Assessment of fecal pollution sources in a small northern-plains watershed using PCR and phylogenetic analyses of Bacteroidetes 16S rRNA gene

Regina Lamendella1, Jorge W. Santo Domingo2, Daniel B. Oerther1, Jason R. Vogel3 & Donald M. Stoeckel4

1Department of Environmental Engineering, University of Cincinnati, Cincinnati, OH, USA; 2US Environmental Protection Agency, Office of Research and Development, National Risk Management Research Laboratory, Cincinnati, OH, USA; 3US Geological Survey, Lincoln, NE; and 4US Geological Survey, Columbus, OH, USA

Correspondence: Jorge W. Santo Domingo, US Environmental Protection Agency, Office of Research and Development, National Risk Management Research Laboratory, Martin Luther King Drive, Cincinnati OH. Tel.: +1 513 569 7085; fax: +1 513 569 7328; e-mail: santodomingo.jorge@epa.gov

Received 31 March 2006; revised 23 June 2006; accepted 11 July 2006. First published online 27 October 2006. DOI:10.1111/j.1574-6941.2006.00211.x

Editor: Michael Wagner

Keywords: microbial source tracking; Bacteroidetes; 16S rRNA gene; water quality; fecal pollution.

Abstract

We evaluated the efficacy, sensitivity, host-specificity, and spatial/temporal dynamics of human- and ruminant-specific 16S rRNA gene Bacteroidetes markers used to assess the sources of fecal pollution in a fecally impacted watershed. Phylogenetic analyses of 1271 fecal and environmental 16S rRNA gene clones were also performed to study the diversity of Bacteroidetes in this watershed. The host-specific assays indicated that ruminant feces were present in 28–54% of the water samples and in all sampling seasons, with increasing frequency in downstream sites. The human-targeted assays indicated that only 3–5% of the water samples were positive for human fecal signals, although a higher percentage of human-associated signals (19–24%) were detected in sediment samples. Phylogenetic analysis indicated that 57% of all water clones clustered with yet-to-be-cultured Bacteroidetes species associated with sequences obtained from ruminant feces, further supporting the prevalence of ruminant contamination in this watershed. However, since several clusters contained sequences from multiple sources, future studies need to consider the potential cosmopolitan nature of these bacterial populations when assessing fecal pollution sources using Bacteroidetes markers. Moreover, additional data is needed in order to understand the distribution of Bacteroidetes host-specific markers and their relationship to water quality regulatory standards.

Introduction

The microbiological impairment of water is traditionally assessed by monitoring concentrations of fecal indicator bacteria such as fecal coliforms and enterococci. These microorganisms are associated with fecal material from warm-blooded animals and their presence in water indicates the potential incidence of enteric pathogens that could cause illness in exposed individuals (Dufour, 1984). Screening for members of the Bacteroidetes phylum has advantages over screening for traditional indicators, including fecal coliforms and enterococci. For example, Bacteroidetes are more abundant in the feces of warm-blooded animals than Escherichia coli, constituting up to one half of fecal bacterial community (Holdeman et al., 1976; Fiksdal et al., 1985). Additionally, Bacteroidetes might be useful in predicting recent fecal contamination, as they are obligate anaerobes and are assumed to survive for only short periods of time outside the intestinal tract, while E. coli and enterococci are facultative anaerobes able to proliferate in soil, sand and sediments (Duerden, 1980; Allsop & Stickler, 1985; Franks et al., 1998; Suau et al., 1999; Sghir et al., 2000; Hopkins et al., 2001; Hold et al., 2002).

Current standard methods for measuring fecal pollution in water cannot be used to identify sources of pollution, as coliforms are common to all mammalian hosts (Gordon & Cowling, 2003). Recently, numerous phenotypic, genotypic and chemical based microbial source tracking (MST) methods have been developed with the intent of determining the source of fecal pollution in various environments (U.S. Environmental Protection Agency 2000; Scott et al., 2002; Simpson et al., 2002). In general terms, MST tools could be classified as library and nonlibrary-based methods, where library-based methods require the development of fingerprint databases of cultured bacteria, while library-independent methods do not rely on the development of large culture.
collections. However, limited research has been conducted in
testing the efficacy of these MST tools on a spatial and
temporal scale (U.S. Environmental Protection Agency 2005).

Several studies have suggested that fecal members of the
phylum Bacteroidetes are ecologically diverse, numerically
important members of the intestinal microbial communities
and exhibit some host specificity (Salyers, 1984; Gherna &
Woese, 1992; Wood et al., 1998; Boone & Castenholz, 2001;
Daly et al., 2001; Leser et al., 2001; Hold et al., 2002). Recently,
Bernhard & Field (2000a, b) developed PCR-based, 16S rRNA
gene assays targeting Bacteroidetes which were capable of
distinguishing between human and ruminant sources (Bern-
hard & Field, 2000a, b; Bernhard et al., 2003; Field et al.,
2003). While these assays are rapid, specific and relatively
sensitive, the spatial and temporal analysis and host-specifi-
city of these PCR-based Bacteroidetes assays have not been
widely tested for their applicability to different large-scale
graphic scenarios. Furthermore, phylogenetic analysis of
the gene sequences recovered from a diverse array of environ-
mental samples is needed to further clarify host distributions
of fecal Bacteroidetes bacteria in environmental systems.

Plum Creek, located in south-central Nebraska, contains
elevated concentrations of fecal coliform bacteria and is one
of the main tributaries contributing to the 303(d) listed,
impaired Platte River (http://www.deq.state.ne.us/Publica-
nsf/Pages/WAT021). Cattle are among the potential primary
sources of fecal pollution; however, this has not been con-
formed with any of the available source tracking methods.
Other animals (e.g. wildlife, horse and human) may also be
contributing to the pollution in this watershed. Cattle fecal
pollution is of concern, as some of the microorganisms shed
in cattle feces can be pathogenic to humans, such as E. coli
parvum and Giardia spp. (Cotruvo et al., 2004).

The goal of this study was to assess the applicability of
host-specific, Bacteroides-targeted, 16S-rRNA gene PCR
assays in identifying sources of fecal pollution in the Plum
Creek watershed by challenging total, ruminant and human-
targeted 16S rRNA gene Bacteroides markers against fecal,
water and sediment samples collected over a 6-month
period. The specificity of host-specific Bacteroides-targeted
16S rRNA gene assays was determined against target and
nontarget fecal samples, as well as by performing sequencing
analyses of clone libraries from fecal, water and sediment
samples.

Materials and methods

Study site: Plum Creek

Plum Creek drains 224 square miles of rural land and is a
perennial tributary to the middle of the Platte River, which
is located near the town of Smithfield, in south-central

Nebraska. Segment MP2-20000 of the Platte River was listed
on the 2004 Nebraska Section 303(d) list, impaired for
primary contact recreation due to excessive levels of fecal
coliform bacteria (http://www.deq.state.ne.us/Publica-
nsf/Pages/WAT021). Fecal coliform data from 2001 indicated
that Plum Creek was the most contaminated of the peren-
nial tributaries to this segment of the middle of the Platte
River (P. Obrien, USGS, pers. commun.). Grassland was the
most extensive land cover in the Plum Creek watershed,
with irrigated row crops, hay and small grains comprising
secondary land use. Six sampling sites were established along
Plum Creek. Site 1 was located near the headwaters of the
perennially flowing portion of Plum Creek and was used to
record background levels of pollution, due to the previously
observed lower levels of fecal indicator bacteria at this site.
Sites 2 through 6 were located intermittently downstream
on the main stem of Plum Creek, with Site 6 located in close
proximity to the confluence of Plum Creek with the Platte
River. An estimated 1200 cattle were located near the
sampled portion of Plum Creek, while humans (c. 15 septic
systems representing < 100 humans), horses (< 10) and
wildlife (e.g. deer, waterfowl, raccoon and opossum) are
among the other likely fecal sources possibly impacting this
watershed.

Sample collection

All fecal samples were collected aseptically during the last
week of July 2004 between Site 1 and Site 2 and transferred
into cryogenic-safe tubes, transported to the laboratory in
coolers and stored at −80 °C prior to analyses. A total of
101 individual cattle fecal samples (collected between Site 1
and Site 2), 66 wildlife fecal samples, six horse fecal samples
and 10 septic dips (from 10 different households represent-
ing c. 50 humans) were used to evaluate the specificity of the
assays. Wildlife fecal samples were collected in the riparian
area directly adjacent to Plum Creek and likely comprise
samples from raccoons and opossums. Water samples were
collected in duplicate, biweekly from May 2004 through
November 2004 at each site along the Plum Creek
(n = 26–28 per site) following previously published standard
operating procedures (U.S. Geological Survey 2000). Briefly,
water samples were collected in sterilized Nalgene bottles and
transported to the laboratory in ice coolers. Water samples
(i.e. 30–100 mL) were then filtered on to 0.45 μm polycarbo-
nate filters which were stored at − 80 °C until genomic DNA
extractions were performed. Water blanks were collected at
all 14 sampling dates as part of the quality assurance protocol
to determine the potential of cross-contamination due to
sample handling. Sediment bed cores were collected from
two different intervals (0–10 cm, 10–20 cm and 20–30 cm)
at Site 2 during the latter half of the sampling period (i.e.
Genomic DNA extraction

Prior to genomic DNA extractions, fecal slurries were made consisting of 0.10 g of fecal sample and 3.5 mL of magnesium phosphate rinse buffer (American Public Health Association, 1998). Fecal genomic DNA extractions were performed with the UltraClean Fecal DNA Isolation Kit according to the manufacturer’s instructions (MO BIO Laboratories, Inc., Carlsbad, CA) using 250 μL of each fecal slurry. For water samples, DNA was extracted directly from whole filters using UltraClean Soil DNA Isolation Kit (MO BIO Laboratories, Inc.). Sediment genomic DNA was also extracted using the Ulaclean Soil DNA Isolation Kit. Genomic DNA was eluted in 50 μL of 10 mM Tris and quantified using a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE). Fecal genomic DNA was diluted using ultrapure water to achieve a final concentration of 1.5 ng μL⁻¹. To determine the potential impact of PCR inhibitors in the environmental DNA extracts, PCR was performed with 10-fold dilutions (using ultrapure water) of genomic DNA extracted from the water and sediment samples. To test for possible cross-contamination, control blanks for genomic DNA extractions were performed using ultrapure water samples as templates. Extracted genomic DNA was stored at −20°C until further processing.

Bacteroides 16S rRNA gene PCR assays

PCR analyses were performed on all fecal, water and sediment samples using four different primers sets, targeting bacteria closely related to total Bacteroides and Prevotella (Bac32; herein referred as total Bacteroides-like) and using primers specific to ruminant Bacteroides (CF128 and CF193) or human Bacteroides (HF134 and HF 183). All reactions were performed using Bac708R as the reverse primer as previously described (Bernhard & Field, 2000b) with the following modifications. The final PCR solutions contained 2.5 μL of Takara Ex Taq 10X buffer (20 mM Mg²⁺), 2 μL of dNTP mixture (2.5 mM each), 0.4 μL of 4% BSA, 17 μL of UltraPure water (2.5 mM each), 0.4 μL of primer at 25 pmol μL⁻¹ concentration and 0.625 U of Ex Taq DNA polymerase (TAKARA Mirus Bio, Madison, WI). Reactions were conducted on a DNA Engine 2 Tetrad thermalcycler (Bio-Rad Laboratories, Inc., Hercules, CA). Template concentrations for fecal samples were adjusted to 0.25 ng μL⁻¹ and 1.5 ng μL⁻¹ for the total Bacteroides-like and the host-specific Bacteroides assays, respectively. The PCR conditions for the total Bacteroides-like PCR assay included a denaturing step at 94°C for 2 min, followed by 30 cycles (for fecal samples) or 35 cycles (for sediment and water samples) of 94°C for 1 min, 53°C for 1 min and 72°C for 1 min, followed by a final extension step at 72°C for 7 min. The PCR conditions for CF128 and CF193 ruminant Bacteroides assays included denaturing at 94°C for 2 min followed by 30 cycles (fecal samples) or 35 cycles (sediment and water samples) of 94°C for 1 min, 62°C for 1 min and 72°C for 1.5 min, followed by a final extension step of 72°C for 7 min. The PCR conditions for the human assays, HF134 and HF183, were similar to the aforementioned ruminant assays with the exception that the annealing step was performed at 63°C for 1 min and extension at 72°C for 1.5 min.

All PCR products were visualized using 1% agarose gels and GelSTAR nucleic acid stain (Cambrex BioScience, East Rutherford, NJ). Detection limit assays were conducted for the general Bacteroides-like, ruminant and human 16S rRNA gene targeted primer sets using composite samples from each potential fecal source type in order to determine the lowest concentration of detectable fecal pollution (Table 1). To determine assay sensitivity, known concentrations of fecal DNA composites were serially diluted using ultrapure water. Aliquots (1 μL) of the serial dilutions were then used as template in 35 PCR cycles for each primer set tested.

PCR purification, cloning and sequencing analyses

All PCR products were purified using the QIAquick PCR Purification Kit according to the manufacturer’s instructions for the microcentrifuge protocol (QIAGEN, Valencia, CA). Representative PCR products from total Bacteroides-like and ruminant CF128 Bacteroides assays were cloned into a pCRII-TOPO vector and transformed into Mach-T1 chemically competent E. coli cells as described by the manufacturer (Invitrogen, Carlsbad, CA). Clone libraries were developed from PCR products generated with genomic DNA from septic tanks, cattle, wildlife and horse fecal samples, water samples from all six Plum Creek sampling sites and sediment cores from Site 2. Individual E. coli clones were then subcultured into 300 μL of Luria Broth containing 50 μg mL⁻¹ ampicillin and screened for inserts using M13 PCR.

M13 PCR products were submitted to Children’s Hospital (Cincinnati, OH) for sequencing. Sequencing data was generated using Big Dye sequencing chemistry (Applied Biosystems, Foster City, California), M13 forward and reverse primers and an Applied Biosystems PRISM 3730XL DNA Analyzer. Sequences were manually cleaned for base calls and aligned, using SEQUENCER 4.5 software. Potential
were not included in further analyses. Sequences were also deposited in the GenBank database under accession numbers DQ886077-DQ886236.

Phylogenetic analysis used ARB software and trees were inferred from 527 sequence positions using two tree-building methods: neighbor-joining using a Kimura correction and maximum parsimony using the Phylip DNAPARS tool (Ludwig et al., 2004). In order to statistically evaluate the confidence of branching, bootstrap values were obtained from a consensus of 100 parsimony trees, as the ARB neighbor-joining bootstrap tool was not functional at the time of analysis. The 16S rRNA gene sequence of Porphyromonas gingivalis (accession # L16492) was considered as the outgroup for all constructed trees (Paster et al., 1994).

Diversity of the clone libraries was also investigated by rarefaction analysis (Hurlbert, 1971; Heck et al., 1975; Simberloff, 1978). Rarefaction curves were produced using the Hurlbert analytical approximation algorithm and 95% estimated confidence intervals. Calculations were performed with the freeware program aRarefactWin (Holland, 2003).

**Nucleotide sequence accession numbers**

Representative sequences generated in this study have been deposited in the GenBank database under accession numbers DQ886077-DQ886236.

**Results and discussion**

**Escherichia coli densities and 16S rRNA gene PCR assays**

Microbiological assessment of surface waters from Plum Creek indicated the geometric mean densities of *E. coli* ranged from 55 CFU 100 mL⁻¹ at Site 1, to 534 CFU 100 mL⁻¹ at Site 4 (Table 2).

Using cattle fecal DNA as the target, the detection limits of the general *Bacteroidetes* primer sets CF128 and CF193 assays were found to be 1 × 10⁻¹⁴, 1 × 10⁻¹² and 1 × 10⁻¹¹ g fecal genomic DNA, respectively (Table 1). The lower sensitivity of the ruminant assays is in agreement with the fact that these assays only target a fraction of the total *Bacteroides*-like bacteria (Bernhard & Field, 2000b). The ruminant CF128 assay also exhibited a higher amplification percentage for fecal (96%) and water samples (53%) than the CF193 primer (90% and 28%, respectively), which is indicative of the higher sensitivity of the CF128 primer and consistent with previous findings (Bernhard & Field, 2000b). Detection limits for the Bac32, HF134 and HF183 assays performed on septic samples were 1 × 10⁻¹², 1 × 10⁻⁹ and 1 × 10⁻¹⁰ g of fecal DNA, respectively. These results indicate that the HF183 marker is more sensitive than the HF134 marker, findings that are in agreements with Bernhard and Field (2000b). Detection limits with wildlife and horse fecal DNA extracts using the Bac32 assay were 1 × 10⁻¹³ and 1 × 10⁻¹⁴ g of fecal DNA, respectively.

All water-collection field blanks and fecal extraction blanks were negative using all PCR assays. None of the no-template controls (i.e. ~10% of the total number of reactions) produced amplification products, implying the lack of artifacts introduced by cross-contamination. All positive controls (~2% of the total number of reactions) produced the expected amplification products. More than 90% of the cattle fecal samples produced positive signals with the general *Bacteroidetes* and the ruminant-specific PCR assays (Table 2). Interestingly, only one-third of the wildlife fecal samples amplified using the general *Bacteroidetes* marker, suggesting that *Bacteroides*-like populations may not be as prominent in the intestinal microbiota of certain wildlife or that the tested primers cannot universally amplify all members of this fecal bacterial group. When ruminant specific assays were challenged against nontarget fecal samples, < 3% showed to cross amplify with wildlife and septic source samples, although it should be noted that all six horse fecal samples amplified with both ruminant-specific PCR assays. Cross amplification with the horse samples suggests that the ruminant assay is not capable of discriminating between horse and cattle pollution in this particular watershed. Similar to cattle, horses are large mammalian herbivores and it is therefore possible for them to host similar *Bacteroidales* populations. Both human-targeted *Bacteroides* PCR assays produced positive amplification signals with 40% of the septic samples. Cross

### Table 1. Limit of detection in fecal sample composites using general and host-specific *Bacteroidetes* primer sets

<table>
<thead>
<tr>
<th>Species</th>
<th>CF128 Bac708</th>
<th>CF193 Bac708</th>
<th>HF134 Bac708</th>
<th>HF 183 Bac708</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cattle (10)</td>
<td>1 × 10⁻¹⁴</td>
<td>1 × 10⁻¹²</td>
<td>1 × 10⁻¹¹</td>
<td>–</td>
</tr>
<tr>
<td>Septic (10)</td>
<td>1 × 10⁻¹²</td>
<td>–</td>
<td>1 × 10⁻⁹</td>
<td>1 × 10⁻¹⁰</td>
</tr>
<tr>
<td>Horse (6)</td>
<td>1 × 10⁻¹⁴</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Wildlife (10)</td>
<td>1 × 10⁻¹³</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

*Number of samples in composite used for detection limit assays.

1All detection limits are in the units of grams of fecal DNA.
amplification with the human HF134 assay was low when challenged against cattle and wildlife fecal DNA (i.e. 4% and 6%, respectively), but frequent with horse samples (i.e. four of six fecal samples). The HF183 assay appeared to be slightly more specific than the HF134 assay, as of all nontargeted samples only one cattle sample cross-amplified with the former assay.

Application of host-specific *Bacteroides* 16S rRNA gene markers uncovered interesting spatial and temporal trends of fecal pollution in the Plum Creek watershed. Nearly all of the water samples (i.e. 94%) analyzed in this study produced positive signals using the general *Bacteroides* assay, while 54% and 28% of the water samples were positive for the ruminant-specific assays CF128 and CF193, respectively (Table 2). In all sites that CF193 produced positive signals, the water samples were also positive when using the CF128 assay, further suggesting the presence of ruminant fecal pollution in these sites. Among the six sites sampled along the Plum Creek, water samples from Sites 1 and 2 were associated with the lowest number of positive ruminant specific signals. The low amount of positive ruminant signals at Site 2 is in contrast with the fact that cattle were present nearby this site. However, water samples downstream from Site 2 showed an increase in amplification percentages when using ruminant-specific assays, with samples from Site 5 achieving 100% amplification using the CF128 assay. On a spatial scale, these PCR-based assays indicated ruminant-fecal pollution incidence generally increased along the downstream tributary gradient of Plum Creek (Fig. 1). This observation may suggest an accumulation of ruminant-associated fecal pollution along the length of Plum Creek reflected by the increasing amount of potential cattle source contributions to Plum Creek along its entirety. Ruminant-associated contamination was observed in all sampling seasons, with summer water samples showing the highest number of positive ruminant signals after short rainfall events that occurred during July 2004 (Fig. 1).

PCR amplification using the HF134 primer (i.e. human-specific assay) was only positive for a limited number of water samples from Sites 4, 5 and 6 (i.e. 5% of all water samples). The HF183 assays supported these results as a limited number of water samples from Site 4 and 6 amplified as well. This low percentage of human-specific amplification may be indicative of the relatively low ratio of humans (i.e. <500) to cattle (i.e. 2000–5000) existing in Plum Creek watershed. However, it should be noted that in this study the human assays amplified only 40% of the septic samples, suggesting that these assays might be underestimating the overall contribution of human sources in this watershed. The low efficiency of the human-targeted *Bacteroides*-based assays has previously been reported, suggesting their somewhat limited value in MST studies (Carson et al., 2005; Fogarty & Voytek, 2005). In contrast, PCR results for sediment showed that nearly 24% and 19% of the sediment samples were positive for the HF134 and HF183 PCR assays, respectively, while only 14% of the sediment samples were positive with the ruminant-specific assays, even though nearly 81% of the sediment samples produced a PCR product with the total *Bacteroides*-like specific assay. Sediment-derived PCR results may be biased, as sediment cores were only collected from one of the six sampling sites along the Plum Creek. No amplification trends were evident among samples from different sediment depths, although all depths amplified at least once with all assays. Higher levels of human-associated fecal pollution in sediment

<table>
<thead>
<tr>
<th>Samples (n)</th>
<th>Bac32</th>
<th>CF128</th>
<th>CF193</th>
<th>HF134</th>
<th>HF183</th>
<th>E. coli</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cattle fecal samples (101)</td>
<td>97*</td>
<td>96</td>
<td>90</td>
<td>4</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>Horse fecal samples (6)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>67</td>
<td>0</td>
<td>–</td>
</tr>
<tr>
<td>Wildlife fecal samples (66)</td>
<td>32</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>–</td>
</tr>
<tr>
<td>Septic samples (10)</td>
<td>90</td>
<td>0</td>
<td>0</td>
<td>40</td>
<td>40</td>
<td>–</td>
</tr>
<tr>
<td>Sediment samples (21)*</td>
<td>81</td>
<td>14</td>
<td>14</td>
<td>24</td>
<td>19</td>
<td>–</td>
</tr>
<tr>
<td>Water samples Site 1 (26)</td>
<td>96</td>
<td>12</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>Water samples Site 2 (26)</td>
<td>96</td>
<td>19</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>270</td>
</tr>
<tr>
<td>Water samples Site 3 (26)</td>
<td>92</td>
<td>38</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>381</td>
</tr>
<tr>
<td>Water samples Site 4 (28)</td>
<td>89</td>
<td>68</td>
<td>21</td>
<td>14</td>
<td>11</td>
<td>534</td>
</tr>
<tr>
<td>Water samples Site 5 (26)</td>
<td>96</td>
<td>100</td>
<td>54</td>
<td>4</td>
<td>0</td>
<td>466</td>
</tr>
<tr>
<td>Water samples Site 6 (26)</td>
<td>93</td>
<td>85</td>
<td>50</td>
<td>8</td>
<td>8</td>
<td>302</td>
</tr>
<tr>
<td>Total water samples (158)</td>
<td>94</td>
<td>54</td>
<td>28</td>
<td>5</td>
<td>3</td>
<td>–</td>
</tr>
</tbody>
</table>

*Percent positive total *Bacteroides*-like, ruminant and human-targeted PCR amplification for 101 cattle fecal samples, 66 wildlife fecal samples, six horse fecal samples, 10 septic tank samples representing human fecal samples, 158 total water samples (n = 26–28 per sampling site along Plum Creek). Water samples were collected in duplicate, thus PCR results for each replicate were averaged.

1Sediment samples were collected from Site 2.

2Geometric mean densities of *E. coli* for Sites 1 through 6 measured in CFU 100 mL⁻¹.
samples may also reflect differences in survival rates of Bacteroides species from different hosts. Survivability of Bacteroides species in water is impacted by dissolved oxygen concentration and temperature (Gerba & McLeod, 1976; Howell et al., 1996; Jones et al., 1980; LaLiberte & Grimes, 1982; Gordon & Cowling, 2003); however, the survivability of the different Bacteroides species in sediment is poorly understood. Thus, survivability and accurate identification of fecal bacteria from sediment environments may be imperative in understanding sediment’s contribution to presumed fecal loads in the water column (Hood & Ness, 1982). If indeed populations from different sources have different survival rates in sediments, these populations might not be the best targets to track recent fecal contamination events.

Sequence and phylogenetic analysis

To further understand and assess environmental host-distribution of Bacteroides-like populations in the Plum Creek watershed, clone libraries were constructed using DNA from cattle, horse, wildlife, human (septic samples) hosts, water samples from each site and sediment samples from Site 2. A total of 142 sequences were identified as potential chimeric sequences and were removed from the phylogenetic analysis. The final phylogenetic analysis contained 1271 sequences from fecal samples (405 cattle, 174 wildlife, 167 septic tanks, 87 horse), water (319) and sediment (119) samples. Phylogenetic relationships were inferred from partial 16S RNA gene sequences (527 positions) (E. coli bases 151 to 678) of clones recovered from Plum Creek watershed environmental samples, cultured members of the Bacteroides and Prevotella genera and previously identified uncultivated fecal bacteria of the order Bacteroidales bacteria from different hosts (Paster et al., 1994; Simpson et al., 2004; Dick et al., 2005). Among all the unique clone sequences from the different library types analyzed in this study, 19–64% exhibited high sequence similarity (≥ 97%) to Bacteroidales-like 16S RNA gene sequences in the publicly available databases, while 0–28% showed relatively low sequence similarity (< 93%) (Fig. 2).

The phylogenetic analysis of the clones indicated some interesting relationships among host fecal bacterial distributions in fecal, water and sediment matrices (Fig. 3). Most notably, almost half (46%) of all the sequences in this study (581 sequences) exclusively derived from cattle feces and water samples comprised a cluster of their own, forming a remote sister group to cultured Bacteroides genera. This latter group also clustered with previously identified cattle fecal members (i.e. CF123 and CF151 clusters) (Bernhard & Field, 2000b; Dick et al., 2005). This observation of endemic host-distribution of ruminant Bacteroides is supported by Dick et al. (2005) in which ruminant Bacteroidales sequences shared the same cluster and were not present in clusters of other host species. The unique anatomy and physiology of the ruminant digestive system may have provided a different evolutionary pathway for Bacteroides-like organisms to live than those inhabiting nonruminant hosts (Dick et al., 2005).

This large ruminant-associated cluster also provides further evidence of cattle fecal pollution present in the Plum Creek, as several cattle fecal-derived sequences shared 100% sequence identity to sequences acquired from water sample clone libraries. Additionally, 57% of all water-derived sequences clustered with cattle-derived sequences (Fig. 3). Sequences obtained from water samples from all six sites were represented in this large supercluster, while some sequences from exclusively water clone libraries

![Image](https://example.com/image.png)
formed clusters outside of this supercluster, indicating even this extensive fecal clone library construction did not represent all fecal Bacteroidales present in the water environment. Additionally, because fecal samples used in this study represent only a portion of the hosts contributing to fecal loadings in Plum Creek, fecal samples may not have adequately represented the overall molecular diversity associated with this watershed. Representative sampling of both host fecal material and environmental samples is critical to ensure fecal bacterial diversity is well-characterized.

To a lesser extent, phylogenetic analysis also uncovered relationships among wildlife and human (i.e. septic samples) derived sequences in both water and sediment samples from Plum Creek. Interestingly, 54% of the wildlife sequences affiliated with cultured members of the Bacteroides spp. group. Most notably, 60 wildlife clones clustered with Bacteroides ovatus, while small exclusive wildlife-derived clades formed with Bacteroides acidofaciens and Bacteroides eggerthii. Human and avian clones have been previously observed to be affiliated to these cultured Bacteroides; (Dick et al., 2005) however, this is the first study in which nonwaterfowl wildlife derived sequences are shown to be associated with these Bacteroides clades (Fig. 3). Nearly 60% of all septic clones clustered with cultured Bacteroides spp., with 89 clones composing an exclusive cluster with Bacteroides thetaiotaomicron, a bacteria species primarily associated with human fecal contamination (Holdeman et al., 2004; Simpson et al., 2004; Dick et al., 2005). Unidentified Env. refers to bacterial sequences obtained from Plum Creek water and sediment samples.

Fig. 3. Schematic representation of the host distribution of Bacteroides 16S rRNA gene sequences in Plum Creek based on a neighbor-joining algorithm. Information in parenthesis refers to the number of sequences per library type (i.e. fecal, water and sediment) as compared to cultured Prevotella and Bacteroides species, as well as previously identified nonculturables CATTLE, ELK, HUMAN, CAT, DOG and HORSE Bacteroidales bacterial sequences (Paster et al., 1994; Simpson et al., 2004; Dick et al., 2005). Unidentified Env. refers to bacterial sequences obtained from Plum Creek water and sediment samples.
Many of the remaining wildlife and septic clones clustered in unidentified environmental clades, distinct from the *Bacteroides* and *Prevotella* spp. clades, clustering with 86% of all sediment-derived sequences and nearly one-third of water-derived clones. For example, in an unidentified environmental and fecal clade, 30 wildlife and 16 septic-derived clones affiliated with 32 and 23 sediment and water clones, respectively. Compared to the cattle-like sequences, clones generated in this study from septic and wildlife were found in numerous clades, suggesting a relatively cosmopolitan lifestyle of many of these bacterial populations. In addition, novel clades outside the *Bacteroides* and *Prevotella*-related groups containing sequences from septic, wildlife, sediment and water samples were identified, further suggesting that the *Bacteroidales* is a large and complex bacterial group.

Overall, fewer clone sequences clustered with *Prevotella*-related species. One water sample clone from Site 6 was 100% similar to *Prevotella ruminicola*, a group commonly found in the rumen and hindgut of mammals. Although 13 of the 87 horse-derived bacterial sequences were members of a cosmopolitan clade related to the *Bacteroides* group, the vast majority of the horse clones clustered with *Prevotella* spp. and *Prevotella*-like groups. A horse-specific cluster containing 58 sequences was previously documented by Dick *et al.* (2005). Additionally, three sediment sequences previously described as specific to cat, dog and human formed a small clade within the *Prevotella* group. Very few (two clones) cattle fecal-derived sequences were closely related to cultured *Prevotella* species. These trends were noted by Dick *et al.* (2005), as sequences from nonruminant feces fell within the genera *Bacteroides* and *Prevotella*, while most sequences from ruminant hosts did not cluster with any known species.

Another notable feature of the phylogenetic analysis was the apparent clustering of sediment-derived sequences to culturable *Bacteroides* genera and septic sequences as seen in the *Bacteroides uniformis* and *Bacteroides ovatus*-related groups. These clusters provide some evidence of human-fecal contamination in the sediment samples perhaps originating from surrounding septic systems used in this watershed. This hypothesis however becomes confounded, as the *B. ovatus* cluster also contains several wildlife bacterial clones. These observations may be explained by the cosmopolitan lifestyle of cultured fecal *Bacteroidales* with respect to their host, as explained by the horizontal transmission of fecal bacteria among hosts that may have more interaction with each other, resulting in higher rates of horizontal transfer of fecal bacteria (Dick *et al.*, 2005).

The overall topology of the neighbor-joining phylogenetic analysis of *Bacteroides*-like 16S rRNA gene sequences was supported by parsimony trees with 100 resamplings (data not shown). The branch of the tree containing a majority of the cattle fecal sequences and water-derived sequences was supported in 85% of the parsimony bootstraps, while two branches of the tree containing cultured *Bacteroides* species (*B. acidofaciens* and *Bacteroides fragilis*) and fecal sequences were not supported by high bootstrap values. The low bootstrap values may be a result of the use of partial 16S rRNA gene sequences limiting phylogenetic resolution in a comparative analysis (Hopkins *et al.*, 2001). All other superclusters in the parsimony analysis were supported by bootstrap values of 50% and higher.

Rarefaction analysis was conducted to evaluate whether screening 1271 clones was sufficient to estimate diversity of *Bacteroides*-like sequences derived from fecal and environmental clone libraries from Plum Creek (Fig. 4). The calculated rarefaction curves did not reach saturation, indicating that analysis of an increasing number of clones would have revealed further diversity. Rarefaction analysis did indicate that the richness of species appeared to be higher in sequences derived from water clone libraries (120 operational taxonomic units, herafter OTUs) as compared to cattle fecal sequences (85 OTUs) and other fecal sequences (102 OTUs). This observation may indicate that our fecal source sampling was not extensive enough to capture the overall *Bacteroides*-like diversity in water samples. These results further revealed that the molecular diversity of *Bacteroides*-like bacteria in the environment is high. Attempts to quantify specific groups using techniques like qPCR will benefit from a better understanding of the molecular diversity of fecal bacteria. Quantification of fecal sources is an important step towards determining fecal loads and for the development of accurate microbial risk assessment models (Simpson *et al.*, 2002; U.S. Environmental Protection Agency 2005).

The results from the *Bacteroidetes* host-specific assays and the phylogenetic analyses were able to confirm the prevalence of ruminant contamination in Plum Creek. The results also suggest that other sources (e.g. human, horse and wildlife) are potentially contributing to the fecal pollution of this watershed. However, it should be noted that these host-specific PCR assays do have some limitations, as the assays did show some cross-amplification with a small percentage of wildlife fecal samples and most horse fecal samples. Despite these inherent limitations, assessment of fecal pollution in Plum Creek watershed highlighted the applicability of 16S rRNA gene PCR assays and phylogenetic analysis as tools to differentiate among ruminant and non-ruminant fecal source pollution in fecal, water and sediment matrices. Furthermore, notable spatial and temporal variations of fecal pollution could be indicative of ‘hot spots’ associated to particular sources of fecal pollution. This is true only if some of the specific clusters are endemic. For example, in this study ruminant-specific *Bacteroides* sequences illustrated endemism, while nonruminant sequences were less host-specific. While culture-independent PCR-based assays present the opportunity of characterizing microbial ecosystems independent of microbial cultivation, understanding
Assessment of fecal pollution sources


**Fig. 4.** Rarefaction curves for unique 16S rRNA gene library clones. Rarefaction curves were calculated using the analytical approximation algorithm described by Hurlbert (1971) and 95% confidence intervals estimated as described by Heck et al. (1975). The expected number of operational taxonomic units is plotted vs. the number of unique 16S rRNA gene clones. Rarefaction curves were calculated for cattle fecal, other fecal (wildlife, horse and human), water and sediment clone libraries. The error bars indicate the 95% confidence intervals.

the host distribution of host-specific markers and its relationship to water quality regulatory standards are necessary to effectively apply these assays under regulatory scenarios.

**Acknowledgements**

This research was supported in part by an Augmentation Award to J.S.D. from the National Center for Computational Toxicology of the U.S. EPA, Office of Research and Development, by a grant from the Nebraska Department of Environmental Quality and by the USGS Cooperative Water Program. We are grateful for the cooperation of the landowners in the Plum Creek watershed during this project and for the technical assistance from Cathy Keltty, Jinrang Lu and Christopher Luedeker. Any opinions expressed in this paper are those of the author(s) and do not, necessarily, reflect the official positions and policies of the U.S. EPA. Any mention of products or trade names does not constitute recommendation for use by the U.S. EPA or the USGS.

**References**


