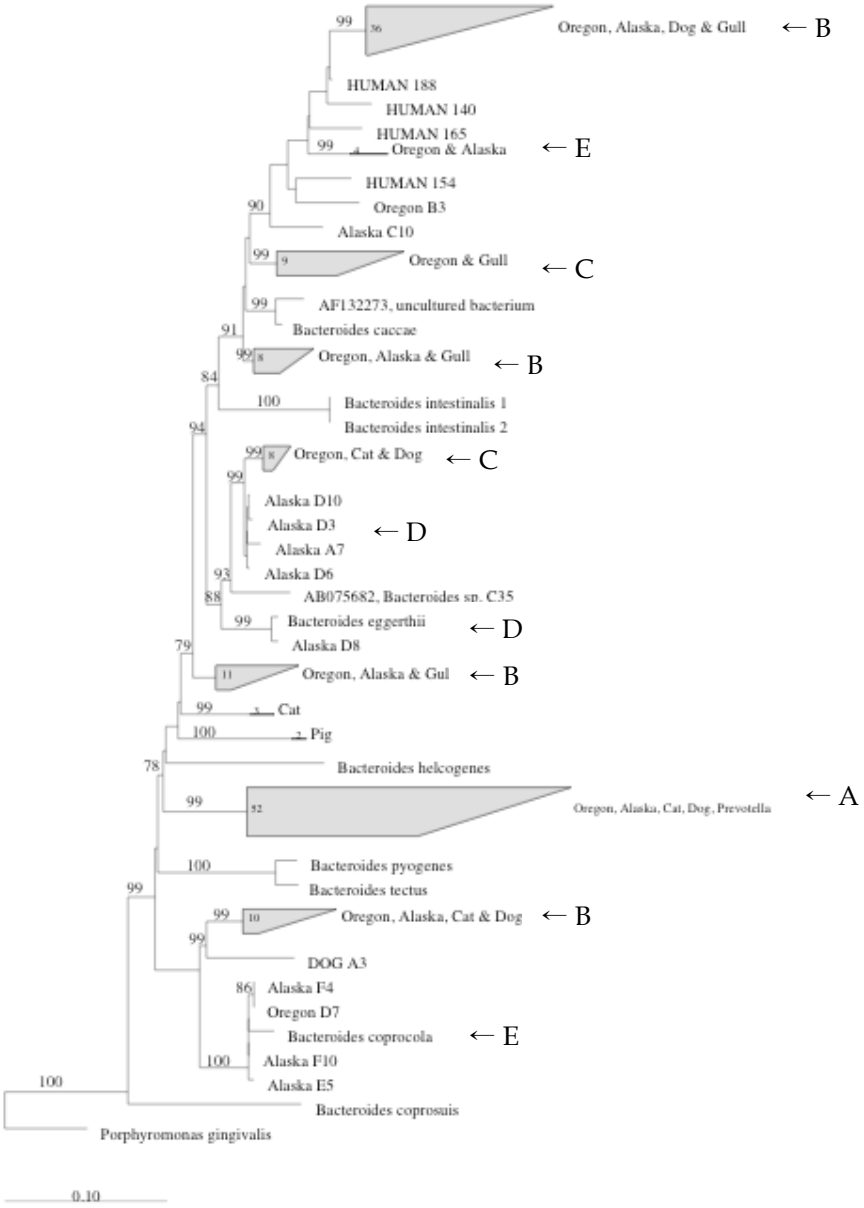
4.4. Phylogenetic Analysis

Figure 4.3 shows a diagrammatic representation of the host distributions from a rooted neighbor-joining tree inferred from the sequences from this study, known *Bacteroides* and *Prevotella* sequences, and sequences from GenBank. Several host-specific clades could be identified; however, sequence clusters were not necessarily monophyletic with respect to host species. Human, cat, dog, and sometimes gull sequences tended to cluster together (Fig 4.3). All cultured *Prevotella* sequences clustered in a large group of sequences that also included cat and dog sequences (Fig. 4.3A). Four clades contained cultured *Bacteroides* sequences along with human sequences from Oregon and Alaska, with 1 clade containing cat and dog sequences, 2 containing gull sequences, and 1 clade containing dog and gull sequences in addition to the human sequences (Fig. 4.3B). Two clades contained cultured *Bacteroides* sequences with Oregon human sequences, with 1 of those clades also containing gull sequences and 1 clade also containing cat and dog sequences (Fig. 4.3C). There were also two clades containing just Alaska sequences with 1 clade containing a cultured *Bacteroides* representative (Fig. 4.3D). There were also 2 clades containing only human sequences from Oregon and Alaska with no cultured representatives (Fig. 4.3E).

Bootstrap values above 70 are shown on the phylogenetic tree, with strong bootstrap values for the abovementioned clades (Fig. 4.3).

Although only nine individuals were sampled in the Oregon fecal clone library, most of the *Bacteroides* sequence diversity uncovered in previous studies was found in this library. Clades containing the cultured species *B. eggerthii*, *B. massiliensis*, and *B. acidofaciens* didn't have representatives from the Oregon library, although the Alaskan library was represented in those clades. In addition, there are several clades with cultivated species for which no uncultivated sequences were found in either the Oregon or Alaskan library. The widespread distribution of the Oregonian sequences in the phylogenetic tree show that nine Oregon individuals adequately represent non-Alaskan diversity. The sequences from the Oregon volunteers are spread throughout the tree with cultured *Bacteroides* and human sequences from previous studies in our lab and elsewhere (9, 18).

Figure 4.3. Rooted neighbor-joining phylogenetic tree with partial and full-length 16S rRNA gene sequences. All named sequences from human samples; sequences marked "HUMAN" are from an earlier study from our laboratory (19, 9, 10); sequences marked "Oregon" and "Alaska" are from human fecal samples in this study; "uncultured" sequences not otherwise identified are from human samples, with the exception of "TB13" which was isolated from Tillamook Bay waters. Cat, dog, gull, and pig sequences are from our previous studies (19). (A) clades containing Oregon and Alaska human sequences, cat, dog, gull, and cultured *Prevotella* sequences; (B) clades containing Oregon and Alaska human sequences with cat, dog, gull, and cultured *Bacteroides* sequences; (C) clades containing Oregon human sequences with cat, dog, gull, and cultured *Bacteroides* sequences; (D) clades containing Alaskan human sequences; (E) clades containing Oregon and Alaska human sequences.


Oregon, Alaska, Dog & Gull
← B
HUMAN 188
HUMAN 140
HUMAN 165
Oregon & Alaska
← E
HUMAN 154
Oregon B3
Alaska C10
Oregon & Gull
← C
AF132273, uncultured bacterium
Bacteroides caccae
Oregon, Alaska & Gull
← B
Bacteroides intestinalis 1
Bacteroides intestinalis 2
Oregon, Cat & Dog
← C
Alaska D10
Alaska D3
Alaska A7
← D
Alaska D6
AB075682, Bacteroides sn. C35
Bacteroides eggerthii
Alaska D8
← D
Oregon, Alaska & Gul
← B
Cat
Pig
Bacteroides helcogenes
Oregon, Alaska, Cat, Dog, Prevotella
← A
Bacteroides pyogenes
Bacteroides tectus
Oregon, Alaska, Cat & Dog
← B
DOG A3
Alaska F4
Oregon D7
Bacteroides coprocola
← E
Alaska F10
Alaska E5
Bacteroides coprosuis
Porphyromonas gingivalis
0.10

Design and Testing of Novel Host-Specific Bacteroidales 16S rRNA Primers

4.5. Primer Design

Host-specific sequences were identified from the fecal 16S rDNA gene sequences for Alaskan *Bacteroidales*. They were used to design two Alaskan human-specific PCR primers (HF185F and HF1001R). Primer HF185F was coupled with *Bac*708R and primer HF1001 was coupled with the general *Bacteroidales* primer *Bac*32F.

4.6. Fecal Pool Constructs

The 9 fecal pools made from non-human sources—cat, dog, cow, horse, sheep, elk, pig, chicken, and gull—all successfully amplified with the general *Bacteroidales* primer pair *Bac*32F and *Bac*708R (data not shown). This showed that these pools contained *Bacteroides* DNA that could be used to screen new primer pairs for host-specificity.

4.7. Primer Testing

Both primer pairs, *Bac*32F coupled with HF1001R and HF185F coupled with *Bac*708R, were highly specific to Alaska human feces using fecal DNA pools from target and non-target host species (Figure 4.4.1. & 4.4.2.). The Alaskan-specific marker HF185F was found in 9 of 10 household fecal samples tested, and the Alaskan-specific marker HF1001R was found in 10 of 10 household fecal samples tested (Figure 4.5.1. & 4.5.2.).

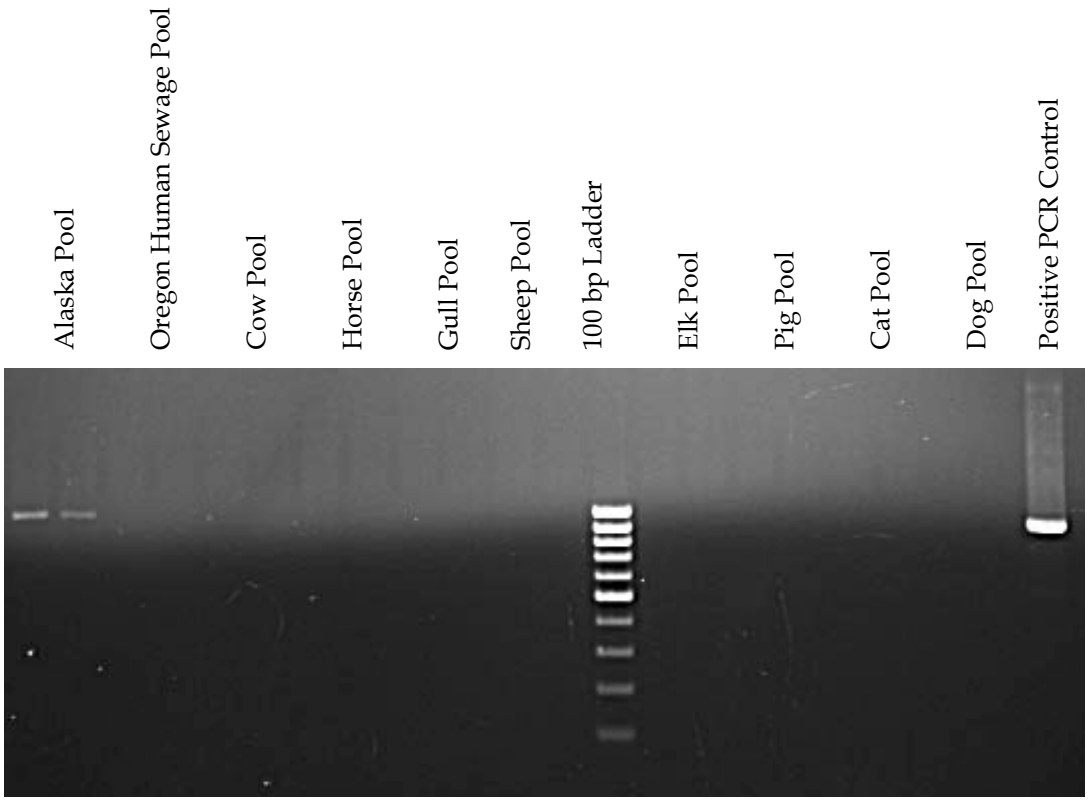


Figure 4.4.1: Multiple hosts amplified with *Bac*32F and HF1001R.

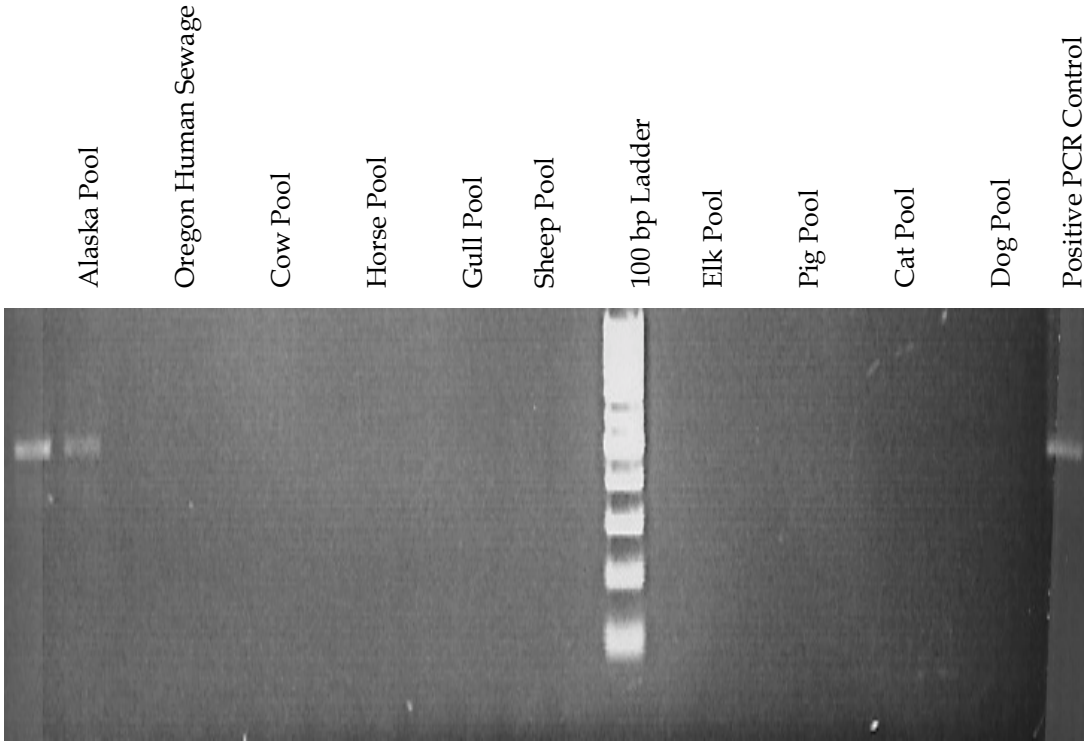


Figure 4.4.2: Multiple hosts amplified with HF185F and 708R.

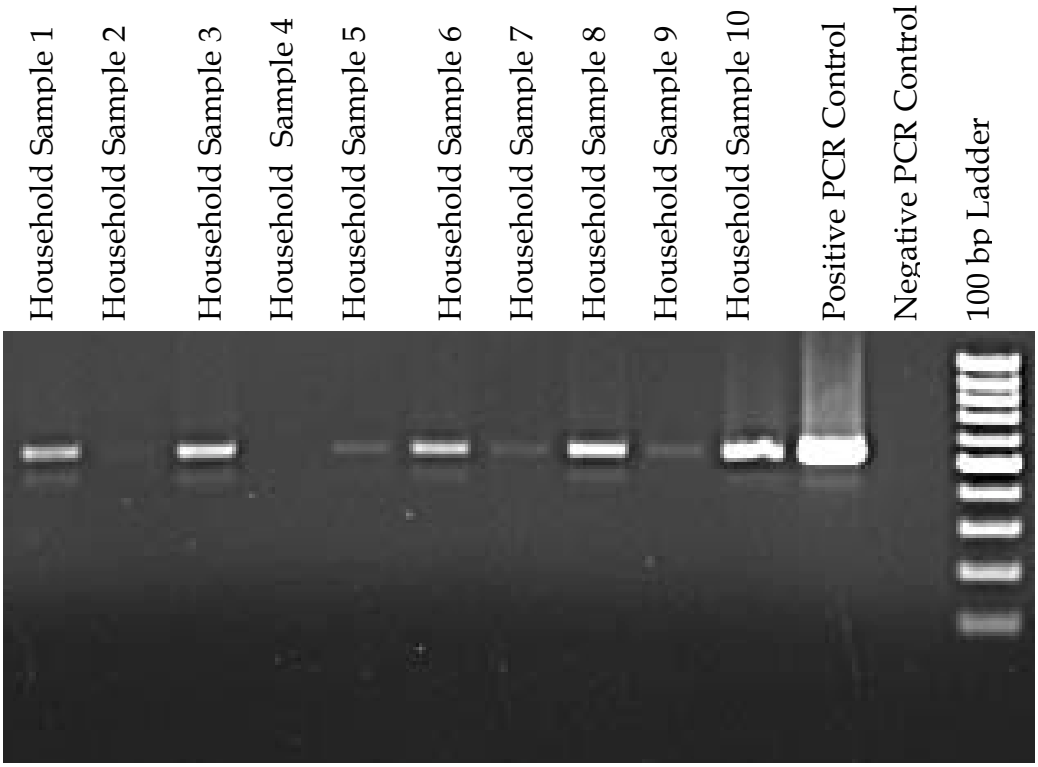


Figure 4.5.1: Alaska human fecal sampes amplified with HF185F and 708R.

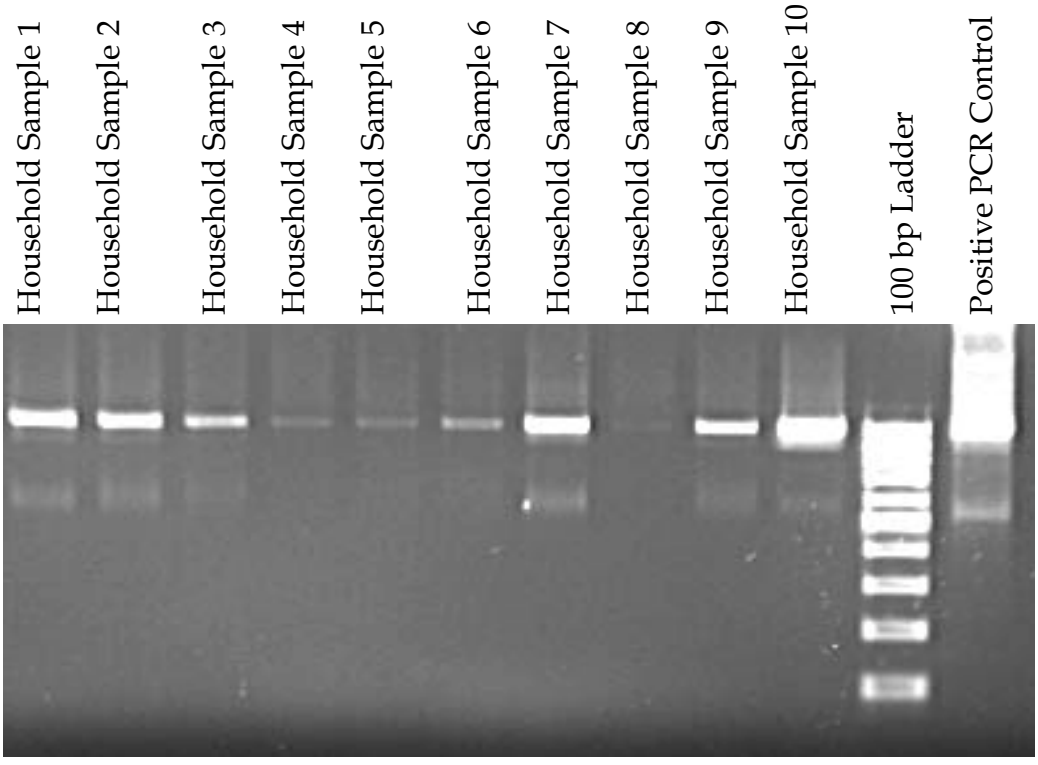


Figure 4.5.2: Alaska human fecal samples amplified with *Bac*32F and HF1001R.

4.8. Limit of Detection

The theoretical limit of detection for both Alaskan-specific markers was determined to be 10 pg of plasmid template (data not shown). This translates into a sensitivity of 1.6×10^6 gene copies or 2.8×10^5 cells.

4.9. Alaskan Water Sample Analysis

The two new Alaska-specific primer pairs were used to amplify the 11 contaminated water samples with no amplification observed (Figure 4.6.1. & 4.6.2.).

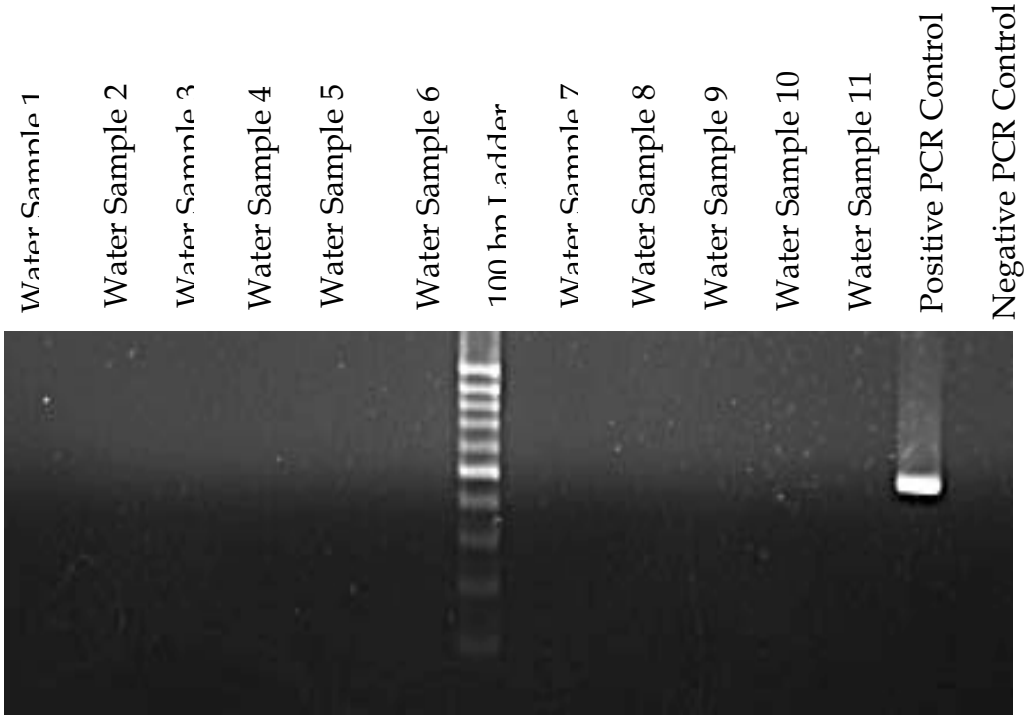


Figure 4.6.1: Alaska water samples amplified with HF185F and *Bac*708R.

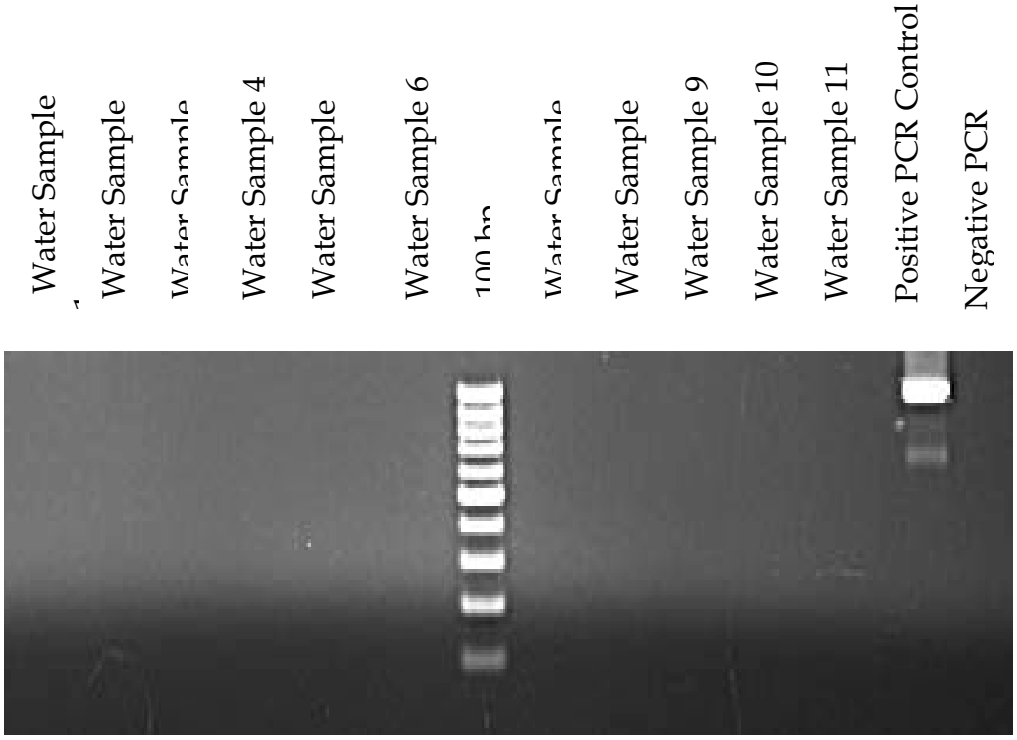


Figure 4.6.2: Alaska water samples amplified with *Bac*32F and HF1001R.

Chapter 5. DISCUSSION

Fecal and Water Sample Analysis

5.1. PCR Analysis

Fecal samples: Because the Alaskan human fecal samples did amplify with the general *Bacteroidales* primer pair but not with either of the human-specific primer pairs, it was concluded that these Alaskan fecal samples did contain *Bacteroidales* sequences, but did not contain sequences from certain common human *Bacteroidales* groups. HF134F and HF183F, two human-specific primers in current use amplify human fecal DNAs in the U. S., Canada, Europe, and New Zealand (19). The inability of these primers to amplify Alaskan human fecal samples means that this isolated human population has different enough fecal bacteria to distinguish them from other non-isolated populations. Specifically, they are missing an otherwise-common human-specific gene cluster. The reasons for this difference could be geographic isolation, population history, or diet.

The successful amplification of the Alaskan dog fecal samples with both the general *Bacteroidales* and dog-specific primer pairs showed that there are *Bacteroidales* sequences in the fecal samples collected from the Alaskan dogs, and the samples contain sequences from common dog *Bacteroidales* groups. This also demonstrates that the dog-specific primer pair could be used to test the water

samples; if there were no amplification in water, it would be because dogs are not the source of fecal contamination.

Water samples: Because of the successful amplification of the Alaskan water samples with the general *Bacteroidales* primer pair, there was definitely fecal contamination with *Bacteroidales* bacteria. This confirmed *E. coli* results; water samples were chosen that had high *E. coli* counts. However, since there was no amplification of the water samples with the dog-specific primer set, dogs could be ruled out as the source of fecal contamination. Humans could not be ruled in or out, since there was no amplification of the Alaskan human fecal samples with either of the human-specific primer pairs. No amplification of the water samples with either human-specific primer pair reinforced the earlier finding that these primers do not work with Alaskan human feces.

Construction and Analysis of 16S rRNA Gene Libraries

5.2. Restriction Digest Analysis

The results of the *in vitro* restriction digests were disappointing. The clones picked for sequencing based on the *in vitro* restriction digests were later restricted *in silico* and found to have different restriction patterns than expected. This was most likely due to incomplete digestion of the PCR products. This might have been avoided by purifying the PCR products before digestion to try and minimize extraneous DNA in the restriction digests. Additionally,

sequencing of the entire clone libraries should be done to ensure complete coverage of the diversity represented in the libraries.

5.3. Phylogenetic Analysis

Figure 4.3 shows that the Alaska *Bacteroidales* 16S rRNA gene sequences are scattered throughout the phylogenetic tree. The human-specific primers HF134F and HF183F were designed in an earlier study from the HUMAN 8 sequence. While there are a few Alaska sequences that were closely related to this sequence, they were not close enough to contain the primer target sequence, which further lends credence to the Alaskan fecal bacteria being distinct from fecal bacteria in other locations.

The widespread distribution of the *Bacteroidales-Prevotella* 16S gene sequences from all hosts suggests that fecal bacteria are passed horizontally among species. This is especially evident when looking at human, cat, and dog *Bacteroidales-Prevotella* 16S gene sequences. The ease of transmission between humans and their cats and dogs explains why most clades contain gene sequences from these 3 hosts. Also, birds have access to human waste through garbage dumps and contaminated of water, resulting in shared 16S gene sequences between humans and birds. Additionally, the high bootstrap values indicate strong support for the branches displayed.

Design and Testing of New Host-Specific Bacteroidales 16S rRNA Primers

5.4. Primer Design

Evaluation of primer specificity was hindered by the fact that although sequences from this study were full-length sequences of the 16S rDNA gene, the sequences from previous studies were only partial sequences of the 16S rDNA gene. Primers could only be evaluated *in silico* based for host-specificity with the non-human sequences if they were within the first half of the 16S rDNA gene. This led to many primers being tested that amplified other non-human species, most notably cats and dogs. All future studies should include sequencing the full-length 16S rDNA gene, which may also alter the phylogenetic resolution of the trees constructed.

5.5. Primer Testing

While 2 new primers were designed that amplified human fecal samples from Alaska, neither primer pair amplified human fecal samples from Oregon, which would have been ideal. A primer pair that would amplify human fecal bacteria from a variety of locations would be useful for future studies and could be applied to more diverse locations than the current human-specific primers.

5.6. Alaskan Water Sample Analysis

There was no amplification in any of the 11 Alaska water samples with either of the new Alaska-specific primer pairs, potentially ruling out humans as

the source of fecal contamination. Both of the new Alaskan primers detected 1.6 x 10^6 gene copies or 2.8 x 10^5 cells. To obtain more sensitive PCR reactions, further optimizations could be done. One gram of feces contains 3 x 10^{11} bacterial cells with 30% (or 10^{11}) of those being *Bacteroides* and there is a 20:1 ratio of *E. coli* to *Bacteroides*, meaning that there is 5 x 10^9 *E. coli* cells per gram of feces (9). Since there was approximately 500 *E. coli* cells per ml of water (72), meaning that there was 1 x 10^{-7} grams of feces per milliliter of water. With 10^{11} *Bacteroides* bacterial cells per gram of feces there is 10,000 *Bacteroides* cells per milliliter of water in the Alaskan water samples. Considering that there was 100 ml of water (1,000,000 *Bacteroides* cells) and with a primer sensitivity of 2.8 x 10^5 cells, these primers would not have been sensitive enough to detect the presence of human *Bacteroides* in the sample. Since dogs were ruled out as the source of fecal contamination the actual source of fecal contamination is not known. The likely source of contamination may be indigenous wildlife for which there are no markers developed. Several potential host sources of fecal contamination include beavers, moose, or several species of birds.

5.7. Future Directions

This study determined that dogs and humans are unlikely to be the source of fecal contamination in surface waters of a rural Alaskan village, but did not identify the actual source, which was disappointing. However, if the source of fecal contamination is indigenous wildlife, remediation of the problem is quite

complex. It is reassuring that neither humans nor dogs are the source considering the various opportunities in which contamination could occur by either species. To identify the host source, the same methods could be applied using fecal samples from other possible hosts like beaver, moose, and birds.

Another area for future study would be to sample human feces on a global level to investigate the distribution of the 2 new human-specific primers and the 2 current primers, HF134F and HF183F. This will reveal whether the new primers are unique to isolated and/or native populations and what other populations, if any, do not amplify with the human-specific primers HF134F and HF183F.

The results of this study have also pointed to a new direction of research that explores the biodiversity of commensals. Since *H. pylori* has been used to trace human migration patterns, perhaps fecal anaerobes have followed a similar distribution pattern. In another study looking at the distribution of pathogenic microorganisms a strong correlation was found between species richness and latitude in some groups of pathogens (34). The researchers found that species richness increased as latitude decreased. Future studies could investigate whether commensals follow the same species richness and latitude correlation as macroorganisms and some pathogenic microorganisms.

In summary, the findings presented in this thesis have shown that neither humans nor dogs are responsible for the fecal contamination of water samples from an isolated Alaskan village. The findings also show that there are

differences in human fecal bacteria between this native population and human fecal bacteria from non-isolated populations in Oregon. Both of these results point to directions for future research that will uncover the source of fecal contamination in this rural village and will examine the distribution of fecal anaerobes.

REFERENCES

1. **Agency, U. S. E. P.** 2001. National coastal condition report EPA-620-R-01-005. U.S. Environmental Protection Agency, Washington, D. C.
2. **Alderisio, K. A., and N. DeLuca.** 1999. Seasonal enumeration of fecal coliform bacteria from the feces of ring-billed gulls (*Larus delawarensis*) and Canada geese (*Branta canadensis*). Applied and Environmental Microbiology **65:**5628-5630.
3. **Allsop, K., and D. J. Stickler.** 1984. The enumeration of *Bacteroides fragilis* group organisms from sewage and natural waters. J. Appl. Bacteriol. **56:**15-24.
4. **Anderson, K. L., J. E. Whitlock, and V. J. Harwood.** 2005. Persistence and differential survival of fecal indicator bacteria in subtropical waters and sediments. Applied and Environmental Microbiology **71:**3041-3048.
5. **Ayebo, A., G. M. Breuer, T. G. Cain, M. D. Wichman, P. Subramanian, and S. J. Reynols.** 2006. Sterols as bio-markers for waste impact and source characterization in stream sediment. Jounal of Environmental Health **68:**46-50.
6. **Bej, A. K., R. J. Stefan, J. DiCesare, L. Haff, and R. M. Atlas.** 1990. Detection of colifrom bacteria in water by polymerase chain reaction and gene probes. Applied and Environmental Microbiology **56:**307-314.
7. **Bellar, T. A., and W. L. Budde.** 1988. Determination of nonvolatile organic compounds in aqueous environmental samples using liquid chromatography/mass spectrometry. Anal. Chem. **60:**2076-2083.
8. **Bendriss, E., N. Markoglou, and I. W. Wainer.** 2000. Liquid chromatographic method for the simultaneous determination of caffeine and fourteen caffeine metabolites in urine. Journal of Chromatography B **746:**331-338.
9. **Bernhard, A. E., and K. G. Field.** 2000. 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:**1587-1594.
10. **Bernhard, A. E., and K. G. Field.** 2000. A PCR assay to discriminate human and ruminant feces on the basis of host differences in *Bacteroides-Prevotella* genes encoding 16S rRNA. Appl. Environ. Microbiol. **66:**4571-4574.
11. **Blaser, M. J.** 2005. An Endangered Species in the Stomach. Scientific American **292:**38-45.
12. **Booth, A. M., C. Hagedorn, A. K. Graves, S. Hagedorn, and K. H. Mentz.** 2003. Sources of fecal pollution in Virginia's Blackwater river. Journal of Environmental Engineering **129:**547-552.
13. **Buchan, A., M. Alber, and R. E. Hodson.** 2001. Strain-specific differentiation of environmental *Eschericia coli* isolates via denaturing

gradient gel electrophoresis (DGGE) analysis of the 16S-23S intergenic spacer region. FEMS Microbiology Ecology **35:**313-321.

14. **Carson, C. A., B. L. Shear, M. R. Ellersieck, and A. Asfaw.** 2001. Identification of fecal *Escherichia coli* from humans and animals by ribotyping. Applied and Environmental Microbiology **67:**1503-1507.
15. **Council, N. R. D.** 2006. Pollution-related beach closings and advisories climb in 2005. http://www.nrdc.org/water/oceans/nttw.asp.
16. **Daly, K., C. S. Stewart, H. J. Flint, and S. P. Shirazi-Beechey.** 2001. Bacterial diversity within the equine large intesting as revealed by molecular analysis of cloned 16S rRNA genes. FEMS Microbiology Ecology **38:**141-151.
17. **Davies, C. M., J. H. Long, M. Donald, and N. J. Ashbolt.** 1995. Survival of fecal microorganisms in marine and freshwater sediments. Appl. Environ. Microbiol. **61:**1888-1896.
18. **Dick, L. K., A. E. Bernhard, T. J. Brodeur, J. W. Santo Domingo, J. M. Simpson, S. P. Walters, and K. G. Field.** 2005. Host Distributions of Uncultivated Fecal Bacteroidales Reveal Genetic Markers for Fecal Source Identification. Appl Environ Microbiol **71:**3184-3191.
19. **Dick, L. K., M. T. Simonich, and K. G. Field.** 2005. Microplate subtractive hybridization to enrich for source-specific Bacteroidales fecal pollution indicators. Appl Environ Microbiol **71:**3179-3183.
20. **Ding, W.-H., J. Wu, M. Semadeni, and M. Reinhard.** 1999. Occurence and behavior of wastewater indicators in the Santa Ana river and the underlying aquifers. Chemosphere **39:**1781-1794.
21. **Falush, D., T. Wirth, B. Linz, J. K. Pritchard, M. Stephens, M. Kidd, M. J. Blaser, D. Y. Graham, S. Vacher, G. I. Perez-Perez, Y. Yamaoka, F. Megraud, K. Otto, U. Reichard, E. Katzowitsch, X. Wang, M. Achtman, and S. S.** 2003. Traces of Human Migration in *Helicobacter pylori* Populations. Science **299:**1582-85.
22. **Farnleitner, A. H., N. Kreuzinger, G. G. Kavka, S. Grillenberger, J. Rath, and R. L. Mach.** 2000. Simultaneous detection and differentiation of *Escherichia coli* populations from environmental freshwaters by means of sequence variations in a fragment of the beta-D-glucuronidase gene. Applied and Environmental Microbiology **66:**1340-1346.
23. **Field, K. G.** 2004. Faecal Source Identification. *In* J. A. Cotruvo, A. Dufour, G. Rees, J. Bartram, R. Carr, D. O. Cliver, G. F. Craun, R. Fayer, and V. P. G. Gannon (ed.), Waterborne Zoonoses: Identification, Causes and Control. World Health Organization, IWA Publishing Alliance House, London.
24. **Field, K. G., E. C. Chern, L. K. Dick, J. A. Fuhrman, J. F. Griffith, P. A. Holden, M. G. LaMontagne, J. Le, B. H. Olson, and M. T. Simonich.** 2003. A comparative study of culture-independent, library-independent genotypic methods of fecal source tracking. J. Water Health **1:**181-194.

25. **Fiksdal, L., J. S. Make, S. J. LaCroix, and J. T. Staley.** 1985. Survival and detection of *Bacteroides* spp., prospective indicator bacteria. Appl. Environ. Microbiol. **49:**148-150.
26. **Fogarty, L. R., S. K. Haack, M. J. Wolcott, and R. L. Whitman.** 2003. Abundance and characteristics of the recreational water quality indicator bacteria *Escherichia coli* and enterococci in gull faeces. Journal Applied Microbiology **94:**865-878.
27. **Garrity, G. M., J. Bell, and T. G. Lilburn.** May, 2004. Taxonomic outline of the prokaryotes. Bergey's Manual of Systemic Bacteriology, Second edition, Release, 5.0.
28. **Geldreich, E. E.** 1970. J Am Water Works Assoc **62:**113-120.
29. **Geldreich, E. E., and B. A. Kenner.** 1969. J Water Pollut Control Fed **41:**R336-352.
30. **Gordon, D. M.** 2001. Geographical structure and host specificity in bacteria and the implications for tracing the source of coliform contamination. Microbiol. **147:**1079-1085.
31. **Gordon, D. M., and F. FitzGibbon.** 1999. The distribution of enteric bacteria from Australian mammals: host and geographical effects. Microbiology **145:**2663-2671.
32. **Griffith, J. F., S. B. Weisberg, and C. D. McGee.** 2003. Evaluation of microbial source tracking methods using mixed fecal sources in aqueous test samples. Journal of Water and Health **01.4:**141-151.
33. **Guan, S., R. Xu, S. Chen, J. Odumeru, and C. Gyles.** 2002. Development of a procedure for discriminating among *Escherichia coli* isolates from animal and human sources. Applied and Environmental Microbiology **68:**2690-2698.
34. **Guernier, V., M. E. Hochberg, and J.-F. Guegan.** 2004. Ecology Drives the Worldwide Distribution of Human Diseases. PLoS Biology **2:**0740-46.
35. **Hagedorn, C., J. B. Crozier, K. A. Mentz, A. M. Booth, A. K. Graves, N. J. Nelson, and R. B. Reneau Jr.** 2003. Carbon source utilization profiles as a method to identify sources of fecal pollution in water. J. Appl. Microbiol. **94:**792-799.
36. **Hartel, P. G., J. D. Summer, J. L. Hill, J. V. Collins, J. A. Entry, and W. I. Segars.** 2002. Geographic variability of *Escherichia coli* ribotypes from animals in Idaho and Georgia. J. Environ. Qual. **31:**1273-1278.
37. **Harwood, V. J., J. Butler, D. Parrish, and V. Wagner.** 1999. Isolation of fecal coliform bacteria from the diamondback terrapin (*Malaclemys terrapin centrata*). Appl. Environ. Microbiol. **65:**865-867.
38. **Harwood, V. J., B. A. Wiggins, C. Hagedorn, R. D. Ellender, J. Gooch, J. Kern, M. Samadpour, A. C. H. Chapman, and B. J. Robinson.** 2003. Phenotypic library-based microbial source tracking methods: efficacy in the California collaborative study. J. Water Health **1:**153-166.

39. **Hatcher, P. G., and P. A. McGillivary.** 1979. Sewage contamination in the New York Bight. Coprostanol as an indicator. Environmental Science and Technology Research **13:**1225-1229.
40. **Havelaar, A. H., K. Furuse, and W. M. Hogeboom.** 1986. Bacteriophages and indicator bacteria in human and animal faeces. J App Bacteriol **60:**255-262.
41. **Havelaar, A. H., and W. M. Hogeboom.** 1984. A method for the enumeration of male-specific bacteriophages in sewage. J App Bacteriol **56:**439-447.
42. **Hold, G. L., S. E. Pryde, V. J. Russell, E. Furrie, and H. J. Flint.** 2002. Assessment of microbial diversity in human colonic samples by 16S rDNA sequence analysis. FEMS Microbiol. Ecol. **39:**33-39.
43. **Hooper, L. V., and J. I. Gordon.** 2001. Commensal host-bacterial relationships in the gut. Science **292:**1115-1118.
44. **Hooper, L. V., T. Midtvedt, and J. I. Gordon.** 2002. How host-microbial interactions shape the nutrient environment of the mammalian intestine. Annu. Rev. Nutr. **22:**283-307.
45. **Horie, H., and K. Kohata.** 2000. Analysis of tea components by high-performance liquid chromatography and high-performance capillary electrophoresis. Journal of Chromatography A **881:**425-438.
46. **Howell, J. M., M. S. Coyne, and P. L. Corneilus.** 1996. Effect of sediment particle size and temperature on fecal bacteria mortality rates and the fecal coliform/fecal streptococci ratio. Journal of Environmental Quality **25:**1216-1220.
47. **Kanazawa, H., R. Atsumi, Y. Matsushima, and J. Kizu.** 2000. Determination of theophylline and its metabolites in biological samples by liquid chromatography-mass spectrometry. Journal of Chromatography A **870:**87-96.
48. **Kariuki, S., C. Gilks, J. Kimari, A. Obanda, J. Muyodi, P. Waiyali, and C. A. Hart.** 1999. Genotype analysis of *Escherichia coli* strains isolatd from children and chickens living in close contact. Appl. Environ, Microbiol. **65:**472-476.
49. **Khatib, L. A., Y. L. Tsai, and B. H. Olson.** 2002. A biomarker for the identification of cattle fecal pollution in water using the LTIIa toxin gene from enterotoxigenic *Escherichia coli*. Applied and Environmental Microbiology **59:**97-104.
50. **Kreader, C. A.** 1995. Design and evaluation of *Bacteroides* DNA probes for the specific detection of human fecal pollution. Appl. Environ. Microbiol. **61:**1171-1179.
51. **Kreader, C. A.** 1998. Persistence of PCR-detectable *Bacteroides distasonis* from human feces in river water. Appl. Environ. Microbiol. **64:**4103-4105.
52. **Leser, T. D., J. Z. Amenuvor, T. K. Jensen, R. H. Lindecrona, M. Bove, and K. Moller.** 2002. Culture-independent analysis of gut bacteria: the pig

gastrointestinal tract microbiota revisited. Appl. Environ. Microbiol. **68:**673-690.

53. **Leser, T. D., R. H. Lindecrona, T. K. Jensen, B. B. Jensen, and K. Moller.** 2001. Cuture-independent analysis of gut bacteria: the pig gastrointestinal tract microbiota revisited. Applied and Environmental Microbiology **68:**673-690.
54. **Long, S. C.** 2005. Assessment of sources and diversity of male-specific coliphages for source tracking. Environmental Engineering Science **22:**367-377.
55. **Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Heier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Forster, L. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. Konig, T. Liss, R. Lubmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K. H. Schleifer.** 2004. ARB: a software environment for sequence data. Nucleic Acids Research **32:**1363-1371.
56. **Maidak, B. L., N. Larsen, M. J. McCaughey, R. Overbeek, G. J. Olsen, K. Fogel, J. Blandy, and C. R. Woese.** 1994. The Ribosomal Database Project. Nucleic Acids Research **22:**3485-3487.
57. **Mazmanian, S. K., C. H. Liu, A. O. Tzianabos, and D. L. Kasper.** 2005. An Immunomodulatory Molecule of Symbiotic Bacteria Directs Maturation of the Host Immune System. Cell **122:**107-118.
58. **Meays, C. L., K. Broersma, R. Nordin, and A. Mazumder.** 2004. Source tracking fecal bacteria in water: a critical review of current methods. Journal of Environmental Management **73:**71-79.
59. **Munian-Mujika, I., M. Calvo, F. Lucena, and R. Girones.** 2003. Comparative analysis of viral and potential indicators in shellfish. Int J Food Microbiol **83:**75-85.
60. **Noblet, J. A., D. L. Young, E. Y. Zeng, and S. Ensari.** 2004. Use of fecal sterols to infer the sources of fecal indicator bacteria in the lower Santa Ana river watershed, California: Sewage is unlikely a significant source. Environmental Science and Technology **38:**6002-6008.
61. **Parveen, S., K. M. Portier, K. Robinson, L. Edmiston, and M. Tamplin.** 1999. Discriminate analysis of ribotype profiles of *Escherichia coli* for differentiating human and nonhuman sources of fecal pollution. Appl. Environ. Microbiol. **65:**3142-3147.
62. **Radke, M.** 2005. Sterols and anionic surfactants in urban aerosol: emissions from wastewater treatment plants in relation to background concentrations. Environmental Science and Technology **39:**4391-4397.
63. **Rakhoff-Nahoum, S., J. Paglino, F. Eslami-Varzaneh, S. Edberg, and R. Medzhitov.** 2004. Recognition of Commensal Microflora by Toll-Like Receptors is Required for Intestinal Homeostasis. Cell **118:**229-241.

64. **Resnick, I. G., and M. A. Levin.** 1981. Assessment of *Bifidobacteria* as indicators of human fecal pollution. Applied and Environmental Microbiology **42:**433-438.
65. **Ritchie, N. J., M. E. Schutter, R. P. Dick, and D. D. Myrold.** 2000. Use of length heterogeneity PCR and fatty acid methyl ester profiles to characterize microbial communities in soil. Applied and Environmental Microbiology **66:**1668-1675.
66. **Samadpour, M., J. Stewart, K. Steingart, C. Addy, J. Louderback, M. McGinn, J. Ellington, and T. Newman.** 2002. Laboratory investigation of an *Escherichia coli* 0157:H7 outbreak associated with swimming in Battle Ground Lake, Vancouver, Washington. Journal of Environmental Health **64:**16-20.
67. **Shanks, O. C., J. W. Santo Domingo, R. Lamendella, C. A. Kelty, and J. E. Graham.** 2006. Competetive metagenomic DNA hybridization identifies host-specific microbial genetic markers in cow fecal samples. Applied and Environmental Microbiology **72:**4054-4060.
68. **Simpson, J. M., J. W. Santo Domingo, and D. J. Reasoner.** 2002. Microbial source tracking: state of the science. Environ. Sci. Technol. **36:**5279-5288.
69. **Standley, L. J., L. A. Kaplan, and D. Smith.** 2000. Molecular tracers of organic matter sources to surface water resources. Environmental Science and Technology **34:**3124-3130.
70. **Stoeckel, D. M., M. V. Mathes, K. E. Hyer, C. Hagedorn, H. Kator, J. Lukasik, T. L. O'Brien, T. W. Fenger, M. Samadpour, K. M. Strickler, and B. A. Wiggins.** 2004. Comparison of seven protocols to identify fecal contamination sources using *Escherichia coli*. Environmental Science and Technology **38:**6109-6117.
71. **Tynkkynen, S., R. Satokari, M. Saarela, T. Mattila-Sandholm, and M. Saxelin.** 1999. Comparison of ribotyping, randomly amplified polymorphic DNA analysis, and pulsed-field gel electrophoresis in typing of *Lactobacillus rhamnosus* and *L. casei* strains. Applied and Environmental Microbiology **65:**3908-3914.
72. **University of Alaska, F.** 2004. Protecting family drinking water in rural Alaska: Preliminary survey and research findings.28p.
73. **Weaver, R. W., J. A. Entry, and A. K. Graves.** 2005. Numbers of fecal streptococci and *Escherichia coli* in fresh and dry cattle, horse, and sheep manure. Can J Microbiol **51:**847-851.
74. **Weigel, S., K. Bester, and H. Huhnerfuss.** 2001. New method for rapid solid-phase extraction of large-volume water samples and its application to non-target screening of North Sea water for organic contaminants by gas chromatography-mass spectrometry. Journal of Chromatography A **912:**151-161.
75. **Wheeler, A. E., J. Burke, and A. Spain.** 2003. Fecal indicator bacteria are abundant in wet sand at freshwater beaches. Water Res. **37:**3978-3982.

76. **Wiggins, B. A.** 1996. 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:**3997-4002.

77. **Wiggins, B. A., R. W. Andrews, R. A. Conway, C. L. Corr, E. J. Dobratz, D. P. Dougherty, J. R. Eppard, S. R. Knupp, M. C. Limjoco, J. M. Mettenburg, J. M. Rinehardt, J. Sonsino, R. L. Torrijos, and M. E. Zimmerman.** 1999. Use of antibiotic resistance analysis to identify nonpoint sources of fecal pollution. Appl. Environ. Microbiol. **65:**3483-3486.

78. **Wiggins, B. A., P. W. Cash, W. S. Creamer, S. E. Dart, P. P. Garcia, T. M. Gerecke, J. Han, B. L. Henry, K. B. Hoover, E. L. Johnson, K. C. Jones, J. G. McCarthy, J. A. McDonough, S. A. Mercer, M. J. Noto, H. Park, M. S. Phillips, S. M. Purner, B. M. Smith, E. N. Stevens, and A. K. Varner.** 2003. Use of antibiotic resistance analysis for representativeness testing of multiwatershed libraries. Appl. Environ. Microbiol. **69:**3399-3405.

79. **Wolf, G.** 2006. Gut Microbiota: A Factor in Energy Regulation. Nutrition Reviews **64:**47-50.