AN ABSTRACT OF THE DISSERTATION OF

Anne E. Bernhard for the degree of Doctor of Philosophy in Microbiology presented on July 28, 2000. Title: Molecular Markers from Fecal Anaerobes to Identify Nonpoint Source Pollution in Coastal Waters

Abstract

Redacted for privacy

Katharine G. Field

Fecal pollution is a serious environmental problem afflicting many coastal regions in the U. S. and worldwide. Despite efforts to minimize fecal input to coastal waters, the problem persists, partly due to the inability to reliably identify nonpoint sources. We describe a new PCR assay to differentiate sources of fecal contamination in coastal waters. We identified 7 host-specific genetic markers derived from human and cow fecal samples by amplifying 16S rDNA fragments from *Bifidobacterium* and the *Bacteroides-Prevotella* group and analyzing them by Length Heterogeneity Polymerase Chain Reaction (LH-PCR) and Terminal Restriction Fragment Length Polymorphism (T-RFLP). Host-specific patterns suggested species composition differences in *Bifidobacterium* and *Bacteroides-Prevotella* populations of human and cow feces. The patterns were highly reproducible among different hosts of the same species. Ease of detection and longer survival in water made *Bacteroides-Prevotella*
better than Bifidobacterium as indicators. Using cloning and sequencing techniques, we
recovered fecal 16S rDNA sequences corresponding to our Bacteroides-Prevotella
markers, none of which matched published Bacteroides or Prevotella sequences
exactly. Phylogenetic analysis of these sequences placed them into three host-specific
gene clusters within the Cytophaga-Flavobacter-Bacteroides division. 16S rDNA
clones from coastal waters recovered with the same Bacteroides-Prevotella primers
clustered with fecal sequences, but did not match them exactly. From the sequence
data, we designed PCR primers specific for each gene cluster. To determine host
specificity of the markers, we tested fecal samples from other animals for the presence
of the genetic markers. Markers derived from human feces were not detected in any
non-human sources. Cow-derived fecal markers were detected in all other ruminants
tested. Using the source-specific primers, we surveyed water samples from Tillamook
Bay, Oregon and its tributaries from June 1998 to April 1999. In most cases, only
areas near sewage treatment facilities and outfalls tested positive for human pollution.
Fecal pollution from ruminant sources was more widely distributed in the bay and
rivers. Although our aim was to identify nonpoint sources of fecal contamination, the
method should be widely applicable for following spatial and temporal fluctuations of
specific bacterial groups in natural environments.
Molecular Markers from Fecal Anaerobes to Identify Nonpoint Source Pollution in Coastal Waters

by

Anne E. Bernhard

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Presented July 28, 2000
Commencement June 2001
Doctor of Philosophy dissertation of Anne E. Bernhard presented on July 28, 2000

APPROVED:

Redacted for privacy

Major Professor, representing Microbiology

Redacted for privacy

Chair of Department of Microbiology

Redacted for privacy

Dean of Graduate School

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

Redacted for privacy

Anne E. Bernhard, Author
ACKNOWLEDGMENTS

I am grateful for the advice and guidance of my major professor, Dr. Katharine Field, and my committee members, Drs. Stephen Giovannoni, James Moore, Evelyn Sherr, and John Fowler. I thank Amanda Barry and Thierry Goyard for assistance with DNA extractions and PCR analyses; Debbie Colbert, Nathan Guzman, James McManus, and James Moore for sample collection; Weerathep Pongprasert and Mike Rappé for their phylogenetic expertise; and Nancy Ritchie and Kevin Vergin for invaluable technical advice. I also thank Abigail Salyers for providing DNA from Bacteroides cultures and Caprice Rosato and Nanci Adair for the GeneScan analyses and sequencing.

This work was partially supported by grant NA76RG0476 (project no. R/ECO-04) from the National Oceanic and Atmospheric Administration to the Oregon State University Sea Grant College Program, by appropriations made by the Oregon State legislature, and by grant R827639-01-0 from the U.S. Environmental Protection Agency. This work was also supported by the Research Council of Oregon State University. Additional funding was provided by the N. L. Tartar Fellowship and the P.F. Yerex and Nellie Buck Yerex Graduate Fellowship.

I also thank my husband, Beto, for always listening and keeping the beer on tap, my parents for keeping the checkbook handy, and, finally, Calvin, Harley, and Zed, for reminding me everyday that there are more important things than work.
CONTRIBUTION OF AUTHORS

Dr. Katharine Field was involved in the design, analysis, and writing of each manuscript.
Thierry Goyard assisted in sample collection, data collection, and analysis for manuscript three.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter 1.</th>
<th>Introduction</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 2.</td>
<td>Identification of Nonpoint Sources of Fecal Pollution in Coastal Waters by Using Host-Specific 16S Ribosomal DNA Genetic Markers from Fecal Anaerobes</td>
<td>14</td>
</tr>
<tr>
<td>2.1 Abstract</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>2.2 Introduction</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>2.3 Materials and Methods</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>2.3.1 Sample collection</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>2.3.2 DNA extractions</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>2.3.3 PCR</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>2.3.4 Restriction endonuclease digestion</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>2.3.5 GeneScan analysis</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>2.3.6 Clone library construction and analysis</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>2.3.7 Sequencing of marker clones</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>2.3.8 Phylogenetic analysis</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>2.3.9 Sensitivity analysis</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2.3.10 GenBank accession numbers</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2.4 Results</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>2.5 Discussion</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>Chapter 3.</td>
<td>A PCR Assay to Discriminate Human and Ruminant Feces Based on Host Differences in Bacteroides-Prevotella 16S Ribosomal DNA</td>
<td>44</td>
</tr>
<tr>
<td>3.1 Abstract</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>3.2 Introduction</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>3.3 Materials and Methods</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>3.3.1 Clones recovered from water samples</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>3.3.2 Primer design</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>3.3.3 PCR sensitivity</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>3.4 Results</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>3.5 Discussion</td>
<td>54</td>
<td></td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS (Continued)

Chapter 4. Application of a New PCR Assay to Discriminate Between Human and Ruminant Fecal Pollution in Tillamook Bay, Oregon ... 56

4.1 Abstract ........................................... 57
4.2 Introduction ........................................ 57
4.3 Materials and Methods ........................................ 59
  4.3.1 Sampling site ...................................... 59
  4.3.2 Sample collection .................................... 62
  4.3.3 DNA extractions ..................................... 62
  4.3.4 PCR ............................................. 62
4.4 Results ................................................. 64
4.5 Discussion .............................................. 69

Chapter 5. Summary ........................................ 73
Bibliography .................................................. 77


### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Map of Tillamook Bay, Oregon</td>
</tr>
<tr>
<td>2.1</td>
<td>LH-PCR analyses of human and cow fecal DNAs amplified with <em>Bacteroides</em>- and <em>Bifidobacterium</em>-specific primers</td>
</tr>
<tr>
<td>2.2</td>
<td>T-RFLP analyses of human and cow fecal DNAs amplified with <em>Bacteroides</em>-specific primers</td>
</tr>
<tr>
<td>2.3</td>
<td>T-RFLP analyses of human and cow fecal DNAs amplified with <em>Bifidobacterium</em>-specific primers</td>
</tr>
<tr>
<td>2.4</td>
<td>Phylogenetic analysis of 16S rDNA clones recovered from human and cow feces</td>
</tr>
<tr>
<td>2.5</td>
<td>T-RFLP analyses of DNAs from Tillamook Bay water samples amplified with <em>Bacteroides</em>-specific primers</td>
</tr>
<tr>
<td>3.1</td>
<td>Phylogenetic analysis of 16S rDNA clones recovered from Tillamook Bay water samples</td>
</tr>
<tr>
<td>4.1</td>
<td>Locations of NPDES and CAFO permitted sites in Tillamook Bay</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Bacteroides and Bifidobacterium-specific primers</td>
</tr>
<tr>
<td>2.2</td>
<td>Potential host-specific genetic markers</td>
</tr>
<tr>
<td>2.3</td>
<td>Presence of genetic markers and levels of fecal coliforms in water samples from Tillamook Bay</td>
</tr>
<tr>
<td>3.1</td>
<td>Host-specific primers</td>
</tr>
<tr>
<td>3.2</td>
<td>Distribution of genetic markers in target hosts</td>
</tr>
<tr>
<td>3.3</td>
<td>Distribution of genetic markers in nontarget hosts</td>
</tr>
<tr>
<td>3.4</td>
<td>Detection limits for Bacteroides-Prevotella 16S rDNA, host-specific genetic markers, and fecal coliforms</td>
</tr>
<tr>
<td>4.1</td>
<td>Primers for human and ruminant-specific genetic markers</td>
</tr>
<tr>
<td>4.2</td>
<td>Detection of source-specific markers in Tillamook Bay from June 1998 to April 1999</td>
</tr>
<tr>
<td>4.3</td>
<td>Distances from sample site to the nearest NPDES or CAFO permitted site</td>
</tr>
<tr>
<td>4.4</td>
<td>Precipitation data</td>
</tr>
</tbody>
</table>
Molecular Markers from Fecal Anaerobes to Identify Nonpoint Source Pollution in Coastal Waters

Chapter 1

Introduction

Anne E. Bernhard
Fecal pollution is a persistent problem, affecting many coastal and inland waters in the United States. Despite improvements in wastewater treatment technology and management practices, fecal contamination from many sources finds its way into our waters, jeopardizing the health of the ecosystems and everything that depends on them. The problem continues partly because current methods are unable to identify the source. Fecal contamination may be introduced into natural waters from a variety of sources, including ineffective sewage treatment, leaking septic systems, illegal dumping, recreational boaters, agricultural runoff, and wildlife.

Point sources, such as sewage treatment plants, can be easily monitored, and appropriate actions taken to reduce or eliminate the discharge. Conversely, nonpoint sources, such as runoff from urban and rural areas, are much more difficult to trace, and often constitute a significant portion of the contamination. Prior to the 1970's, pollution from agriculture and wildlife were considered "natural and uncontrollable" (Martin 1997). However, several high profile incidents involving fish kills and shellfish bed closures increased public awareness, and in 1971, the United States Army Corps of Engineers began issuing permits for discharging wastes into navigable streams and their tributaries (Martin 1997). Since then, numerous studies have investigated the effects of agricultural runoff on bacterial pollution in natural waters (Patni et al. 1985; Fernandez-Alvarez et al. 1991; Niemi and Niemi 1991; Miner and Dorsey-Kramer 1994; Edwards et al. 1997; Stoddard et al. 1998; Baudart et al. 2000). Runoff from agricultural land is now considered one of the major sources of nonpoint pollution.

One of the more obvious ways that fecal pollution interferes with human activities is the introduction of pathogens commonly found in feces. Bacterial pathogens include organisms such as Salmonella, Shigella, E. coli O157:H7, Cryptosporidium, and Giardia. Viruses such as Hepatitis, Norwalk, and other enteroviruses are also often associated with feces and are commonly detected in environmental water samples (Havelaar et al. 1993; Paul et al. 1997; Griffin et al. 1999; Vantarakis and Papapetropoulou 1999). Many of
these bacteria and viruses survive in the environment (Rhodes et al. 1983; Sinton et al. 1994; Mezrioui et al. 1995; Bosch et al. 1997; Murrin and Slade 1997; Pallin et al. 1997; Sinton et al. 1999), and thus, are cause for concern.

In addition to pathogens carried in feces, large amounts of nutrient-rich material are introduced into receiving waters. This sudden surge of nutrients, such as nitrogen, phosphorus, and organic carbon, can cause severe imbalances in the ecosystem. For example, Cloern and Oremland (Cloern and Oremland 1983) found increases in decomposition and nitrification, leading to anoxia, following the discharge of a large volume of primary-treated sewage into San Francisco Bay. Other studies have focused on the effects of nutrient enrichment from agricultural runoff. Phosphorus found in runoff from agricultural land was shown to increase as the proportion of land used in agricultural practices, such as dairy waste application, increased (McFarland and Hauck 1999). This is significant because of the potential for eutrophication in phosphorus-limited freshwaters (Daniel et al. 1998). Additionally, allochthonous material from nonpoint sources can induce algal blooms, some of which may be harmful (see Paerl 1988 for review).

Because of these imminent threats, fecal pollution is frequently monitored in many coastal waters, especially those areas used for shellfisheries and recreation. The most commonly used measure of fecal pollution is the number of viable fecal coliforms in a water sample (American Public Health Association 1992). Fecal coliforms are defined as gram negative, non-sporulating, rod-shaped bacteria that ferment lactose with gas formation within 24 hours at 44.5°C (American Public Health Association 1992). They have been used as indicator bacteria for many years and are currently the Environmental Protection Agency standard for assessing water quality.

Despite the popularity of fecal coliform tests, many researchers and water resource managers have begun to question the validity of these measurements (see Toranzos 1991 for review). Although these bacteria typically originate from fecal material, fecal coliforms can survive and grow outside of enteric habitats (Flint 1987; Alkan et al. 1995; Davies et
al. 1995; Mezrioui et al. 1995; Bogosian et al. 1996; Stoddard et al. 1998). They often settle in sediments, where they can grow, and then be resuspended during mixing events (Sherer et al. 1988; Sherer et al. 1992; Davies et al. 1995). Thus, measurements of fecal coliforms may not accurately reflect recent contamination.

The uncertainty of fecal coliform measurements has led researchers to seek new indicator organisms, hoping to avoid some of the drawbacks of coliform detection. Other fecal organisms such as fecal streptococci, and anaerobic bacteria such as Clostridium, Bifidobacterium, and Bacteroides spp. have been proposed and evaluated as indicators in various environments with varying degrees of success (Resnick and Levin 1981; Allsop and Stickler 1985; Carrillo et al. 1985; Fiksdal et al. 1985; Sorenson et al. 1989; Kreader 1995; Straub and Dixon 1997; Bradley et al. 1999; Griffin et al. 1999). Fecal streptococci are prone to some of the same problems as coliforms, namely that they can survive outside of the intestinal tract (Kibbey et al. 1978; Chandler et al. 1981; Toranzos 1991; Sinton et al. 1993). Anaerobic organisms, however, are less likely to survive once released into the environment compared to facultative organisms like fecal coliforms and fecal streptococci. Although Clostridium spp. are common in feces, they are also common inhabitants of the soil microflora, and thus, their presence does not necessarily indicate fecal contamination. Additionally, Clostridia can form spores that are resistant to harsh environmental conditions, and can survive many years in the environment (Madigan et al. 1997). Evaluation of Bifidobacterium spp. as fecal indicators suggests that these bacteria survive for only a few hours, and thus may be useful only for very recent contamination (Resnick and Levin 1981; Carrillo et al. 1985). Bacteroides, however, can survive for several days after release into the environment (Avelar et al. 1998; Kreader 1998), and are more abundant in feces than coliforms (Savage 1977). Although some of these alternative indicators hold promise as a replacement for coliform detection, none has been evaluated sufficiently for widespread use.
Some of the most frequently suggested candidates for fecal indicators are bacteriophages, including coliphages (Goyal et al. 1980; Singh and Gerba 1983; Palmateer et al. 1991), F-specific phages (Havelaar et al. 1993), and phages specific for *Bacteroides fragilis* (Jofre et al. 1986; Tartera and Jofre 1987; Tartera et al. 1989). Bacteriophages are viruses that infect bacteria, and are typically very specific for their host. Evidence suggests that bacterial indicators are not always correlated with the presence of enteric viruses (Berg et al. 1978; Sobsey 1989; Ferguson et al. 1996), so viral indicators might be better predictors of the presence of pathogenic enteric viruses (Grabow et al. 1983; Stetler 1984). Field studies, however, have not consistently borne out this expected relationship (Havelaar et al. 1993; Ferguson et al. 1996; Griffin et al. 1999). In some cases, phages may be able to infect other host bacteria in the environment, thus increasing their chances of replication (Ferguson et al. 1996). Confounding these problems, distribution of bacteriophages is variable. For example, Tartera and Jofre (Tartera and Jofre 1987) detected *B. fragilis* phages in only 10% of human fecal samples.

In addition to the uncertainties regarding phage ecology, there are some problems associated with the methodology used to detect them. The most frequently used method to detect bacteriophages is the plaque assay, which requires an appropriate host for infection, and strain specificity is not always known. If the correct host is not used, contamination may be over- or underestimated, if the host is too inclusive or restrictive, respectively (Rose et al. 1997). Additionally, phage recovery efficiency can vary dramatically, ranging from 11 to 100% (Goyal et al. 1980; Singh and Gerba 1983). Based on these results, it seems unlikely that bacteriophages are an adequate alternative to fecal coliforms.

Perhaps some of these alternative indicators may emerge as adequate measures of fecal pollution, but none of these methods, including fecal coliforms, allows for consistent, reliable identification of the pollution source. Many fecal bacteria are widely distributed among humans and other animals, making it impossible to identify the source of the
contamination based on numbers of bacteria. Since the ultimate goal is not simply monitoring the amount of fecal contamination, but actually reducing or eliminating it, methods must be developed that can identify the origin of the contamination. By determining the source, resources can be allocated appropriately to correct the problem.

Consequently, source identification has been the subject of much research, leading to several methods that have been used with varying degrees of success. It was initially thought that the ratio of fecal coliforms to fecal streptococci could distinguish human from animal sources (Geldreich 1976), but further research found these ratios were frequently skewed because of survival differences of these organisms (Patni et al. 1985; American Public Health Association 1992; Howell et al. 1996; Edwards et al. 1997). Others have found correlations between the presence of certain bacteriophages and human fecal contamination (Osawa et al. 1981; Jofre et al. 1986; Tartera and Jofre 1987), but these methods have not been consistent and are reliable only under certain conditions.

Using a different approach, several researchers have investigated the potential of antibiotic resistance patterns of fecal coliforms (Kaspar et al. 1990; Parveen et al. 1997) and fecal streptococci (Wiggins 1996; Wiggins et al. 1999) to discriminate human and animal sources. Both methods require creating a large database by testing thousands of fecal isolates from potential sources against as many as thirteen antibiotics at multiple concentrations. The accuracy of classification depends on the antibiotics used and the number of isolates analyzed, and varies from 64 to 100%. The extent to which survival of fecal coliforms and streptococci affects the results of these tests has not been addressed.

In a recent field study in Virginia, antibiotic resistance patterns of fecal streptococci were used to classify fecal pollution sources, identifying cattle as the predominant source (Hagedorn et al. 1999). Based on these results, cattle access to streams was restricted by fencing, resulting in a 94% reduction in fecal pollution, and less than 45% of the fecal streptococci isolates were classified as being from cattle after fencing. Despite the variability in accuracy, the method proved useful in reducing fecal pollution entering the
waterways. However, the method is very labor-intensive, leaving room for better, less arduous methods.

The application of molecular methods is perhaps the most promising approach to source differentiation of fecal bacteria. Molecular methods have the advantage of being specific, sensitive, and rapid. Oligonucleotide probes and primers that target ribosomal RNA genes can range in specificity from the kingdom level down to the strain level (Pace et al. 1986), making these methods very versatile as well. Many researchers have developed gene probes to monitor microbes in environmental samples (see Sayler and Layton 1990; Pickup 1991; Atlas et al. 1992; Pompepuy and Le Guyader 1998 for reviews), including Cryptosporidium (Johnson et al. 1995), E. coli (Bej et al. 1990; Bej et al. 1991; Tsai et al. 1993; Farnleitner et al. 2000), Salmonella (Knight et al. 1990), and several enteric viruses (Bosch et al. 1997; Murrin and Slade 1997; Pallin et al. 1997; Griffin et al. 1999; Vantarakis and Papapetropoulou 1999). These methods have proven to be superior in specificity and sensitivity compared to traditional culture methods.

Gene probes are very useful for detecting single cells and species from a complex sample, but other methods have emerged recently that enable microbial ecologists to characterize whole communities in a single analysis. One of these methods, Length Heterogeneity PCR (LH-PCR) (Rappé et al. 1998; Suzuki et al. 1998), discriminates among mixtures of bacterial gene sequences by detecting differences in the number of base pairs in a particular gene fragment (Avaniss-Aghajani et al. 1994). The method first requires selection and amplification of specific gene sequences using fluorescently-labeled primers and the PCR. Gene fragments are analyzed by separating them by size on an automated DNA sequencer equipped with a fluor-detector that registers the size of the fragments, measured by their electrophoretic mobilities, and their relative abundance, measured by the relative fluorescence. These data can then be translated into an electropherogram using GeneScan software, with the position of each peak indicating the
fragment size and the area under the peak the relative abundance of that fragment. The benefits of this technology are that it is rapid and has the potential to be quantitative.

Several studies have shown the utility of this method to characterize complex microbial communities from aquatic and terrestrial habitats. Two studies characterized marine communities of bacterioplankton (Suzuki et al. 1998) and phytoplankton (Rappé et al. 1998) by using LH-PCR to generate small subunit rRNA gene profiles. Ritchie and others (Ritchie et al. 2000) found LH-PCR to be highly reproducible and used the method to identify microbial communities indicative of particular regions or soil management practices. Fisher and Triplett (Fisher and Triplett 1999) found distinctly different community profiles when they used the method to analyze differences in the rRNA intergenic spacer region of microbial community DNA from several different freshwater environments. These studies show that LH-PCR is effective for quickly and reproducibly characterizing microbial communities.

Terminal Restriction Fragment Length Polymorphism (T-RFLP) is a variation of LH-PCR, in which the PCR amplicons are cut with restriction endonucleases and only the labeled terminal fragments are detected. Due to the sequence specificity of each restriction endonuclease, T-RFLP analysis adds another level of sequence discrimination compared to LH-PCR. Computer simulation of length variation in terminal restriction fragments of PCR products from the 16S rRNA gene indicates that this method can be used to characterize complex bacterial populations (Brunk et al. 1996). In fact, this method has proven very powerful for characterizing complex microbial communities at spatial and temporal scales in a variety of habitats (Bruce 1997; Liu et al. 1997; Clement et al. 1998; Moeseneder et al. 1999; Flynn et al. 2000; Scala and Kerkhof 2000). Although most studies have analyzed fragments of the 16S rRNA gene, the method has proven equally useful for analyses of other genes, such as denitrifying genes (Scala and Kerkhof 2000) and mercury resistance genes (Bruce 1997). Additionally, some investigators were able to make preliminary identifications of community members when analyses from multiple
endonuclease digestions were compared to predicted restriction fragments from sequences in public databases (Liu et al. 1997). These studies demonstrate the potential application of gene fragment length polymorphisms for microbial community analyses in complex environments.

We applied LH-PCR and T-RFLP to characterize the communities of Bacteroides and Bifidobacterium in human and cow feces, and to identify differences between the two hosts. Although different host species may harbor many of the same major bacteria, we hypothesized that there would be some bacterial species that were particular to each host. Other researchers have shown potential host-specific species among fecal bacteria (Kreader 1995; Wang et al. 1996). Once identified, these unique bacterial species would then be used to identify the source of fecal contamination in coastal waters.

We chose to target Bacteroides and Bifidobacterium for several reasons. First, they are both strict anaerobes, which should reduce their chances of survival in natural waters, and thus, would likely represent only recent contamination. Second, they are commonly isolated from human and animal feces, and in the case of Bacteroides, make up a significant portion of the fecal microflora (Macy and Probst 1979; Macy 1981). These organisms have not been used in the past because they are more difficult to grow than facultative anaerobes, such as fecal coliforms.

Bacteroides spp. are anaerobic, gram-negative, nonsporulating rods that inhabit regions of the gastrointestinal tract, such as the mouth, cecum, colon, and rumen. They are known to metabolize polysaccharides, and to transform bile salts and steroids into various metabolites, some of which are implicated in colon cancer (Macy 1981). Many of the species of Bacteroides are nonpathogenic, although some species, such as B. fragilis and B. thetaiotaomicron are opportunistic pathogens commonly isolated from human clinical specimens (Finegold 1977). Because of their intestinal and clinical importance, many methods have been developed to identify and detect Bacteroides spp. (Salyers et al. 1983; Kuritza et al. 1986; Roberts et al. 1987; Attwood et al. 1988; Morotomi et al. 1988;
Rocha and Smith 1995; Doré et al. 1998; Miyamoto and Itoh 1999). Most of these methods have focused on molecular technology, using oligonucleotide probes and primers, and DNA:DNA hybridization techniques.

The genus *Bacteroides* is a member of the *Cytophaga-Flavobacter-Bacteroides* (CFB) division, a phylogenetically similar, but physiologically disparate, group (Gherna and Woese 1992). Recently, Shah and Collins (Shah and Collins 1988; Shah and Collins 1989; Shah and Collins 1990) proposed that the genus *Bacteroides* be reclassified and divided into three groups: (i) *Bacteroides* sensu stricto, which consists of the saccharolytic, nonpigmenting species, such as *B. fragilis*; (ii) *Prevotella*, which consists of moderately saccharolytic, bile-sensitive, predominantly oral species; and (iii) *Porphyromonas*, which consists of assaccharolytic, black-pigmenting species. We chose to target *Bacteroides* and *Prevotella* species in our study, since *P. ruminicola* (previously *Bacteroides ruminicola*) has been identified as a common rumen bacterium (Stewart and Bryant 1988).

*Bifidobacterium* is the third most common genus of the gut microflora, after *Bacteroides* and *Eubacterium*. The genus consists of gram-positive pleomorphic rod-shaped bacteria that ferment carbohydrates (Scardovi 1984). The genus *Bifidobacterium*, unlike *Bacteroides*, is a monophyletic group, falling into the high G+C gram positive subdivision (Woese 1987). They are important in maintaining general health of the intestinal microflora (Rasic and Kurmann 1983), and have been the subject of much research for use as probiotics (Rolfe 2000). Because of their importance in human health, many researchers have developed specific probes and primers to detect and quantify members of this intestinal genus (Yamamoto et al. 1992; Langendijk et al. 1995; Matsuki et al. 1999). Studies of the species composition of *Bifidobacterium* in humans have suggested that populations are characteristic of particular hosts (McCartery et al. 1996; Kimura et al. 1997), and their presence in bovine rumens appears to be sensitive to diet, particularly stimulated by high carbohydrate diets (Trovatelli and Matteuzzi 1976).
The abundance and distribution of *Bifidobacterium* and *Bacteroides-Prevotella* in humans and cows make them good choices for potential indicators of fecal contamination in areas where these sources are the primary concerns, such as Tillamook Bay, Oregon. Previous studies in this bay have identified sewage treatment plants, leaking septic systems, and dairy farm waste as the most likely sources of pollution in the bay (Crane and Moore 1986; Plummer 1995; Newell 1998; Strittholt et al. 1998).

Tillamook Bay is located on the northwest coast of Oregon, approximately 80 km west of Portland; it is 15 km long and 8 km wide, and houses one of the state’s largest commercial and recreational shellfish areas (Figure 1.1). The bay is fed by five major rivers. The Miami River empties into the north end of the bay, near the mouth of the estuary. The other four rivers, the Tillamook, the Trask, the Wilson, and the Kilchis, are in the southeast portion of the bay, and have created a large floodplain. This floodplain is mostly developed, with the majority in dairy farms. The remainder of the drainage basin consists mainly of steep forested slopes.

The climate in the Tillamook watershed is under a strong marine influence from the Pacific Ocean, leading to very wet winters and dry summers. Average annual rainfall is approximately 100 inches, with most of this falling during the fall and winter months, especially during large winter storms. As a result of these storms, flooding is a common occurrence during winter, and flood waters often carry large amounts of fecal material into the bay and rivers.

We chose Tillamook Bay as our study site for several reasons. First, Tillamook Bay is a participant in the Environmental Protection Agency’s National Estuary Program, which targets estuaries in need of restoration and protection. In Tillamook Bay, habitat degradation, sedimentation, and bacterial contamination were identified as the three most serious threats to the estuary’s health. Second, there are several ongoing projects in Tillamook Bay, providing excellent opportunities for collaboration with other scientists. And, third, because of the high concentration of dairy farms, the importance of
Figure 1.1. Tillamook Bay and its tributaries.
shellfisheries in the bay, and the large amount of rainfall, Tillamook Bay is a prime site for
developing methods of source differentiation of fecal contamination.
Chapter 2

Identification of Nonpoint Sources of Fecal Pollution in Coastal Waters by Using Host-Specific 16S Ribosomal DNA Genetic Markers from Fecal Anaerobes

Anne E. Bernhard and Katharine G. Field

Published in *Applied and Environmental Microbiology* 66: 1587-1594
American Society for Microbiology, Washington, D. C.
2.1 ABSTRACT

We describe a new PCR-based method to discriminate human and cow fecal contamination in coastal waters without culturing indicator organisms, and show that the method allows bacterial marker sequences to be tracked in complex environments. We identified 2 human- and 5 cow-specific genetic markers from fecal samples by amplifying 16S rDNA fragments from Bifidobacterium and the Bacteroides-Prevotella group and analyzing them by Length Heterogeneity Polymerase Chain Reaction (LH-PCR) and Terminal Restriction Fragment Length Polymorphism (T-RFLP). Host-specific patterns suggested species composition differences in Bifidobacterium and Bacteroides-Prevotella populations of human and cow feces. The patterns were highly reproducible among different hosts of the same species. Additionally, all host-specific genetic markers were detected in water samples collected from areas frequently contaminated with fecal pollution. Ease of detection and longer survival in water made Bacteroides-Prevotella better than Bifidobacterium as indicators. Fecal 16S rDNA sequences corresponding to our Bacteroides-Prevotella markers comprised closely-related gene clusters, none of which exactly matched published Bacteroides or Prevotella sequences. Our method detected host-specific markers in water at concentrations ranging from $2.8 \times 10^{-5}$ to $2.8 \times 10^{-7}$ g dry weight feces/L and $6.8 \times 10^{-7}$ g dry weight sewage/L. Although our aim was to identify nonpoint sources of fecal contamination, the method should be widely applicable to follow spatial and temporal fluctuations of specific bacterial groups in a natural environment.
2.2 INTRODUCTION

Fecal pollution is a serious environmental problem afflicting many coastal regions in the U. S. and worldwide. Pathogens associated with fecal pollution can lead to human disease and economic loss in industries that depend on coastal waters, such as shellfisheries (Crane and Moore 1986). Despite efforts to minimize fecal input to coastal waterways, the problem persists, partly due to inability to reliably identify nonpoint sources. These sources may include inefficient sewage treatment plants, leaking septic systems, agricultural runoff, or wildlife (Strittholt et al. 1998). Knowing the source of the contamination is crucial to effective resource management and, ultimately, cessation of the problem.

The most widely used method for measuring fecal pollution is to quantify viable fecal coliforms by culturing them (American Public Health Association 1992). This method, however, does not identify the source of the fecal contamination. In addition, the extent that fecal coliforms settle, grow and are resuspended after release into receiving waters remains controversial (Davies et al. 1995), leaving the method's accuracy in question. Thus, there is a need for a reliable method to identify the source of fecal pollution that does not rely on measuring coliform concentrations.

_Bacteroides_ and _Bifidobacterium_ have been suggested by several researchers as possible indicator organisms (Resnick and Levin 1981; Allsop and Stickler 1985; Fiksdal et al. 1985). Both are strict anaerobes, restricted to warm-blooded animals and, unlike coliforms, make up a significant portion of fecal bacteria. Additionally, because they are strict anaerobes, they do not survive very long once released into receiving waters
(Resnick and Levin 1981; Carrillo et al. 1985; Avelar et al. 1998; Kreader 1998). Their use as indicators, however, has been limited because strict anaerobes are often difficult to culture. The difficulty of growing strict anaerobes can be circumvented by using molecular, rather than culture-based, methods to detect them.

Molecular approaches have become popular and efficient methods to characterize and track changes in community structure of microbial populations (e.g. Murray et al. 1996; Suzuki et al. 1998; Glockner et al. 1999). Our approach, however, was to identify and track spatial and temporal fluctuations of individual bacterial markers in a natural environment. We used recently developed technologies that discriminate among mixtures of bacterial gene sequences by detecting differences in the number of base pairs in a particular gene fragment (Avaniss-Aghajani et al. 1994; Brunk et al. 1996). Length Heterogeneity Polymerase Chain Reaction (LH-PCR) (Suzuki et al. 1998) and Terminal Restriction Fragment Length Polymorphism (T-RFLP) (Bruce 1997; Liu et al. 1997; Clement et al. 1998) analyze differences in lengths of gene fragments due to insertions and deletions and estimate the relative abundance of each fragment. To track a bacterial marker sequence, it first must be uniquely identified using a combination of specific primers for PCR and appropriate restriction enzymes. Once a reliable identification has been made, the signal can be followed easily and quickly from many samples.

Our goals were first, to develop 16S rDNA markers, based on fecal anaerobes, that discriminate human from cow fecal pollution; and second, to show that these markers could be recovered and identified from natural waters, indicating their potential utility in measuring and discriminating fecal pollution.
Starting with cow and human fecal samples, we used LH-PCR and T-RFLP to characterize the community profiles of *Bacteroides-Prevotella* and *Bifidobacterium*, looking for unique host-species-specific patterns or markers. We were able to identify host-specific markers that were highly reproducible among hosts of the same species, but the *Bifidobacterium* markers were more difficult to detect.

To show that these markers could also be recovered from natural waters, we analyzed water samples from Tillamook Bay, Oregon. Pollution in the bay, which houses a major shellfish industry, is thought to be mostly of dairy cow origin, but human fecal contamination from septic tanks and sewage treatment cannot be ruled out. We were able to recover our host-specific markers from water samples, and sequence analysis confirmed their identities.

### 2.3 Materials and Methods

#### 2.3.1 Sample collection

*Fecal samples:* Human fecal samples were donated by healthy adult and child volunteers from Corvallis, Oregon, including Caucasian, Asian and Hispanic individuals. Samples were collected in sterile containers and stored at -80°C. Fresh cow fecal samples were collected from healthy Holstein dairy cows from two farms in Corvallis, Oregon, and three farms in Tillamook County, Oregon. We collected Corvallis cow fecal samples during 3 different seasons from 1996 to 1998 and Tillamook County cow fecal
samples during Fall 1996. Samples were collected with sterile utensils and placed in sterile 50 ml tubes, kept on ice for transport to the lab, and stored at -80 °C.

**Water samples:** Samples were collected from multiple bay and river sites in the Tillamook Watershed and from sewage treatment facilities in Corvallis and Tillamook, OR. We selected sites representing three rivers, the Tillamook, the Trask, and the Wilson, that are frequently contaminated with fecal pollution, and four sites along a north-south transect starting near the confluence of the Tillamook and Trask rivers and ending at a site near the mouth of the estuary. Water samples were collected in sterile 1-liter containers from surface waters and stored on ice during transport to the lab. Upon return to the lab, we filtered water samples through 0.2 μm Supor-200 filters. Filters were placed in separate plastic bags or 50 ml disposable centrifuge tubes with 5 mls of lysis buffer (20 mM EDTA, 400 mM NaCl, 750 mM sucrose, 50 mM Tris, pH 9) and stored at -80°C. We measured fecal coliforms according to standard methods (American Public Health Association 1992).

### 2.3.2 DNA extractions

**Fecal samples:** We extracted DNA by bead-beating according to the method of Gray and Herwig (Gray and Herwig 1996) with the following modifications: 0.5 g of 0.1 mm glass beads (acid-washed and baked) were used, polyvinylpolypyrrolidone was omitted from the lysis buffer, and crude extracts were ethanol-precipitated, the resulting pellets dried under vacuum and resuspended in InstaGene Matrix (BioRad, Hercules, CA) or TE (10 mM Tris, 1 mM EDTA, pH 8). The DNA extracts were purified by
phenol/chloroform extractions followed by ethanol precipitation and resuspension in 
TE.

*Water samples:* DNA was extracted based on the method of Giovannoni
(Giovannoni *et al.* 1990) except the CsTFA purification steps were omitted. Instead,
samples were cleaned using one of the following methods: 1) one volume of 20% 
polyethylene glycol 8000 in 2.5 M NaCl was added and the samples were incubated for 
15 minutes at 37°C, centrifuged for 10 minutes at 12,500 rpm, and the resulting pellets 
were washed twice with ice-cold 80% ethanol; 2) guanidine thiocyanate (Fluka, Buchs,
Switzerland) purification based on the method of Pitcher and colleagues (*Pitcher et al.* 
1989), or 3) polyvinylpolypyrrolidone (Aldrich, Milwaukee, WI) spin columns 
(Berthelet *et al.* 1996).

### 2.3.3 PCR

Approximately 2-4 ng of fecal DNAs from individual humans and cows were 
amplified by the PCR. In addition to analyzing individual samples, we also analyzed 
pooled PCR products from multiple individuals from each host species. DNAs from 
fourteen human samples were amplified with both *Bacteroides-Prevotella* and 
*Bifidobacterium* primers (Table 2.1). DNAs from 8 Corvallis and 8 Tillamook cows 
were amplified with *Bacteroides-Prevotella* primers, but only 4 each with 
*Bifidobacterium* primers. Each 50 µl PCR contained 1X *Taq* polymerase buffer, 10 µM 
each primer, 200 µM each dNTP, 1.25 units of *Taq* polymerase, 640 ng/µl BSA
(Kreader 1996), and 1.5 mM MgCl$_2$. Bif601R and Bac32F were labeled with the fluorophore 6-FAM (GenSet, La Jolla, CA). Non-fluorophore labeled primers (Bac303R, Bac708R, and Bif164F) were synthesized by Gibco BRL (Gaithersburg, MD). New primers (Bif601R, Bac32F, and Bac708R) were designed using the Probe Design function of ARB (Strunk and Ludwig 1996) and confirmed using CHECK_PROBE analysis of the Ribosomal Database Project (Maidak et al. 1994) and Probe Match of ARB. We established primer specificity using DNA from cultured Bacteroides and Bifidobacterium. A thermal mini-cycler (MJ Research, Watertown, MA) was used for all reactions with the following conditions: 35 cycles of 94°C for 30s, 53°C for 1 minute, 72°C for 2 minutes followed by a final 6 minute extension at 72°C. We quantified the products in a 1% agarose gel by comparing the band intensity to a low molecular weight DNA mass ladder (Gibco BRL).

2.3.4 Restriction endonuclease digestion

We chose restriction enzymes based on analysis of published sequences in GenBank using Mapsort (Genetics Computer Group, Wisconsin). Enzymes that produced the greatest number of different length terminal restriction fragments within the Bacteroides-Prevotella or Bifidobacterium 16S rDNA sequences were tested empirically. Enzymes were purchased from New England Biolabs (Beverly, MA). PCR products amplified using Bac32F and Bac708R were digested overnight at 37°C with either Acil or HaeIII. PCR products amplified using Bif164F and Bif601R were
digested overnight with HaeIII (at 37°C) or TaqI (at 65°C). Each 20 µl digestion contained 20-40 ng of PCR products, 10 units of enzyme, 1X enzyme buffer, and 100 µg/ml BSA (for TaqI only).

Table 2.1 Primers used in this study. Bac=\textit{Bacteroides-Prevotella}, Bif=\textit{Bifidobacterium}; numbering corresponds to the \textit{E. coli} 16S rRNA gene.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’-3’)</th>
<th>Target</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bac32F</td>
<td>AACGCTAGCTACAGGCTT</td>
<td>\textit{Bacteroides-Prevotella}</td>
<td>This study</td>
</tr>
<tr>
<td>Bac303R</td>
<td>CCAATGTGGGGGACCTTC</td>
<td>\textit{Bacteroides-Prevotella}</td>
<td>Modified from Manz \textit{et al.} 1994</td>
</tr>
<tr>
<td>Bac708R</td>
<td>CAATCGGAGTCTTTCCGTG</td>
<td>\textit{Bacteroides-Prevotella}</td>
<td>This study</td>
</tr>
<tr>
<td>Bif64F</td>
<td>GGGTGGTAATGCCGGATG</td>
<td>\textit{Bifidobacterium}</td>
<td>Langendijk \textit{et al.} 1995</td>
</tr>
<tr>
<td>Bif601R</td>
<td>TAAGCGATGGACCTTCACACC</td>
<td>\textit{Bifidobacterium}</td>
<td>This study</td>
</tr>
</tbody>
</table>

2.3.5 GeneScan analysis

All analyses were done using samples from individuals as well as host-specific community samples. Approximately 25 fmols of PCR products or restriction digest products were resolved on a Long Ranger polyacrylamide gel (FMC, Rockland, ME) on an ABI 377 automated DNA sequencer using GeneScan software (Applied Biosystems Inc., Fremont, CA). The internal size standard, GENESCAN2500-ROX (ABI) was loaded in each lane. Fragment sizes were estimated using the Local Southern Method in GeneScan software v. 2.1 (ABI).
2.3.6 Clone library construction and analysis

Fecal DNAs from individual cows or humans were amplified with Bac32F and Bac708R and amplicons from 10 individuals from each host species were pooled. PCR products were gel purified using the Qiaquick Gel Extraction Kit (Qiagen, Valencia, CA) and cloned using the pGEM-T Easy Cloning Kit (Promega, Madison, WI) according to the manufacturer's directions. One hundred ninety-two transformants were randomly selected from each library and inoculated into 100 μl LB broth with 100 μg/ml ampicillin in 96-well microtiter plates. After a six-hour incubation, two replica plates were made from each original microtiter plate. All plates were incubated overnight at 37°C. The following day, clones from each row in a microtiter replica plate and clones from each column in another microtiter replica plate were pooled.

DNAs from the pooled rows and columns were amplified using Bac32F and either Bac303R or Bac708R. Bac32F was labeled with the fluorophore 6-FAM. PCR products amplified with Bac32F and Bac303R were analyzed by LH-PCR. PCR products amplified with Bac32F and Bac708R were digested with the restriction enzymes HaeIII or AcIi as described above and analyzed by T-RFLP. The clones corresponding to each genetic marker in the LH-PCR or the T-RFLP were identified by locating the intersection of a positive result in a row with a positive result in a column.
2.3.7 Sequencing of marker clones

Plasmid DNAs from overnight cultures were prepared using the Qiaprep Spin Column Purification Kit (Qiagen) according to the manufacturer's directions. DNA was quantified spectrophotometrically on a Shimadzu UV/Vis spectrophotometer. Bidirectional sequences were obtained using T7 and SP6 priming sites on either side of the insert. Sequences were determined on an ABI 377 DNA Sequencer using dye terminator chemistry.

2.3.8 Phylogenetic analysis

Sequences were submitted to BLAST v. 2.0 to obtain preliminary closest phylogenetic neighbors. We aligned the sequences manually to sequences from the Cytophaga-Flavobacter-Bacteroides group obtained from GenBank using the DNA sequence editor in GCG v.10 (Genetics Computer Group, Wisconsin). Sequences and alignments were verified by comparisons to the 16S rRNA secondary structure of Bacteroides fragilis and to Bacteroides signature sequences (Gherna and Woese 1992). Evolutionary distances were calculated using the DNADIST program with the Kimura 2-parameter model for nucleotide change and a transition/transversion ratio of 2.0 (Kimura 1980). Phylogenetic trees were inferred by the neighbor-joining algorithm (Saitou and Nei 1987) using the NEIGHBOR program in PHYLIP 3.5c (Felsenstein 1989). Regions of ambiguous alignment were excluded from the analyses. To check the consistency of the resulting tree, we randomly resampled the sequences 100 times.
(bootstrapping) and obtained a consensus tree (Felsenstein 1985). Similarities were calculated using Similarity Matrix v.1.1 from the Ribosomal Database Project.

2.3.9 Sensitivity analysis

Serial dilutions of fresh cow feces or raw sewage were added to 1-L samples of filter-sterilized bay water. Final concentrations in the 1-L samples ranged from \(2 \times 10^{-7}\) mg (wet weight)/L to 2.0 mg/L. Samples were filtered onto a 0.2 µm Supor filter and stored in lysis buffer at \(-80°C\) as described above. The percent solids of the fecal samples was estimated by weighing replicate samples of wet feces and drying with heat until no more weight was lost. To estimate percent solids of raw sewage, we collected the solids by centrifugation, decanted the supernatant, and dried the samples overnight with heat. DNAs extracted from the filters were amplified using Bac32F and Bac708R as above, and the PCR products were visualized in a 1% agarose gel. Products were digested as described above. We analyzed all samples by T-RFLP, using 25 fmols of the most concentrated dilution (2.0 mg/L), and equivalent volumes from all other dilutions.

2.3.10 GenBank accession numbers

The GenBank accession numbers are as follows: AF233400, AF233401, AF233402, AF233403, AF233404, AF233405, AF233406, AF233407, AF233408, AF233409, AF233410, AF233411, AF233412, and AF233413.
2.4 RESULTS

We amplified human and cow fecal DNAs with primers specific for the fecal anaerobes *Bacteroides-Prevotella* and *Bifidobacterium*. We separated the amplified fragments by size on an ABI DNA sequencer with GeneScan software, allowing us to identify DNA fragment lengths unique to either humans or cows. From these analyses, seven potential host-specific 16S rDNA genetic markers from human and cow fecal DNAs were identified (Table 2.2). To be considered a host-specific genetic marker, the gene fragment had to be present in all samples from that host and be absent from all samples from the other host.

Table 2.2 Potential host-specific genetic markers identified by LH-PCR or T-RFLP analysis of human and cow fecal DNAs. Markers are located within the 16S rRNA genes from *Bacteroides* or *Bifidobacterium*.

<table>
<thead>
<tr>
<th>Host specificity</th>
<th>Primer Pair</th>
<th>Enzyme Used</th>
<th>Size of marker fragment (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>human</td>
<td>Bac32F-Bac7O8R</td>
<td><em>HaeIII</em></td>
<td>119</td>
</tr>
<tr>
<td>cow</td>
<td>Bac32F-Bac7O8R</td>
<td><em>HaeIII</em></td>
<td>222</td>
</tr>
<tr>
<td>cow</td>
<td>Bac32F-Bac7O8R</td>
<td><em>AciI</em></td>
<td>227</td>
</tr>
<tr>
<td>cow</td>
<td>Bac32F-Bac3O3R</td>
<td>none</td>
<td>276</td>
</tr>
<tr>
<td>cow</td>
<td>Bifl64F-Bif6OlR</td>
<td><em>HaeIII</em></td>
<td>142-152</td>
</tr>
<tr>
<td>human</td>
<td>Bifl64F-Bif6OlR</td>
<td><em>TaqI</em></td>
<td>313</td>
</tr>
<tr>
<td>cow</td>
<td>Bifl64F-Bif6OlR</td>
<td>none</td>
<td>453</td>
</tr>
</tbody>
</table>
LH-PCR analysis, which detects length differences in PCR amplicons, revealed *Bacteroides-Prevotella* and *Bifidobacterium* cow-specific genetic markers (Figure 2.1). LH-PCR analysis of 16S rDNA amplicons amplified with Bac32F and Bac708R from human and cow feces identified a peak at 276 bp as a potential cow-specific gene fragment, but no human-specific genetic markers were detected. LH-PCR analysis of 16S rDNA amplicons amplified with Bif164F and Bif601R revealed a cow-specific genetic marker at 453 bp.

We identified five additional host-specific genetic markers by cutting *Bacteroides-Prevotella* or *Bifidobacterium* PCR amplicons with restriction endonucleases and analyzing the fluorescently-labeled terminal end fragments for size differences (T-RFLP). Host-specific peaks, corresponding to terminal end fragments, were identified in T-RFLP analyses of human and cow fecal DNAs (Figures 2.2 and 2.3). When PCR products amplified with Bac32F and Bac708R were digested with AcI, one cow-specific peak was found at 227 bp (Figure 2.2a). There were additional host-specific peaks, which, upon sequence analysis, were discovered to be artifacts produced by partial digestion. Analysis of 16S rDNA amplicons from human feces and sewage amplified with Bac32F and Bac708R and digested with HaeIII revealed a 119 bp human-specific peak and a cow-specific peak of 222 bp (Figure 2.2b).

T-RFLP analysis of 16S rDNA genes amplified from cow feces using Bif164F and Bif601R and cut with HaeIII revealed a cow-specific cluster of peaks ranging from 142-152 bp (Figure 2.3a). Analysis of 16S rDNA amplicons from human feces and sewage amplified with Bif164F and Bif601R and digested with TaqI produced a
Figure 2.1 LH-PCR analysis of 16S rDNA gene fragments amplified with Bac32F-FAM and Bac303R (Panel A) and Bif164F and Bif601R-FAM (Panel B). Solid lines represent community profiles from human fecal DNA; dotted lines represent community profiles from cow fecal DNA. Samples are mixtures of DNA from 7-8 individuals. The arrows indicate cow-specific gene fragments.
Figure 2.2 T-RFLP analysis of 16S rDNA gene fragments amplified with Bac32F-FAM and Bac708R and cut with AcII (Panel A) or HaeIII (Panel B). Solid lines represent community profiles from human fecal DNA; dotted lines represent community profiles from cow fecal DNA. Arrows indicate host-specific genetic markers.
Figure 2.3 T-RFLP analysis of 16S rDNA gene fragments amplified with Bif64F and Bif601R-FAM and cut with HaeIII (Panel A) or TaqI (Panel B). Solid lines represent community profiles from human fecal DNA; dotted lines represent community profiles from cow fecal DNA. Arrows indicate host-specific genetic markers.
human-specific peak at 313 bp, but no cow-specific peaks were detected in the amplicons from cow feces (Figure 2.3b).

Comparison of Bacteroides-Prevotella and Bifidobacterium communities in sewage samples from two Oregon cities, Corvallis and Tillamook, and feces from 14 humans revealed very similar community profiles for both Bacteroides-Prevotella and Bifidobacterium, although there were sometimes differences in proportions of LH-PCR and T-RFLP peaks present (data not shown). Similarly, DNA from cow feces collected at different times of the year, from different farms and different towns, revealed very similar patterns. These results suggest that although there may be slight intraspecies variation, at the level of variability detected by these markers, the host-specific patterns are the same.

To verify the identities of the genetic markers, we constructed 16S rDNA clone libraries from cow and human fecal DNAs amplified with the Bacteroides-Prevotella-specific primers Bac32F and Bac708R. We screened 192 clones from each host and sequenced those clones that had the LH-PCR or T-RFLP pattern of interest. Because Bacteroides-Prevotella is a more promising indicator group (see discussion), we cloned 16S rDNA genes from this group only and not from Bifidobacterium. We found six different clones in the library from human feces corresponding to the human-specific marker of 119 bp (Figure 2.4). Further analysis of these sequences revealed that the fragment size estimated by T-RFLP was 1 bp smaller than the actual size (120 bp) determined from the sequences. Four of these sequences (HF8, HF102, HF117 and HF145), although not identical, were >98.9% similar to each other and were 97.5-98.0%
Figure 2.4 Phylogenetic relationships among partial 16S rDNA sequences (558 positions) of human (HF) and cow (CF) host-specific genetic markers identified from fecal clone libraries. The tree was inferred by neighbor-joining. Numbers above the internal branches are percentages of bootstrap replicates that support the branching order. Bootstrap values below 50% are not shown. Cytophaga fermentans was used to root the tree.
similar to \textit{B. vulgatus}. These sequences formed the closely-related HF8 gene cluster (Figure 2.4), but did not match any published sequences exactly. HF74 was 93.9-94.9% similar to the clones in the HF8 cluster and 93.2% similar to \textit{B. vulgatus}. One other human fecal clone, HF10, was 97.7% similar to \textit{B. uniformis}.

None of the cow-specific clones were closely related to any characterized microorganisms. They formed two distinct gene clusters within the genus \textit{Bacteroides} (Figure 2.4). We recovered seven clones from cow feces that produced the 227 bp size fragment when amplified with Bac32F and Bac708R and cut with \textit{Acil}. Partial 16S rDNA sequencing revealed five different sequences, each with the same T-RFLP profile, which formed the CF123 gene cluster. Fragment sizes estimated by T-RFLP analysis were about 2 bases larger than the size determined from the sequences (225 bp). Similarities ranged from 91.6-95.2% within this cluster. Sequence analysis of the clones corresponding to the 222 bp (T-RFLP with \textit{HaeIII}) and the 276 bp (LH-PCR) cow-specific markers revealed that these markers represented the same sequences. We found four clones representing three different sequences that corresponded to these two markers. These three sequences were 92-94.4% similar and were all included in the CF151 gene cluster (Figure 2.4). Again, T-RFLP and LH-PCR estimated fragment sizes were 1-2 bases different from the sizes predicted from the sequences.

We tested a variety of river and estuarine water samples for the presence of \textit{Bacteroides-Prevotella} and \textit{Bifidobacterium} DNA and also for the marker genes. Fecal coliforms in these samples ranged from 0 to 120 CFU/100mls (Table 2.3). \textit{Bacteroides-Prevotella} DNA was detected in all 8 samples tested, but \textit{Bifidobacterium} DNA was
detected in only 2 of these samples. Additionally, the product yield of *Bifidobacterium* amplicons detected in these water samples was considerably less than that obtained from the same samples using *Bacteroides* primers. All seven host-specific genetic markers were detected in at least one water sample (Table 2.3, Figure 2.5). In subsequent experiments, we used sequence data to validate the identities of the *Bacteroides-Prevotella* markers from water samples. We recovered sequences for *Bacteroides-Prevotella* markers belonging to the HF8, CF123, and CF151 gene clusters (data not shown).

**Table 2.3** Fecal coliform measurements and presence/absence of *Bifidobacterium* and *Bacteroides-Prevotella* host-specific markers in water samples collected from Tillamook Bay, Oregon and three of its tributaries. “+” indicates the peak size was detected in the LH-PCR or T-RFLP analysis; “-” indicates no peak was detected of that size; “nd” indicates no data.

<table>
<thead>
<tr>
<th>Date Collected</th>
<th>Sample Type</th>
<th>Fecal Coliforms (CFU/100mls)</th>
<th><em>Bifidobacterium</em></th>
<th><em>Bacteroides-Prevotella</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>12/12/97</td>
<td>river</td>
<td>2</td>
<td>+</td>
<td>nd</td>
</tr>
<tr>
<td>12/12/97</td>
<td>river</td>
<td>6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5/17/98</td>
<td>river</td>
<td>120</td>
<td>-</td>
<td>nd</td>
</tr>
<tr>
<td>10/30/98</td>
<td>river</td>
<td>36</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>10/30/98</td>
<td>mouth</td>
<td>112</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10/30/98</td>
<td>estuary</td>
<td>35</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10/30/98</td>
<td>estuary</td>
<td>57</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>10/30/98</td>
<td>estuary</td>
<td>0</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>
Figure 2.5 T-RFLP analyses of 16S rDNA gene fragments amplified from DNA extracted from Tillamook Bay water samples. DNA was amplified using Bac32F and Bac708R and digested with AciI (Panel A) or HaeIII (Panels B and C). Arrows indicate host-specific markers. Panels A and B show cow-specific markers 227 and 222, respectively. Panel C shows the 119-bp human-specific marker.
We evaluated the sensitivity of detecting host-specific DNA in water samples by conducting assays using filter-sterilized bay water amended with fresh feces or raw sewage. We used feces and sewage rather than cultured organisms since fecal organisms were the intended targets. The limit of detection of host-specific markers varied, with the 222-bp cow-specific marker being the least sensitive \((2.8 \times 10^{-5} \text{ g dry feces/L})\), followed by the 119-bp human-specific marker \((6.8 \times 10^{-7} \text{ g dry sewage/L})\) and the 227-bp cow-specific marker the most sensitive \((2.8 \times 10^{-8} \text{ g dry feces/L})\).

2.5 DISCUSSION

We identified species composition differences in the *Bacteroides-Prevotella* and *Bifidobacterium* populations between human and cow feces. These differences could be useful for identifying fecal pollution nonpoint sources in coastal waters. The human-specific genetic markers within the *Bacteroides-Prevotella* group were closely related, but not identical, to *Bacteroides* species commonly found in human intestines and feces, *B. uniformis* and *B. vulgatus* (Salyers 1984). The cow-specific genetic markers formed new gene clusters within the *Bacteroides-Prevotella* group, which is in the *Cytophaga-Flavobacter-Bacteroides* phylum, a physiologically diverse, but phylogenetically similar, grouping (Paster *et al.* 1985; Gherna and Woese 1992). Gene clusters are sets of gene sequences more closely related to each other than to any characterized species; they have been found in many diverse natural bacterial
populations (e.g. Giovannoni et al. 1990; Hales et al. 1996; Ohkuma and Kudo 1996; Field et al. 1997)

The discovery of new gene clusters within the Bacteroides-Prevotella group from cows reflects the lack of characterization of diversity within this habitat. Conversely, the human intestinal flora is a better characterized habitat, due to the clinical significance of these bacteria. Microbial diversity of human fecal and colonic bacteria has been the subject of many culture-based studies, but only since the application of molecular techniques have researchers had the tools to assess the diversity more accurately. Although culture bias may be less of a problem in enriched, highly selective environments such as feces, it is still likely to occur, especially for anaerobic bacteria that may be difficult to grow (Amann et al. 1995). Comparisons of 16S rDNA diversity with that assessed by culture methods in human feces (Wilson and Blitchington 1996; Wood et al. 1998) and bovine rumens (Krause and Russell 1996) suggest an underestimation of diversity by culturing alone.

The clones comprising the HF8 cluster were >99% similar, with the exception of HF102. These three clones (HF8, HF117, HF145) varied by only 1-2 nucleotides over a 700 base sequence, which falls within predicted Taq polymerase error rates (Saiki et al. 1988). Three of the six deviant nucleotides were consistent with common Taq errors (Dunning et al. 1988; Tindall and Kunkel 1988) and two others were incompatible with secondary structure, suggesting PCR or sequencing errors. We think it is possible that these three sequences are actually the same. Although HF102 was in the same gene cluster, it had 9 to 11 nucleotide differences from the other three sequences in this
cluster. These differences, however, were in a hypervariable region of the gene that was not included in the phylogenetic analysis because of ambiguous alignment.

LH-PCR and T-RFLP proved to be highly reproducible methods, although the estimated peak size often deviated by 1-2 base pairs from the size predicted from the sequence. These discrepancies between peak sizes and predicted sizes may have several explanations. First, there may be slight differences in the electrophoretic mobility of the ROX-labeled standard and the FAM-labeled samples. Second, our observations from analyses of single clones suggest that the addition of an adenine to the end of the PCR products by Taq polymerase is variable, leading to products exactly one base pair different. Third, differences in sequences may cause products to migrate anomalously compared to the same size standard fragment. These variables may all have contributed to the 1-2 base pair variations we observed between the peak sizes and the sizes predicted from the sequences. We also observed that the deviations appeared to increase with the size of the fragment. Despite these discrepancies, the methods were reproducible, with variances of ± 0.3 bp for fragments up to 350 bp.

Our comparisons of Bacteroides-Prevotella and Bifidobacterium gene profiles among 14 humans and 16 cows suggested insignificant intraspecies variation, mostly manifested as differences in proportions rather than species present. Human feces were collected from coworkers and their families, so it is possible that these individuals share intestinal flora (Caugant et al. 1984; Mehta et al. 1999). However, sewage samples from Tillamook and Corvallis (cities separated by 100 miles and a mountain range) also showed nearly identical patterns, suggesting that the host-specific patterns were widely
distributed. This does not mean that the commensal bacterial communities are identical in individuals from geographically distinct populations. Instead, it demonstrates that the method does not reveal variability at the level of the individual, but does reveal variability between host species.

Previous analyses of human fecal flora using culture techniques did not show major differences in bacterial species composition even when populations with different diets were compared (Finegold et al. 1983), although relative frequencies varied among individuals (Holdeman et al. 1976). Other studies, however, suggest major differences in community composition of Bifidobacterium and Lactobacillus in humans (McCartney et al. 1996; Kimura et al. 1997). Our data suggested low intraspecific variation within a bacterial population. These discrepancies may be explained by the differences in the methods used. Culturing bacteria from samples discriminates to the species or even the strain-level. Methods based on sizes and compositions of gene fragments such as LH-PCR and T-RFLP may discriminate only to the phylogenetic group or gene cluster level. It is possible that individuals harbor different species or strains of bacteria within a particular gene cluster, which would not necessarily be detected using the methods presented in this paper.

We concluded that species within the Bacteroides-Prevotella group were better indicators than Bifidobacterium in coastal waters. Although we detected host-species differences in the Bifidobacterium populations, these genetic markers proved to be less robust than those from Bacteroides-Prevotella. First, we were sometimes unable to detect Bifidobacterium in cow fecal samples. Bifidobacterium has been shown to be
numerous in the rumen (Biavati and Mattarelli 1991), but its prevalence in feces may be affected by acidic conditions in the stomach or by the actions of certain antibiotics (Dennis et al. 1981; Stewart and Bryant 1988). We collected samples from only those cows that were not currently being given antibiotics, but the antibiotic history of the individual cows was not considered at the time of collection.

Secondly, detecting *Bifidobacterium* via PCR was troublesome in water samples as well as fecal samples. It is possible that the signal was simply too weak to be detected via PCR. Resnick and Levin (Resnick and Levin 1981) found that *Bifidobacterium* could not be cultured after 5 or 10 hours depending on whether the cells were in fresh or salt water, respectively. This short survival time would make it difficult to detect *Bifidobacterium* in waters where fecal pollution may be much older than 10 hours. Carrillo and colleagues (Carrillo et al. 1985) also found very low survival of *Bifidobacterium adolescentis* in a tropical environment and suggested that *Bifidobacterium* could only be used to detect very recent pollution.

Conversely, host-specific markers within the *Bacteroides-Prevotella* group were detected in all water samples from Tillamook Bay and its tributaries and all fecal samples. Additionally, we did not detect any fecal markers in water samples collected from the Sargasso Sea or Crater Lake, Oregon, neither of which would be expected to have human or cow fecal pollution. Our assay was not designed to quantify fecal pollution, but rather to identify the presence or absence of a particular source; therefore direct comparisons to fecal coliforms are inappropriate. The presence of fecal markers in samples with no detectable coliforms (Table 2.3) is likely the result of differences in
sensitivity or viability of the coliforms. The sample with no detectable coliforms was from the mouth of the estuary, where salinity is highest and coliforms would likely be stressed, or dead. We did not apply methods for resuscitating stressed organisms, so it is unlikely that we would detect stressed organisms by culturing them. DNA from Bacteroides released into water, however, has been detected from several days to two weeks if conditions are optimal (Kreader 1998).

Results from our sensitivity assays are comparable to other studies that used PCR to detect single Bacteroides species in feces (Kreader 1996; Wang et al. 1996; Kreader 1998). The 119-bp human-specific marker and the 227-bp cow-specific marker appear to be more sensitive, as assayed by T-RFLP using general Bacteroides-Prevotella PCR primers. It is possible that by designing primers specific to these markers, the sensitivity may be increased. The 222-bp cow-specific marker, which represents the same sequences as the 276-bp marker, was the least sensitive by 3-4 orders of magnitude. We also observed some samples that tested positive for one, but not the other (Table 2.3). Since the sensitivity for these markers is much lower, it is possible that the source contamination was at or near the limit of detection; therefore, inconsistent detection of these two markers in the same water sample is not surprising.

Bacteroides makes up as much as 30% of fecal isolates (Holdeman et al. 1976) and 62% of eubacterial fecal rDNA (Wood et al. 1998) and is found in both humans and cows (Macy and Probst 1979). Franks and colleagues (Franks et al. 1998) found that the Bacteroides population from one human fluctuated less over time than populations of Bifidobacterium. Moreover, Bacteroides cells have been isolated from environmental
water samples for at least several days after their dispersal in water (Straub and Dixon 1997; Avelar et al. 1998; Kreader 1998). Survival of *Bacteroides* is dependent primarily on temperature and predation (Kreader 1998) and it can survive up to 6 days under oxygen stress (Avelar et al. 1998).

Thus, the *Bacteroides-Prevotella* group is a promising indicator for source identification of fecal contamination in water samples. We have identified 1 human-specific and 2 cow-specific gene clusters of fecal markers from *Bacteroides-Prevotella*, and demonstrated that these markers can be recovered from natural freshwater and saltwater samples. We have also identified these marker genes phylogenetically as members of the *Bacteroides-Prevotella* group, but representing uncharacterized species. Armed with these sequence data, we are currently designing new primers specific for each cluster of genetic markers. These primers will then identify the most likely sources of fecal contamination in natural water samples. Additionally, with the introduction of real-time quantitative PCR methods, we have the potential to develop a quantitative assay.

Our study has focused on two host species, and thus, we can only identify the absence of a particular pollution source with certainty. We have not yet investigated the distribution of the genetic markers in other animals. It is possible that some of these markers may not be limited to humans or cows. Future studies will test feces from other potential pollution sources such as swine, waterfowl, and other common wildlife.

We have demonstrated that LH-PCR and T-RFLP can be used to identify and track bacterial markers in complex natural environments. These methods have the advantage
of being specific, rapid, and sensitive to changes as subtle as one base pair (Liu et al. 1997; Clement et al. 1998). Potential applications of these methods include tracking environmentally important species, genetically engineered species released in the environment, and pathogens in clinical specimens.
Chapter 3

A PCR Assay to Discriminate Human and Ruminant Feces Based on Host Differences in Bacteroides-Prevotella 16S Ribosomal DNA

Anne E. Bernhard and Katharine G. Field

Submitted to Applied and Environmental Microbiology
American Society for Microbiology, Washington, D. C.
April 2000
3.1 Abstract

Our purpose was to develop a rapid, inexpensive method of diagnosing the source of fecal pollution in water. In previous research, we identified Bacteroides-Prevotella LH-PCR and T-RFLP ribosomal DNA markers that distinguished cow from human feces. Here, we recovered 16S rDNA clones from estuarine and river waters that were close phylogenetic relatives of the markers. From these sequence data, we designed specific PCR primers that discriminate between human and ruminant sources of fecal contamination.

3.2 Introduction

The inability to reliably identify the source of the contamination is partly to blame for the persistent problem of fecal pollution in coastal and inland waters. Although methods exist to quantify fecal pollution, none quickly and accurately identifies the animal source. Antibiotic resistance patterns of fecal streptococci (Wiggins 1996; Hagedorn et al. 1999; Wiggins et al. 1999) and E. coli ribosomal DNA tracking (Samadpour and Chechowitz 1995; Akre and Wilcox 1998) have recently emerged as potentially useful, but labor-intensive, solutions to the problem. Their reliability, however, may be considerably less than 100 percent (Wiggins 1996; Wiggins et al. 1999).

Unlike these methods, which require culturing indicator organisms, detection of host-specific molecular markers does not require culturing, and holds promise as a
reliable, rapid method for identifying sources of fecal contamination. The *Bacteroides-Prevotella* group is one of several non-coliform bacterial groups that has been proposed as an alternative fecal pollution indicator (Allsop and Stickler 1985; Fiksdal *et al.* 1985; Kreader 1995), partly because of its abundance in feces. The use of molecular methods makes it more feasible to use potentially difficult to grow anaerobic bacteria, such as *Bacteroides-Prevotella*, as indicators.

We recently identified host-specific *Bacteroides-Prevotella* 16S rDNA markers for humans and cows by screening fecal DNAs by LH-PCR (Length Heterogeneity PCR) (Suzuki *et al.* 1998) or T-RFLP (Terminal Restriction Fragment Length Polymorphism) (Liu *et al.* 1997) analysis (Bernhard and Field 2000). Cloning and sequencing experiments revealed that each marker comprised multiple sequences, forming host-specific gene clusters. Here, we have identified additional clones, recovered from water samples, that cluster with the fecal clones. Using the sequences from fecal and water clones, we developed primers specific for each cluster that can discriminate human and ruminant feces.

### 3.3 Materials and Methods

#### 3.3.1 Clones recovered from water samples

To identify fecal *Bacteroides-Prevotella* rDNA markers in water, we collected six 1-L water samples from areas in Tillamook Bay, Oregon that are frequently contaminated with fecal pollution. We processed the samples as previously described...
DNAs from each water sample were amplified with *Bacteroides-Prevotella* specific primers (Bac32F and Bac708R) as described (Bernhard and Field 2000). Equal portions of PCR products from all water samples were pooled, and cloned into pGEM T-Easy vectors according to the manufacturer’s directions (Promega).

To locate marker clones, we screened the clones by LH-PCR and T-RFLP analyses as described (Bernhard and Field 2000). Clones with host-specific patterns were sequenced on an ABI 377 DNA sequencer using dye terminator chemistry. All sequences were checked for chimeric structure with CHECK_CHIMERA of the RDP (Maidak *et al.* 1994) and by comparisons to other clones in our study. Similarities were calculated using the Distance function in GCG v.10 (Genetics Computer Group, Madison, Wisconsin) with the Kimura 2-parameter correction. We aligned these clones with the fecal clones from our previous study and inferred a phylogenetic tree with the neighbor-joining algorithm (Saitou and Nei 1987) in PHYLIP 3.5c (Felsenstein 1989). Details of the phylogenetic analysis are described in Bernhard and Field (Bernhard and Field 2000).

### 3.3.2 Primer design

To develop a PCR assay for identifying sources of fecal bacteria in water, we designed primers specific for each cluster and for clone HF10 (Table 3.1). We established specificity and optimal annealing temperatures for all primer pairs by using plasmid DNAs from target and closely related non-target sequences as well as
Bacteroides DNA from cultures. Additional confirmation of specificity was obtained through PROBE MATCH of the RDP. PCR reaction mixtures are described in Bernhard and Field (Bernhard and Field 2000). A thermal mini-cycler (MJ Research, Watertown, MA) was used for all reactions with the following conditions: 25 cycles of 94°C for 30s, appropriate annealing temperature (Table 3.1) for 30s, 72°C for 1 minute followed by a final 6 minute extension at 72°C. 1 µl of each PCR product was reamplified using the same conditions. PCR products were visualized in a 1% agarose gel stained with 1 µg/ml ethidium bromide.

Table 3.1 Primers used in this study. Bac = Bacteroides-Prevotella, HF = human specific; CF = cow specific. Numbers correspond to the numbers of the E. coli 16S rRNA gene. All forward primers (except HF134F) were paired with Bac708R. Annealing temperatures were empirically determined for each primer pair as described in the text. HF134F was paired with HF654R.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5'→3’)</th>
<th>Target</th>
<th>Anneal. Temp. (°C)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bac32F</td>
<td>AACGCTAGCTACAGGC TT</td>
<td>Bacteroides-Prevotella</td>
<td>53</td>
<td>(Bernhard and Field 2000)</td>
</tr>
<tr>
<td>Bac708R</td>
<td>CAATCGGAGTCTTCTCGTG</td>
<td>Bacteroides-Prevotella</td>
<td>-</td>
<td>(Bernhard and Field 2000)</td>
</tr>
<tr>
<td>CF128F</td>
<td>CCAACYTCCCCGWTACTC</td>
<td>CF123 cluster</td>
<td>58</td>
<td>This study</td>
</tr>
<tr>
<td>CF193F</td>
<td>TATGAAAGCTCCGGCC</td>
<td>CF151 cluster</td>
<td>55</td>
<td>This study</td>
</tr>
<tr>
<td>HF134F</td>
<td>GCCGTCTACTCTTGCC</td>
<td>HF10</td>
<td>61</td>
<td>This study</td>
</tr>
<tr>
<td>HF183F</td>
<td>ATCATGAGTTCACATGTCGG</td>
<td>HF8 cluster, HF74</td>
<td>59</td>
<td>This study</td>
</tr>
<tr>
<td>HF654R</td>
<td>CCTGCCTCTACTGTACTC</td>
<td>HF10</td>
<td>61</td>
<td>This study</td>
</tr>
</tbody>
</table>
Host-specific primers were further tested by amplifying fecal DNAs from target and non-target hosts. DNAs from human and cow feces and sewage were collected and processed according to methods described elsewhere (Bernhard and Field 2000). Samples from all other animals were collected with sterile utensils and placed in sterile 50 ml tubes or plastic bags, kept on ice for transport to the lab, and immediately stored at -80°C. Fecal DNAs were extracted using the Fast DNA Kit for Soil (Bio101, Vista, Calif.) following the manufacturer's directions. Samples were tested for marker genes by PCR.

3.3.3 PCR sensitivity

Sensitivity of the PCRs was evaluated by amplifying marker genes from serial dilutions of plasmid DNAs from clones CF123, CF68, and HF145. We also tested the sensitivity using serial dilutions of cow feces or raw sewage. Sensitivity assays were as described (Bernhard and Field 2000). DNAs from each dilution were tested for the markers by PCR. We measured fecal coliforms in each dilution according to standard methods (American Public Health Association 1992).

3.4 Results

Sequence analysis of clones recovered from water samples revealed 7 unique clones that corresponded to human or cow genetic markers previously identified. All of the clones were very similar, but not identical, to clones recovered from human and cow
fecal samples (Figure 3.1). To confirm that the clones recovered from water samples were fecal in origin, we designed primers specific to two of the water clones, TB141 and TB147, and amplified 16S rRNA genes from cow fecal DNAs. Sequence analysis of the PCR products confirmed that the sequences were the same as the sequences of the two clones.

Six of the seven clones recovered from water samples clustered with human- or cow-specific sequences identified in our earlier study (Figure 3.1). TB13 corresponded to the human-specific cluster, HF8, and was greater than 99% similar to other clones in this cluster. The TB13 sequence differed by only 1-2 bases from HF8, HF117, and HF145. The remaining clones corresponded to the cow-specific markers. TB141 had the same T-RFLP pattern as CF46, CF68, and CF151 and was 84.7 to 90.4% similar to the other CF151 clones. TB101, TB106, TB135 and TB146 had the same T-RFLP pattern as the other clones in the CF123 cluster and were 93.3 to 96.1% similar. The T-RFLP pattern of TB147 matched the patterns of the CF123 cluster, but the sequence grouped with the CF151 cluster (Figure 3.1). Additionally, TB147 had the highest similarity with CF17 (88.2%), which is in the CF123 cluster. Bootstrap values for the CF151 cluster dropped considerably when TB147 was included in the analysis.

Using PCR primers specific for each gene cluster, we confirmed host-specificity by testing fecal DNAs from human and cow feces and sewage samples. We detected genes corresponding to the HF8 cluster in 11 out of 13 human fecal samples, all of the sewage samples, and none of the cow fecal samples (Table 3.2). Using the HF10 primers, we detected PCR product in less than half of the sewage and human fecal samples and
Figure 3.1 Phylogenetic relationships among partial 16S rDNA sequences (558 positions) of clones recovered from Tillamook Bay water samples (TB). HF and CF are host-specific genetic markers identified from human or cow fecal clone libraries, respectively. The tree was inferred by neighbor-joining. Numbers above the internal branches are percentages of bootstrap replicates that support the branching order. Bootstrap values below 50% are not shown. Bootstrap values for a and b dropped from 68 to 47 and 76 to 40, respectively, when TB147 was added to the analysis. *Cytophaga fermentans* was used to root the tree.
detected product in one cow fecal sample. Because HF8 genes were more widely
distributed among the humans, and primers for HF10 were not as specific as desired, we
tested only for HF8 genes in subsequent analyses. Genes from the CF151 and CF123
clusters were detected in all cow samples, but in none of the human or sewage samples.

Table 3.2 Distributions of host-specific genetic markers in feces from the targeted hosts.
Numbers of positive PCR results (2 rounds of 25 cycles each) out of N samples tested are
reported.

<table>
<thead>
<tr>
<th>Target</th>
<th>N</th>
<th>HF8 Cluster</th>
<th>HF10 Cluster</th>
<th>Cow Markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>13</td>
<td>11</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Sewage</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Cow</td>
<td>19</td>
<td>0</td>
<td>1</td>
<td>19</td>
</tr>
</tbody>
</table>

To further determine the host-specificity of these primers, we tested fecal samples
collected from other animals in Tillamook and Corvallis (Table 3.3). HF8 sequences
were not detected in any samples from non-human sources. CF123 and CF151
sequences, however, were detected in all other ruminants tested.

Sensitivity of the PCR and detection of the fecal markers was tested using plasmid
DNAs and serial dilutions of feces or sewage. PCR sensitivity was approximately 1 x
$10^{-12}$ g DNA ($10^5$ gene copies) for all three plasmid DNAs. Detection of Bacteroides-
Prevotella DNA was 2 to 4 times greater than fecal coliform detection (Table 3.4).
Detection of CF123 genes was as sensitive as detection of fecal coliforms. Fecal
coliform detection, however, was one to two times more sensitive than detection of
CF151 genes. The sensitivity assay using cow fecal dilutions was repeated with feces from different cows, and similar results were obtained (Table 3.4).

**Table 3.3** Distribution of host-specific genetic markers in feces from non-target animals. Numbers of positive PCR results (2 rounds of 25 cycles each) out of N samples tested are reported. Animals marked with an "*" are ruminants.

<table>
<thead>
<tr>
<th>Animal</th>
<th>N</th>
<th>Human Marker</th>
<th>Cow Markers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>HF8 Cluster</td>
<td>CF123 Cluster</td>
</tr>
<tr>
<td>Cat</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Deer*</td>
<td>3</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Dog</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Duck</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Elk*</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Goat*</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Llama*</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Pig</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Seagull</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sheep*</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

**Table 3.4** Detection limits for Bacteroides-Prevotella 16S rDNA, host-specific genetic markers, and fecal coliforms. Results are from dilution assays using either cow feces or raw sewage. Cow feces A and B are combinations of feces from different cows (4 cows each). Numbers are reported as g dry feces/L. Bacteroides-Prevotella DNA was detected by PCR as described (Bernhard and Field 2000). ND = not determined.

<table>
<thead>
<tr>
<th>Source of DNA</th>
<th>Bacteroides-Prevotella</th>
<th>HF8 Cluster</th>
<th>CF123 Cluster</th>
<th>CF151 Cluster</th>
<th>Fecal Coliforms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cow feces A</td>
<td>2.8 x 10^-9</td>
<td>ND</td>
<td>2.8 x 10^-7</td>
<td>2.8 x 10^-5</td>
<td>2.8 x 10^-7</td>
</tr>
<tr>
<td>Cow feces B</td>
<td>3.6 x 10^-10</td>
<td>ND</td>
<td>3.6 x 10^-6</td>
<td>3.6 x 10^-5</td>
<td>3.6 x 10^-6</td>
</tr>
<tr>
<td>Sewage</td>
<td>1.4 x 10^-10</td>
<td>1.4 x 10^-6</td>
<td>ND</td>
<td>ND</td>
<td>1.4 x 10^-7</td>
</tr>
</tbody>
</table>
3.5 Discussion

We recovered seven unique clones from natural waters that corresponded to the human and cow fecal markers previously identified (Bernhard and Field 2000). These findings support the hypothesis that host-specific genetic markers can be used to identify nonpoint sources of fecal pollution in coastal waters. Six of the clones were close phylogenetic relatives of the clones recovered from feces. The seventh clone, however, was more distantly related and its phylogenetic placement was not strongly supported by bootstrap analysis (Figure 3.1). All the evidence, except the phylogeny, supports the inclusion of this clone in the CF123 cluster. Our inability to conclusively identify the phylogenetic relationship of this clone is likely due to insufficient sequence data. Our analysis included about one-third of the 16S rRNA gene, and analyses based on incomplete gene sequences may be prone to misclassifications (Young and Haukka 1996; Ritchie and Myrold 1999).

Our initial analysis of the fecal markers suggested that they were human and cow-specific. However, after testing feces from other animals, we discovered that the markers derived from cow feces were actually ruminant-specific. This is not surprising, as ruminants have a unique gastrointestinal system, and would likely share many ruminal microbial flora (Stahl et al. 1988; Amann et al. 1990). Therefore, a positive PCR result for genes belonging to the CF123 or CF151 cluster does not rule out wildlife sources, but land use evaluation of the area being tested could determine the likelihood of an agricultural or wildlife source. Additional experiments are currently underway in
our laboratory to identify genetic markers that discriminate between domestic and wild ruminants.

Determining the sensitivity of detecting the genetic markers proved somewhat inconclusive. We are not currently able to measure the abundance of the marker genes in a fecal sample, thus we cannot control how individual variability may affect the results. Experiments using feces from different cows gave similar, but not identical results, when standardized to g dry feces. Although the results varied slightly, we believe that these differences are not significant. Because of unknown variability of the marker species among individuals, the limits of detection can only be approximated.

Although extensive field testing is required to determine the efficacy of the assays and the geographic distribution of the host-specific markers before these markers can be used for routine water quality monitoring, we believe these PCR assays provide a promising diagnostic for identifying nonpoint sources of fecal pollution. Additionally, our approach for the identification of diagnostic markers can be easily applied to find markers for other animals besides humans and ruminants.
Chapter 4

Application of a New PCR Assay to Discriminate between Human and Ruminant Fecal Pollution in Tillamook Bay, Oregon

Anne E. Bernhard, Thierry Goyard, and Katharine G. Field
4.1 ABSTRACT

We demonstrate the use of a new PCR assay to detect and differentiate human and ruminant sources of fecal pollution in natural water samples. We tested samples collected from Tillamook Bay, Oregon, which has a long history of fecal pollution levels that exceed acceptable standards. The most likely sources are from dairy operations and ineffective sewage treatment. Using a suite of three PCR primer pairs specific for human and ruminant markers, we detected at least one marker in 27 of 55 samples. In most cases, the samples that were positive for a particular marker were within a few kilometers of the nearest potential source, as identified by NPDES (National Pollution Discharge Elimination System) or CAFO (Confined Animal Feeding Operation) permits. Most of the samples that were positive were located on two of the five major rivers that drain the Tillamook basin. These two rivers have the majority of on-site sewage systems, and high concentrations of dairy farms. This combination of host-specific genetic markers holds promise for identifying nonpoint source fecal pollution in coastal waters.

4.2 INTRODUCTION

Water quality in coastal bays and rivers is under constant threat of fecal contamination from human and animal sources. This pollution can pose significant health hazards, affecting recreation, fisheries, and aquatic life. The continuing problem is exacerbated by an inability to identify the source of contamination. Potential sources
may include municipal wastewater treatment plants, onsite septic systems, recreational boaters, runoff from livestock waste, and wildlife. Current methods used to measure fecal pollution levels, such as fecal coliform detection methods, do not discriminate between different source species. If the source could be accurately diagnosed, resources could be more appropriately allocated to reduce the pollution.

We recently developed a PCR assay to discriminate between human and ruminant fecal pollution (Bernhard and Field 2000; Bernhard and Field in review). The assay identifies the presence of host-specific 16S ribosomal DNA markers from fecal Bacteroides and Prevotella species. The PCR assay discriminates between human and ruminant sources of fecal contamination in water samples without the need for culturing indicator organisms, thus circumventing potential culture biases (Amann et al. 1995; Giovannoni et al. 1995). Other studies have shown that the use of molecular diagnostics for detection of pathogens and indicators in water can be accurate, sensitive and time-saving (Bej et al. 1990; Tsai et al. 1993; Johnson et al. 1995; Lleo et al. 1999).

The purpose of this study was to assess our ability to detect these host-specific fecal Bacteroides markers in water samples collected from Tillamook Bay, Oregon and its tributaries, and to relate the results to land use patterns. Fecal pollution in Tillamook Bay has been a continuing problem, causing severe economic losses in the shellfish industry, threatening aquatic habitat, and leading to closure of the bay for both recreation and fisheries (Crane and Moore 1986). In many years the bay has been closed more than 100 days (Department of Environmental Quality 1994). Farm animal waste from dairy operations, sewage treatment plants, and septic systems are all
potential major contributors to this pollution (Plummer 1995; Strittholt et al. 1998). By identifying sources of contamination in water samples collected over a large area of the basin, potential problem areas can be located and management strategies can be developed to reduce or eliminate the sources.

4.3 MATERIALS AND METHODS

4.3.1 Sampling site

We collected water samples from Tillamook Bay, Oregon and its five major tributaries, from June 1998 to April 1999. Tillamook Bay is a shallow estuary on the northwest coast of Oregon (Figure 4.1). It is approximately 3.2 km wide and 11.3 km long, with an estimated 3,590 hectares of surface water at high tide. Its watershed covers nearly 150,000 hectares, and is drained by five major rivers, the Miami, the Kilchis, the Tillamook, the Trask, and the Wilson Rivers. Ninety-one percent of the watershed is forested; about 8% is in farmland, with an estimated 22,000 dairy cattle (Dorsey-Kramer 1995). The bay is also the site of commercial and recreational fisheries, including one of Oregon’s primary oyster growing areas.
Figure 4.1 Sampling sites in Tillamook Bay and its tributaries, and locations of NPDES (National Pollution Discharge Elimination System) or CAFO (Confined Animal Feeding Operation) permitted sites.
Figure 4.1
4.3.2 Sample collection

Surface water samples (1 liter) were collected using a Niskin bottle and stored on ice during transport to the lab. We filtered the samples through 0.2 μm Supor filters (Gelman, Ann Arbor, MI), and stored the filters in lysis buffer (20 mM EDTA, 400 mM NaCl, 750 mM sucrose, 50 mM Tris, pH 9) at -80°C.

4.3.3 DNA extractions

DNA was extracted from water samples according to the methods of Giovannoni and others (Giovannoni et al. 1990), omitting the cesium trifluoroacetate centrifugation step. Briefly, SDS and proteinase K were added to the frozen filters and the filters were incubated at 37°C (30 min) and 55°C (30 min). Cell debris was removed by a phenol/chloroform extraction and crude extracts were purified by ethanol precipitation. We quantified the DNA in a 1% agarose gel by comparing the band intensity to a high molecular weight DNA mass ladder (Gibco BRL).

4.3.4 PCR

Each water sample was tested for the presence of the host-specific markers using host-specific primers (Table 4.1). A forward primer specific for each marker (Bernhard and Field in review) was paired with the general Bacteroides-Prevotella reverse primer, Bac708R (Bernhard and Field 2000). Approximately 2 ng DNA from each water sample was amplified in a mini thermal cycler (MJ Research) under the following
conditions: 25 cycles of 94°C for 30s, annealing temperature for 30s, 72°C for 1 minute, and a final extension at 72°C for 6 minutes. Each 50 μl PCR contained 1X Taq polymerase buffer, 10 μM each primer, 200 μM each dNTP, 1.25 units of Taq polymerase, 640 ng/μl BSA (Kreader 1996), and 1.5 mM MgCl₂. A second round of PCR was done using 1 μl of the PCR product from the first 25 cycles under the same conditions. Negative controls with no DNA were included in each experiment. PCR products were visualized in a 1% agarose gel with 1 μg/ml ethidium bromide. Samples were scored as positives if there was a visible band that was the same size as the appropriate positive control.

Table 4.1 Primers used in this study. Bac = Bacteroides-Prevotella; HF = human-specific; CF = ruminant specific. Numbers correspond to the position within the E. coli 16S rRNA gene. All forward primers were paired with Bac708R.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’-3’)</th>
<th>Target</th>
<th>Anneal. Temp. (°C)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bac708R</td>
<td>CAATCGGAGTTCTTCGTG</td>
<td>Bacteroides-Prevotella</td>
<td>-</td>
<td>(Bernhard and Field 2000)</td>
</tr>
<tr>
<td>CF128F</td>
<td>CCAACYTTCCCGWTACTC</td>
<td>Ruminant-specific fecal 16S rDNAs</td>
<td>58</td>
<td>(Bernhard and Field in review)</td>
</tr>
<tr>
<td>CF193F</td>
<td>TATGAAAGCTCCGGCC</td>
<td>Ruminant-specific fecal 16S rDNAs</td>
<td>55</td>
<td>(Bernhard and Field in review)</td>
</tr>
<tr>
<td>HF183F</td>
<td>ATCATGAGTTCACATGTCCG</td>
<td>Human-specific fecal 16S rDNAs</td>
<td>59</td>
<td>(Bernhard and Field in review)</td>
</tr>
</tbody>
</table>
4.4. RESULTS

Using a sensitive and specific PCR assay, we tested for the presence of human- and ruminant-specific genetic markers in fifty-five water samples collected from Tillamook Bay and its tributaries on five different sampling dates. On some dates, due to inclement weather or time constraints, we could not collect samples from all sites. This lack of data for all sites on some dates made comparisons difficult. We detected at least one host-specific fecal marker in 27 of the 55 water samples (Table 4.2). Sixteen samples were positive for the human-specific marker HF8. For the ruminant-specific markers, eleven were positive for CF123, and twenty were positive for CF151. CF123 was not detected in any bay samples; HF8 and CF151 were found in both bay and river samples.

There was no significant correlation between sites positive for a marker and the distance to the nearest permitted source. Distances from sample sites to the nearest NPDES (National Pollution Discharge Elimination System) site ranged from 0.2 to 8.7, and 0.2 to 11.9 km to the nearest CAFO (Confined Animal Feeding Operation) site (Table 4.3). Fecal markers were not detected at sites furthest away from a permitted site, but the distance varied for each marker. HF8 was never detected in samples more than 4.3 km from the nearest NPDES permitted site. Of the 18 samples collected from sites downstream of the nearest NPDES site, only five were positive for HF8. CF123 was never detected in samples more than 2.7 from the nearest CAFO permitted site, and was found in only three of the samples downstream of the nearest CAFO site. CF151
Table 4.2 Detection of human- (HF8) and ruminant- (CF123 and CF151) specific markers in water samples from Tillamook Bay and its tributaries. Site numbers correspond to those in Figure 1. B = bay sites, K = Kilchis River sites, M = Miami River sites, R = Trask River sites, T = Tillamook River sites. "+" indicates a positive PCR result, "-" indicates a negative PCR result. Not all sites were sampled on all sampling dates.

<table>
<thead>
<tr>
<th>Sampling Date</th>
<th>Site</th>
<th>HF8</th>
<th>CF123</th>
<th>CF151</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-24-98</td>
<td>B1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6-24-98</td>
<td>B4</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6-24-98</td>
<td>B5</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6-24-98</td>
<td>B9</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10-17-98</td>
<td>B1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10-17-98</td>
<td>B2</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>10-17-98</td>
<td>B3</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>10-17-98</td>
<td>B4</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>10-17-98</td>
<td>B5</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10-17-98</td>
<td>B7</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10-17-98</td>
<td>B8</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10-17-98</td>
<td>R5</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10-17-98</td>
<td>T4</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10-17-98</td>
<td>W4</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11-29-98</td>
<td>B1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11-29-98</td>
<td>B3</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11-29-98</td>
<td>B4</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11-29-98</td>
<td>K1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11-29-98</td>
<td>K2</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11-29-98</td>
<td>M1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11-29-98</td>
<td>R1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11-29-98</td>
<td>R5</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11-29-98</td>
<td>T4</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11-29-98</td>
<td>W1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 4.2 Continued.

<table>
<thead>
<tr>
<th>Sampling Date</th>
<th>Site</th>
<th>HF8</th>
<th>CF123</th>
<th>CF151</th>
</tr>
</thead>
<tbody>
<tr>
<td>11-29-98</td>
<td>W4</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2-27-99</td>
<td>R1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2-27-99</td>
<td>R3</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2-27-99</td>
<td>R4</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2-27-99</td>
<td>T2</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2-27-99</td>
<td>T3</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2-27-99</td>
<td>W1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2-27-99</td>
<td>W2</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2-27-99</td>
<td>W3</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4-2-99</td>
<td>B1</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>4-2-99</td>
<td>B2</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>4-2-99</td>
<td>B3</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>4-2-99</td>
<td>B4</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>4-2-99</td>
<td>B5</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>4-2-99</td>
<td>B6</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>4-2-99</td>
<td>B7</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>4-2-99</td>
<td>B8</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4-2-99</td>
<td>B9</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4-2-99</td>
<td>R1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4-2-99</td>
<td>R3</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>4-2-99</td>
<td>R4</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4-2-99</td>
<td>R5</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>4-2-99</td>
<td>R6</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>4-2-99</td>
<td>R7</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4-2-99</td>
<td>T1</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>4-2-99</td>
<td>T2</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>4-2-99</td>
<td>T3</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>4-2-99</td>
<td>T4</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>4-2-99</td>
<td>W1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4-2-99</td>
<td>W3</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4-2-99</td>
<td>W4</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 4.3 Distances (km) from each sampling site to the nearest permitted pollution source. NPDES = National Pollution Discharge Elimination System; CAFO = Confined Animal Feeding Operation. Numbers in parentheses indicate that the permitted sites are upstream from the sampling site.

<table>
<thead>
<tr>
<th>Sampling Site</th>
<th>Nearest NPDES Permit</th>
<th>Nearest CAFO Permit</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>(3.9)</td>
<td>(2.3)</td>
</tr>
<tr>
<td>B2</td>
<td>4.3</td>
<td>(2.7)</td>
</tr>
<tr>
<td>B3</td>
<td>3.3</td>
<td>(2.8)</td>
</tr>
<tr>
<td>B4</td>
<td>2.5</td>
<td>(3.1)</td>
</tr>
<tr>
<td>B5</td>
<td>(1.3)</td>
<td>(3.6)</td>
</tr>
<tr>
<td>B6</td>
<td>(1.8)</td>
<td>(4.0)</td>
</tr>
<tr>
<td>B7</td>
<td>1.5</td>
<td>(5.9)</td>
</tr>
<tr>
<td>B8</td>
<td>0.6</td>
<td>(6.7)</td>
</tr>
<tr>
<td>B9</td>
<td>(1.2)</td>
<td>(8.2)</td>
</tr>
<tr>
<td>K1</td>
<td>8.7</td>
<td>5.1</td>
</tr>
<tr>
<td>K2</td>
<td>(1.9)</td>
<td>0.9</td>
</tr>
<tr>
<td>M1</td>
<td>7.5</td>
<td>11.9</td>
</tr>
<tr>
<td>R1</td>
<td>2.8</td>
<td>0.8</td>
</tr>
<tr>
<td>R2</td>
<td>0.2</td>
<td>(0.4)</td>
</tr>
<tr>
<td>R3</td>
<td>(0.2)</td>
<td>(0.5)</td>
</tr>
<tr>
<td>R4</td>
<td>(2.0)</td>
<td>(0.9)</td>
</tr>
<tr>
<td>R5</td>
<td>1.2</td>
<td>(0.6)</td>
</tr>
<tr>
<td>R6</td>
<td>1.0</td>
<td>(1.9)</td>
</tr>
<tr>
<td>T1</td>
<td>1.6</td>
<td>(0.2)</td>
</tr>
<tr>
<td>T2</td>
<td>4.8</td>
<td>(0.6)</td>
</tr>
<tr>
<td>T3</td>
<td>2.1</td>
<td>0.7</td>
</tr>
<tr>
<td>T4</td>
<td>2.0</td>
<td>(1.8)</td>
</tr>
<tr>
<td>W1</td>
<td>8.6</td>
<td>1.9</td>
</tr>
<tr>
<td>W2</td>
<td>3.1</td>
<td>(0.4)</td>
</tr>
<tr>
<td>W3</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td>W4</td>
<td>(4.1)</td>
<td>(2.1)</td>
</tr>
</tbody>
</table>
was not detected in any samples over 5.9 km from the nearest CAFO site, and was found in five of the samples collected downstream of the nearest CAFO site.

Overall, the sites that were most frequently positive for human or ruminant markers were located in the bay, and on the Tillamook and Trask Rivers. No markers were detected in any samples from the Kilchis or Miami Rivers, and only one sample on the Wilson River, W4 on April 2, 1999, tested positive for the HF8 marker. No sites on the Wilson were positive for either of the ruminant-specific markers. Of all the sites in the bay and on the Tillamook and Trask Rivers, only two, R1 and B9, did not test positive for any of the markers on any sampling date.

Precipitation during sampling varied from 0 to almost 3 inches on the day of sampling (Table 4.4). Total precipitation during the five days prior to the sampling date ranged from 0.8 to over 10 inches. Rainfall was heaviest during the February sampling.

<table>
<thead>
<tr>
<th>Sampling Date</th>
<th>Precipitation Data (in inches)</th>
<th>Five Days Prior*</th>
</tr>
</thead>
<tbody>
<tr>
<td>June 24, 1998</td>
<td>0.77</td>
<td>0.8</td>
</tr>
<tr>
<td>October 17, 1998</td>
<td>0.19</td>
<td>2.19</td>
</tr>
<tr>
<td>November 29, 1998</td>
<td>0.77</td>
<td>7.43</td>
</tr>
<tr>
<td>February 27, 1999</td>
<td>2.87</td>
<td>10.59</td>
</tr>
<tr>
<td>April 2, 1999</td>
<td>0.0</td>
<td>2.14</td>
</tr>
</tbody>
</table>

* includes the day of sampling and the five days prior to sampling
4.5 DISCUSSION

We successfully applied a sensitive and specific PCR diagnostic test to identify human- and ruminant-specific fecal contamination in river and estuarine water samples. Our results indicate that the PCR assay is a rapid method for identifying nonpoint source pollution in coastal waters.

We detected fecal markers in at least one sample on all sampling dates, except during November. Precipitation on the sampling date during November and the five days prior was quite high, averaging more than an inch per day. Heavier rains earlier in the week may have washed significant portions of fecal bacteria from the fields, thus leaving low numbers on the day of sampling. Additionally, the large amount of rain may have created a dilution effect, so that the concentrations of the markers were below the detection limits. Others have observed a decrease in indicator bacterial concentrations with increased rainfall (Patni et al. 1985).

We did not observe the same pattern in February, however, when rains were even higher. This may be attributed to the intensity of the storm. The rain on the day of sampling in February was much greater than in November (2.87 vs 0.77 inches, respectively), and the winds were quite high. The weather was severe enough to prevent sampling in the bay. The samples collected that day were sediment-laden and even the DNA extracts were brown, indicating high concentrations of contaminating substances. Other studies have documented increased fecal bacterial concentrations during storm events, due to resuspended sediments and increased survival of bacteria associated with sediments (Sherer et al. 1992; Davies et al. 1995; Baudart et al. 2000).
The relationship between bacterial loads and rainfall, however, is not simple, and is affected by a complex set of interacting factors (see Baxter-Potter and Gilliland 1988 for review). In addition to the intensity and amount of rainfall, other factors include slope of the land, sunlight, temperature, vegetation cover, soil properties, distance from the source to the sampling location, presence of toxic substances, and manure management practices. The age of the manure deposits may also be a significant factor.

Kress and Gifford (Kress and Gifford 1984) studied the effect of varying intensities of rainfall on standard cowpies and found that the effect of rain intensity was related to dryness (or age) of the cowpie. They also found that release of fecal bacteria from 100-day old deposits was minimal compared to fresher fecal material. Because of the complex factors involved, we cannot say with certainty how these factors may have affected our results.

There was a weak relationship between the distance from the nearest permitted fecal pollution source (NPDES and CAFO permits) to the sampling site and presence of the markers, but again the relationship was not simple. Samples collected furthest from the permitted sites were always negative for the fecal markers, but the distances varied for the three markers. The sites that tested positive for the human-specific marker, HF8, could be attributed to sewage outfalls or sewage treatment plant effluents, but septic systems may also be contributing in some cases. There are over 1700 on-site sewage systems in the Tillamook Basin, with the majority concentrated on the Tillamook, the Trask, and the Wilson Rivers (Newell 1998). Surveys of these on-site sewage systems between 1988 and 1996 revealed a failure rate of 6 to 7% (Newell
Because samples were collected on a high tide, we would predict sampling sites located upstream of the source to have higher levels of fecal pollution than sites located downstream of the source. With a few exceptions, the detection of fecal markers was consistent with this prediction.

One of the ruminant markers, CF123, was found only in freshwater locations and river mouths, where salinity was less than 1 ppt. The sensitivity of detection of this marker under laboratory conditions is greater than the sensitivity of detection of the other ruminant marker, CF151 (Bernhard and Field, in review). Salinity effects, however, were not investigated. We detected CF151 in nine samples that were negative for CF123. The majority of these samples were located in the bay, which is affected by saline ocean waters. Our samples were collected during high tides, when the influence of saltwater would be greatest. It is possible that the *Bacteroides* species or strains that comprise CF123 die off more rapidly upon exposure to saline waters, as has been observed with other fecal bacteria (Sinton *et al.* 1994; Mezrioui *et al.* 1995). Studies of the effect of salinity on freshwater bacteria have shown that many riverine bacteria do not survive even in oligohaline conditions (Valdes and Albright 1981; Bordalo 1993; Painchaud *et al.* 1995). Although it is possible that not all fecal pollution from ruminants may carry both markers, previous studies found both markers in all ruminant fecal samples tested (Bernhard and Field, in review).

CF123 and CF151 detect ruminant-specific fecal pollution, so it is possible that deer and elk feces might have contributed to the signal. However, the land use patterns suggest that these sources would be insignificant in most cases. Most of our sampling
sites were located in rural areas with a high density of agricultural operations. Additional evidence for the lack of wildlife contribution is that in sites upriver, such as R1 and W1, where wildlife would be most concentrated, no fecal markers were ever detected. Others have found that in pristine areas, the contribution of wildlife, such as deer and elk, may contribute to fecal pollution in waterways, but the amount is insignificant compared to fecal pollution in agricultural areas (Niemi and Niemi 1991). Despite the improbability of a significant wildlife contribution to the fecal pollution problem, we are currently developing additional host-specific genetic markers that discriminate between domestic and wildlife ruminants.

In summary, we have demonstrated the use of a new method to identify nonpoint sources of human and ruminant fecal contamination in coastal waters. Because the method detects nucleic acids directly from water samples, we circumvent the need for culturing and the problems associated with culturing microbes from natural environments (Manz et al. 1994). Additionally, this method is specific, sensitive, and very rapid. We believe that this method is a promising approach to solving the problem of identifying the source of nonpoint pollution in coastal waters.
Chapter 5

Summary

Anne E. Bernhard
We successfully applied molecular methods to identify and characterize host-specific genetic markers that can distinguish between human and ruminant sources of fecal contamination. LH-PCR and T-RFLP proved to be very powerful tools for identifying host-specific differences in human and cow feces. Other researchers have also found these methods to be very useful for comparing microbial community gene profiles (Bruce 1997; Liu et al. 1997; Clement et al. 1998; Rappé et al. 1998; Suzuki et al. 1998; Moeseneder et al. 1999; Flynn et al. 2000; Ritchie et al. 2000; Scala and Kerkhof 2000). The combination of these methods with gene cloning and sequencing allowed us to characterize the unique enteric Bacteroides and Prevotella species in humans and cows, based on their 16S rRNA gene sequences.

Results of phylogenetic analyses of these unique 16S rDNA sequences revealed host-specific gene clusters. The sequences recovered from cow feces were not related to any published sequences and formed two novel gene clusters. Since the application of molecular methods to diversity studies, the discovery of novel gene clusters has become quite common (see, for e.g., Ueda et al. 1995; Hales et al. 1996; Stephen et al. 1996; Field et al. 1997; Wright et al. 1997; Stephen et al. 1999; Suau et al. 1999). The abundance of such novel lineages in microbial ecology underscores the plethora of undescribed diversity in many habitats.

Using the sequences for the host-specific Bacteroides-Prevotella lineages, we designed oligonucleotide PCR primers for each host-specific gene cluster. Initially, we thought the sequences recovered from cow feces were specific to cows only, but after testing other animals, these sequences proved to be ruminant-specific. Considering that
ruminants have a unique digestive system, it is not surprising that they harbor similar microflora. Experiments are currently underway in our laboratory to find genetic markers that can discriminate domestic from wildlife ruminant sources.

After testing the primers for specificity and sensitivity, we applied our PCR assay to samples collected from Tillamook Bay. These tests were designed to evaluate the source-specific fecal markers for ease of detection and their relation to land use patterns. We demonstrated that the method is very quick and the patterns of the presence of the markers were loosely correlated with distances from the site to the nearest potential source, as indicated by NPDES and CAFO permitted sites.

We believe that this assay can be quite useful in identifying the sources of fecal contamination. However, its biggest limitation is that it is only qualitative, rather than quantitative. Our estimates of its sensitivity are only approximations because we cannot currently measure the abundance of the markers, and the individual variability is unknown. We are currently designing experiments, using real-time PCR technology, to try to address this concern. Additionally, we do not know the distribution of the markers in populations outside of the area tested.

Our approach to identify the source-specific markers and characterize them by sequence identity should be widely applicable to develop markers for other animals, such as swine, or waterfowl. This method can also be applied to compare microbial communities from polluted and unpolluted sites, to identify indicator species or markers for particular pollutants. The combination of LH-PCR and T-RFLP with sequence
identification is a powerful tool in microbial ecology and should prove very useful in a variety of habitats.
Bibliography


environmental nucleic acid samples from the Pacific and Atlantic coasts of the United States. *Applied and Environmental Microbiology* 64: 294-303.


