The Journal of Microbiology (2011) Vol. 49, No. 4, pp. 585-594
DOI 10.1007/s12275-011-0530-6

Microbial Community Analysis and Identification of Alternative Host-Specific Fecal Indicators in Fecal and River Water Samples Using Pyrosequencing

Ju-Yong Jeong[1,4], Hee-Deung Park[2], Kyong-Hee Lee[1], Hang-Yeon Weon[3], and Jong-Ok Ka[4*]

[1]*Department of Water Quality Research, Gyeonggi-do Institute of Health and Environment, Suwon 440-290, Republic of Korea*
[2]*School of Civil, Environmental and Architectural Engineering, Korea University, Seoul 136-713, Republic of Korea*
[3]*Korean Agricultural Culture Collection (KACC), National Academy of Agricultural Science (NAAS), Rural Development Administration (RDA), Suwon 441-853, Republic of Korea*
[4]*Department of Agricultural Biotechnology and Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 151-742, Republic of Korea*

(Received December 20, 2010 / Accepted March 31, 2011)

It is important to know the comprehensive microbial communities of fecal pollution sources and receiving water bodies for microbial source tracking. Pyrosequencing targeting the V1-V3 hypervariable regions of the 16S rRNA gene was used to investigate the characteristics of bacterial and Bacteroidales communities in major fecal sources and river waters. Diversity analysis indicated that cow feces had the highest diversities in the bacterial and Bacteroidales group followed by the pig sample, with human feces having the lowest value. The Bacteroidales, one of the potential fecal indicators, totally dominated in the fecal samples accounting for 31%-52% of bacterial sequences, but much less (0.6%) in the river water. Clustering and Venn diagram analyses showed that the human sample had a greater similarity to the pig sample in the bacterial and Bacteroidales communities than to samples from other hosts. Traditional fecal indicators, i.e., *Escherichia coli*, were detected in the human and river water samples at very low rates and *Clostridium perfringens* and enterococci were not detected in any samples. Besides the Bacteroidales group, some microorganisms detected in the specific hosts, i.e., *Parasutterella excrementihominis*, *Veillonella* sp., *Dialister invisus*, *Megamonas funiformis*, and *Ruminococcus lactaris* for the human and *Lactobacillus amylovorus* and *Atopostipes* sp. for the pig, could be used as potential host-specific fecal indicators. These microorganisms could be used as multiple fecal indicators that are not dependent on the absence or presence of a single indicator. Monitoring for multiple indicators that are highly abundant and host-specific would greatly enhance the effectiveness of fecal pollution source tracking.

***Keywords*:** bacteroidales, microbial source tracking, potential fecal indicator, pyrosequencing

Bodies of water contaminated with feces may potentially contain pathogens, which pose a great risk to human health. Traditionally, microorganisms such as total fecal coliforms and *Escherichia coli* have been used as fecal indicators. The detection of the microorganisms in a water sample is an indication of fecal contamination. However, these indicator microorganisms appear to persist and grow in aquatic environments (Walk *et al.*, 2007). Additionally, these test methods cannot determine the actual sources of contamination. Thus, it is difficult to quantify and trace the sources of fecal contamination with a test based only on indicator microorganisms. More reliable and practical fecal source identification and quantification methods are needed to carry out proper management practices to control fecal contamination sources, to protect people from waterborne pathogens and to conserve the integrity of drinking water supplies.

Over the past decade, culture-independent molecular approaches have been widely used for microbial source tracking. These approaches are mostly based on identification of a host-specific gene marker or host-specific molecular fingerprint from test samples. Generally, the 16S rRNA gene is used to design host-specific gene markers or fingerprints. The design starts with the extraction of total DNA in fecal samples, followed by PCR amplification of the 16S rRNA gene using universal or group-specific primers, cloning and sequencing of the gene. Next, host-specific primer sets targeting potential gene markers are designed based on comparison of clone libraries *in silico*. Finally, the effectiveness of the design is confirmed experimentally *in vitro*. Using the protocol, numerous human-, cow-, and pig-specific fecal indicators have been developed (Layton *et al.*, 2006; Okabe *et al.*, 2007).

However, these source tracking methods were developed with a relatively small number of 16S rRNA sequences (less than 1,000 sequences). When considered in light of the up to 100 trillion (10^{14}) microorganisms in the human intestine, the number of 16S rRNA sequences for designing host-specific gene markers or fingerprints appears to be very low. As a result, many assays with host-specific primer sets do not have reliable specificities for identifying and quantifying fecal pollution sources (McLain *et al.*, 2009), which requires alternative fecal indicators to increase the choice of tools and accurately identify the fecal pollution source in water (Marti *et al.*, 2010).

Recently, a high-throughput sequencing technology called

* For correspondence. E-mail: joka@snu.ac.kr; Tel.: +82-2-880-4673; Fax: +82-2-871-2095

pyrosequencing has begun to be actively used in the life sciences, including microbial ecology and environmental microbiology. Pyrosequencing can generate hundreds of thousands of short sequences from hypervariable regions of the rRNA genes (Dethlefsen *et al.*, 2008). Pyrosequencing can increase the molecular sampling size required to detect rare taxa and to enhance the effectiveness of fecal source tracking (McLellan *et al.*, 2009), which appears to be an effective feature for designing host-specific gene markers or fingerprint signatures. Several researchers have used pyrosequencing to investigate health-related microbial communities in the human gut (Dethlefsen *et al*., 2008), as well as those found in cattle feces (Dowd *et al*., 2008a), pig ileum (Dowd *et al*., 2008b), sewage (McLellan *et al*., 2009) and the activated sludge of wastewater treatment plants (Sanapareddy *et al*., 2009). Recently, pyrosequencing techniques were used to identify candidate alternative human fecal indicators and multiple signatures by comparing microbial communities of human feces, human-derived sewage, and surface water (McLellan *et al*., 2009). However, that study did not conduct a comprehensive analysis of cow and pig sources, both major fecal pollution sources (Jeong *et al.*, 2008; Mieszkin *et al.*, 2009), in order to compare microbial communities.

It is vital to use large numbers of fecal source sequences (>several thousand) and to systematically compare the sequences with those of major fecal sources (i.e., cow, pig, and human) for an effective design of host-specific gene markers or fingerprints. In this study, we investigated the microbial communities of major fecal sources including humans, pigs, beef cattle, and dairy cattle using pyrosequencing that targeted the V1, V2, and V3 hypervariable regions of the 16S rRNA gene to determine how these microbial communities are constituted.

The specific research objectives of this study were 1) to evaluate microbial community structures while emphasizing the Bacteroidales group, one of the most dominant microorganisms in the feces of humans and animals and a promising host-specific indicator; 2) to compare the microbial diversity and similarities of fecal and river water samples; and 3) to propose alternative fecal indicators.

Materials and Methods

Fecal samples and DNA extraction

Fecal samples were collected from ten humans including eight adults and two children, ten cows including five dairy cows and five pigs from livestock farms in Icheon City, Gyeonggi-do, Korea. River water samples were collected from 11 sites on four rivers in May 2009. The four rivers were the Anseong, Jinwi, Osan, and Hwangguji that compose the Anseong river watershed and are the national water quality monitoring sites operated by the Gyeonggi-do Institute of Health and Environment. Two sites were selected from the Anseong and three sites, upstream, midstream, and downstream, were chosen from the others. The Anseong river basin has a population greater than 2 million inhabitants, 100,000 cattle and 500,000 pigs. There are many small-scale livestock farms around the basin. Streams in this area are heavily polluted except for the upstream segment of the Jinwi River. The fecal pollution sources of these rivers are likely to be mainly human, with some contribution from cattle and pigs (Fig. 1).

Fecal DNA was extracted and purified from about 200 mg of fecal sample using the QIAamp stool DNA Mini kit (QIAGEN, USA) according to the manufacturer's instructions. To extract DNA from the river water samples, about 200 ml of water was filtered through a 0.22-μm pore size Durapore membrane filter (Millipore, Ireland). After filtration of the samples, DNA was extracted and purified using the xanthogenate-sodium dodecyl sulfate method (Tillett *et al*., 2001). The stool kit and xanthogenate-sodium dodecyl sulfate method showed efficient DNA extraction and purification for feces and water samples, respectively.

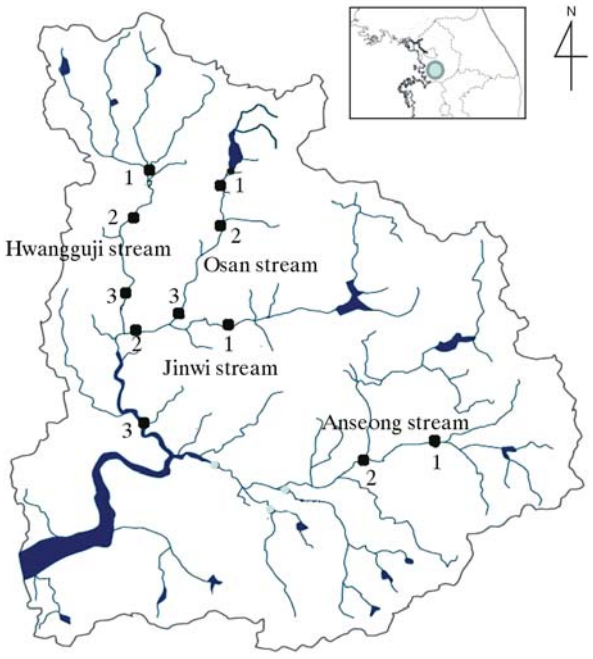

Fig. 1. Map of sampling sites. 1, upstream; 2, midstream; 3, downstream; ■, Sampling site.

Pyrosequencing

In this study, pooled samples were used because our previous studies demonstrated low variability among individual samples in the bacterial community (Jeong *et al*., 2008, 2010). Dick *et al.* (2005) also used pooled samples to identify host-specific Bacteroidales genetic markers. Equal amounts of DNA from fecal sources and river water sources were pooled to create representations for the beef cow, dairy cow, human, pig, and river water pyrosequencing samples. In our previous study, it was demonstrated that individual samples had similar PCR efficiencies in a real-time PCR assay targeting the Bacteroidales group (Jeong *et al*., 2010). The primer set, 27F (GAGTTTGATCMTGGCTCAG) and 518R (WTTACCGCGGCTGCTGG) targeting the V1, V2, and V3 hypervariable regions, was used to amplify the *E. coli* positions 27-518. Approximately 1 μg of the final PCR product DNA was used to generate DNA libraries for use with the Genome Sequencer FLX (Roche, Germany). The ends of the PCR products were blunted and one short adapter (14 bp long) was ligated onto each end. The adapters provide priming sequences for both amplification and sequencing of the sample library fragments and also the sequencing key, a short sequence of four nucleotides used for base calling. Single effective copies of template species from the DNA library to be sequenced were hybridized to DNA capture beads, as

Table 1. Statistics for pyrosequencing data

Sample	Beef cow	Dairy cow	Human	Pig	River water
No. of reads	8,189	9,718	9,835	7,614	9,007
Max. read length (bp)	534	539	547	531	514
Median read length (bp)	478	479	481	490	469
Min. read length (bp)	300	300	300	300	300
% of *Bacteroidales* group	19.2	21.3	43.7	54.9	0.6

one fragment per bead and amplified using emulsion-based clonal amplification. After amplification, the beads were placed in the wells of a PicoTiterPlate device (PTP), along with the proper reagents and enzymes. The loaded PTP was then inserted into the Genome Sequencer FLX Instrument (Roche) for pyrosequencing. All of the procedures were performed in accordance with the manufacturers' instructions and were conducted at Macrogen (Korea).

Data analysis

The raw sequences were first processed using the GL FLX software (Roche) for sorting according to key and to trim low-quality sequences and the primers. In addition, the sequences more than 300 bases in length were selected for further analysis. Next, the 16S rRNA gene fragments were phylogenetically classified from phylum to genus using the RDP classifier tool (Wang *et al.*, 2007) with an 80% confidence threshold. The sequences assigned to the order Bacteroidales were also extracted from each sequence library using the FASTA sequence selection tool in RDP's pyrosequencing pipeline. The two merged clone libraries, containing the bacterial or Bacteroidales sequences, were aligned with the aligner in the RDP's pyrosequencing pipeline. The alignments were trimmed to consider the overlapping alignment columns. Then, the column-formatted distance matrices were created and microbial diversity was analyzed using the Mothur software (Schloss *et al.*, 2009).

Each sequence was assigned to operational taxonomic units (OTUs) based on distances of 0.03. At each dissimilarity level, coverage percentages were calculated using Good's method as G=1-n/N, where n is the number of singletons and N is the total number of sequences in the sample. Rarefaction analysis of each library was then conducted and nonparametric diversity indices, such as Chao1, ACE, and Shannon, were calculated to measure the diversity of each library. In addition, community trees were generated to describe the similarity among the bacterial and Bacteroidales communities in each sample. Groups were clustered using the UPGMA algorithm, based on a pairwise distance matrix between samples (Schloss *et al.*, 2009). Venn diagram analysis was used to compare the richness based on the observed OTUs among the samples.

Identification of alternative potential fecal indicators

Venn diagram analysis was used to identify alternative potential host-specific indicators. The shared OTUs between samples, with the sequence difference defining OTUs set at 3%, were omitted to obtain the host-specific OTUs. The resulting sequences were phylogenetically assigned using the RDP classifier. Then, considering the sequences assigned to the specific genus level and excluding rarely-occurring sequences, the potential host-specific indicators were identified. In addition, the EzTaxon server 2.1 (Chun *et al.*, 2007) was used to identify the shared genera among samples to the species level.

Nucleotide sequence accession numbers

The 16S rRNA gene sequences determined in this study were deposited into GenBank under accession numbers GU599989 to GU644351.

Results

Taxonomic distribution of fecal and river water microbial communities

After filtering the poor-quality sequences that had little or no similarity to the 16S rRNA gene sequences in the GenBank database, 8,189 16S rRNA gene sequences were obtained for the beef cow, 9,718 for dairy cattle, 9,835 for humans, 7,614 for pigs, and 9,007 from the river water samples. Table 1 sum-

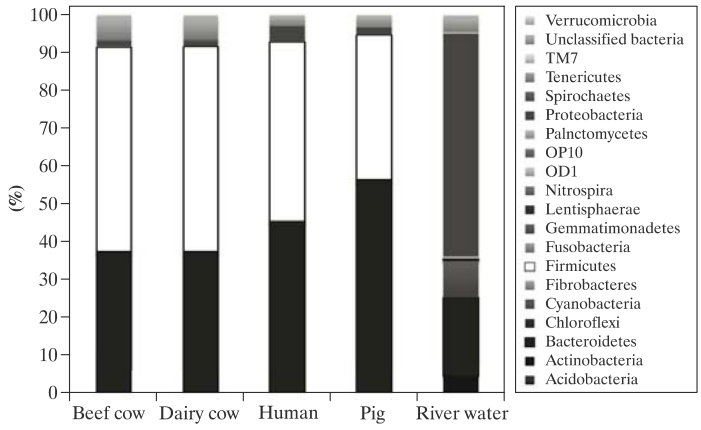

Fig. 2. Bacterial phylum diversity based on 16S rRNA gene clone libraries and the RDP classification in the fecal and river water samples.

Table 2. Distributions of Bacteroidales group in the fecal and river water samples assigned by RDP classifier

Taxonomy		% of Sequences				
Family	Genus	Beef cow	Dairy cow	Human	Pig	River water
Bacteroidaceae		**3.30**	**3.33**	**47.91**	**0.19**	**32.14**
	Bacteroides	3.30	3.33	47.91	0.19	32.14
Prevotellaceae		**7.68**	**11.16**	**36.78**	**90.46**	**7.14**
	Prevotella	0.57	1.16	34.50	72.04	3.57
	Paraprevotella	1.78	2.66	1.47	0.02	N.D.
	Hallella	0.19	0.05	N.D.	0.02	N.D.
	Xylanibacter	N.D.	N.D.	N.D.	0.29	N.D.
	Unclassified	5.14	7.30	1.86	18.08	3.57
Porphyromonadaceae		**9.14**	**10.49**	**6.47**	**3.47**	**55.36**
	Parabacteroides	0.76	0.48	3.19	0.26	5.37
	Paludibacter	0.25	0.39	N.D.	N.D.	26.79
	Odoribacter	0.06	N.D.	0.16	N.D.	N.D.
	Barnesiella	N.D.	0.05	1.72	N.D.	N.D.
	Butyricimonas	N.D.	N.D.	0.30	N.D.	N.D.
	Unclassified	8.06	9.57	1.09	3.20	25.00
Rikenellaceae		**16.12**	**17.45**	**3.19**	**N.D.**	**3.57**
	Alistipes	14.53	15.90	3.19	N.D.	3.57
	Rikenella	0.76	0.72	N.D.	N.D.	N.D.
	Unclassified	0.82	0.82	N.D.	N.D.	N.D.
Unclassified_Bacteroidales		**63.77**	**57.56**	**5.65**	**5.88**	**1.79**
Number of sequence		1,563	2,055	4,299	4,183	56

ND, Not detected

marizes the number of reads and the maximum, median, and minimum read lengths for each sample. The percentages of sequences falling into the Bacteroidales group were relatively high (i.e., 19.2-54.9% of the total sequences in the fecal samples).

When the RDP classifier was used to assign phylogenetic affiliations to each sequence with an 80% bootstrap cutoff, 95.4% of the sequences were classified to a specific phylum. Figure 2 shows the phylum-level taxonomic compositions of each sample. The numbers of phyla detected in the fecal samples of the beef cow, dairy cow, human and pig samples were 14, 12, 7, and 7, respectively, and 17 were detected in the river water sample. The phyla Bacteroidetes and Firmicutes accounted for 37-57% and 38-54% of the bacterial sequences in the fecal samples, respectively. Nearly 97% of the Bacteroidetes corresponded to the class Bacteroidia in the human and pig feces, with the relevant percentages being 51.5% and 56.9% for the beef cow and dairy cow samples, respectively. The class Clostridia accounted for 88-99% of the Firmicutes sequences in the fecal samples. In addition, the ratios of the phyla Bacteroidetes/Firmicutes were substantially different between the fecal samples, i.e., 0.69, 0.69, 0.96, and 1.48 for the beef cow, dairy cow, human, and pig samples, respectively. The phylum Spirochaetes occupied about 1.0% of bacterial se-

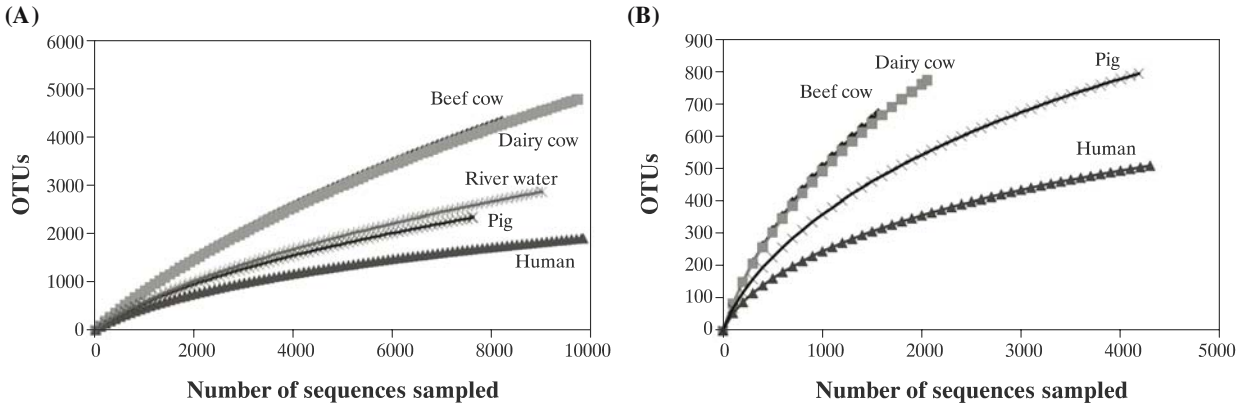

Fig. 3. Rarefaction curves describing the dependence of discovering novel OTUs defined by 0.03 distance cutoff in (A) bacteria and (B) Bacteroidales group. The y axis shows the number of OTUs that would be found after sampling the number of reads or sequences indicated on the x axis.

Table 3. Comparison of diversities of bacterial and Bacteroidales group based on a sequence difference of 3%

Sample	Reads	Bacterial					Reads	*Bacteroidales*				
		OTUs	Chao1	ACE	Shannon	Coverage		OTUs	Chao1	ACE	Shannon	Coverage
Beef cow	8,189	4.347	10,519	16,962	7.97	0.64	1,576	673	1,380	1,873	5.97	0.74
Dairy cow	9,718	4,793	11,002	19,067	8.01	0.67	2,069	776	1,564	2,304	6.06	0.78
Human	9,835	1,916	3,028	3,180	6.06	0.91	4,298	511	742	765	4.62	0.95
Pig	7,614	2,349	4,876	7,412	6.67	0.81	4,181	797	1,283	1,324	5.41	0.91
River water	9,007	2,876	5,967	8,947	6.82	0.81	56	29	67	57	3.13	0.70

quences in the beef cow and dairy cow samples but was not found in human feces and occurred at a low rate in the pig sample (0.05%). In contrast to the Bacteroidetes and Firmicutes results, Proteobacteria accounted for only 0.7-4.1% of the total sequences in the fecal samples. Although many sequences, especially those of the beef and dairy cows, were not assigned to a specific genus, the genus *Oscillibacte*r was the most dominant in the beef (6.6%) and dairy cow (7.9%) samples. The genera *Prevotella* and *Bacteroides* accounted for about 14.6% and 20.9%, respectively, of the organisms in the human sample. In addition, the genus *Prevotella* was also predominant in the pig sample (39.6%). In the river water sample, the genera *Bacillariophyta* and *Flavobacterium* constituted about 9.0% and 6.2% of the microorganisms, respectively.

The order Bacteroidales, which incorporates the most frequently used microorganisms in microbial source tracking using culture-independent methods, were dominant in the fecal samples, accounting for 19-55% of the total sequences (Table 2). The taxonomic distributions of the Bacteroidales group exhibited different profiles in each sample; approximately 60% of this group was assigned to unclassified Bacteroidales in the beef cow and dairy cow samples. However, the unclassified Bacteroidales accounted for only about 6% in the human and pig samples. In addition, the Prevotellaceae bacteria dominated in the pig Bacteroidales sequences (90%); the Bacteroidaceae and Prevotellaceae bacteria accounted for 48% and 37%, respectively, in the human Bacteroidales sequences. The detailed information on the Bacteroidales classification is represented in Table 2.

Microbial community diversity

The Mothur software estimated 4,347, 4,793, 1,916, 2,349, and 2,876 bacterial OTUs and 673, 776, 511, 797, and 29 Bacteroidales OTUs for beef cow, dairy cow, human, pig and river water samples, respectively, when set at a sequence difference of 3%. Rarefaction curve analysis was used to evaluate the degree of completion in each taxonomic sampling. Estimates of OTUs increased with the number of reads and a plot of OTUs versus the number of reads indicated that the dairy cow sample had the greatest richness, followed by that of the beef cow sample. Human feces showed a significantly lower richness than the beef cow and dairy cow feces (Fig. 3).

The nonparametric methods (i.e., ACE, Chao1, Shannon) provided estimates that also varied with the sample. The nonparametric estimators (i.e., Chao1) of richness were 58-130% and 45-131% higher than the observed richness for the bacteria and Bacteroidales, respectively. While rarefaction and richness estimators consider survey completion from the perspective of identified and unidentified taxa, coverage considers completion from the perspective of individual reads or sequences. Good's coverage, the estimated likelihood that a randomly chosen sequence from the sample will belong to an OTU that has already been identified in the dataset, was 74%, 78%, 95%, 91%, and 70% for Bacteroidales in the beef cow, dairy cow, human, pig and river water samples, respectively; more sequences are needed to detect a new OTU at a 3% cutoff for the human and pig samples than for the beef cow, dairy cow and river water samples (Table 3).

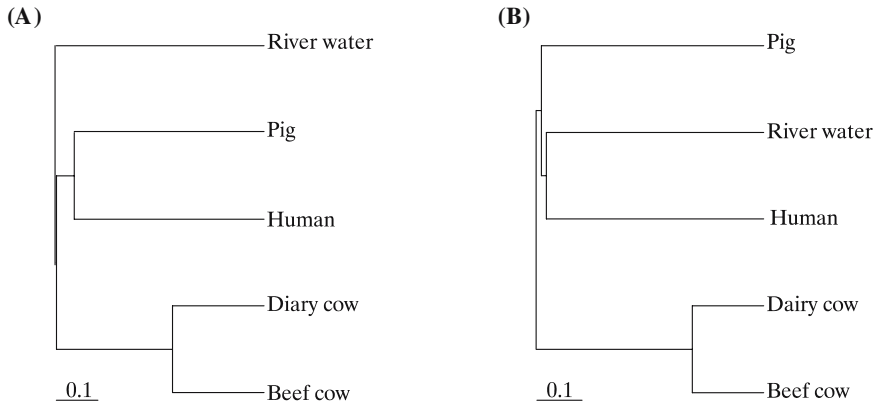

Fig. 4. Community tree representing the bacterial (A) and Bacteroidales (B) communities' sample similarity based on the structure-based Yue & Clayton θ similarity coefficient at OTU cutoff of 3%. The scale bar indicates genetic distance (0.1 nucleotide substitution per site).

Table 4. Potential host-specific and conventional indicator occurrences in the fecal and river water samples

Taxonomy		% of Sequences				
Order/Family	Genus/Species	Human	Beef cow	Dairy cow	Pig	River water
Bacteroidaceae	*Bacteroides vulgatus*	3.13	ND	ND	ND	0.06
Porphyromonadaceae	*Odoribacter*	0.09	ND	ND	ND	ND
	Barnesiella	0.90	ND	0.01	ND	ND
	Butyricimonas	0.16	ND	ND	ND	ND
Rikenellaceae	*Alistipes*	1.67	2.36	3.35	ND	0.02
Alcaligenaceae	*Sutterella*	2.96	ND	ND	0.01	ND
	Parasutterella	0.08	ND	ND	ND	ND
Veillonellaceae	*Veillonella*	0.20	ND	ND	ND	ND
	Dialister	4.47	ND	ND	1.08	0.04
	Megamonas	0.39	ND	ND	ND	ND
	Mitsuokella	ND	ND	ND	0.85	ND
	Selenomonas	ND	ND	ND	2.26	ND
Lactobacillaceae	*Lactobacillus*	0.01	ND	ND	1.38	ND
Carnobacteriaceae	*Atopostipes*	ND	ND	ND	0.83	ND
Clostridiaceae	*C. perfringens*	ND	ND	ND	ND	ND
Enterobacteriaceas	*E.coli*,	0.46	ND	ND	ND	0.02
Bifidobacteriaceae	*Bifidobacterium*	ND	ND	ND	ND	0.01
Enterococcaceae	*enterococci*	ND	ND	ND	ND	ND

ND, Not detected

Comparison of microbial communities between samples
Community trees were constructed to describe the dissimilarity (1-similarity) among the fecal and river water samples. Groups were clustered using the UPGMA algorithm, with the distance between communities calculated using the bacteria and Bacteroidales similarities in community membership and structure. Figure 4A shows the comparison of microbial profiles between the fecal and river water samples for OTUs defined at 0.03, in which the human-pig and beef cow-dairy cow samples were clustered, respectively. When we clustered the Bacteroidales group with the membership-based Jaccard coefficient calculated using the Chao1 estimated richness values, the human sample was clustered with the pig sample at the 3% OTU cutoff. The Jaccard's index is most frequently used to compare two communities, (i.e., the ratio of the number of shared OTUs to the number of distinct OTUs in two communities). Jaccard's index can be expressed as $\theta_J=(s_0/s_1+s_2-s_0)$, where s_i is the number of OTUs in community i, i=1, 2 and S_0 is the number of shared OTUs for the two communities (Yue and Clayton, 2005). A Venn diagram analysis was used to calculate the shared OTUs with a 3% difference among the samples; these samples are illustrated in Fig. 6. The majority of OTUs represented by the bacterial or Bacteroidales communities were host-specific. That is, the human samples shared 9.9% and 4.7% of OTUs at a 3% cutoff with pig samples for the bacteria (Fig. 6A) and Bacteroidales group (Fig. 6B), respectively, while the beef cow samples shared 34% and 46% of OTUs with the dairy cow samples.

Microbial community characteristics of river water
Most 16S rRNA gene sequences from the river water were assigned to Proteobacteria, Bacteroidetes, Cyanobacteria, and Actinobacteria (Fig. 2). The Bacteroidales group occupied only 0.6% of the total sequences, compared with 19.2-54.9%

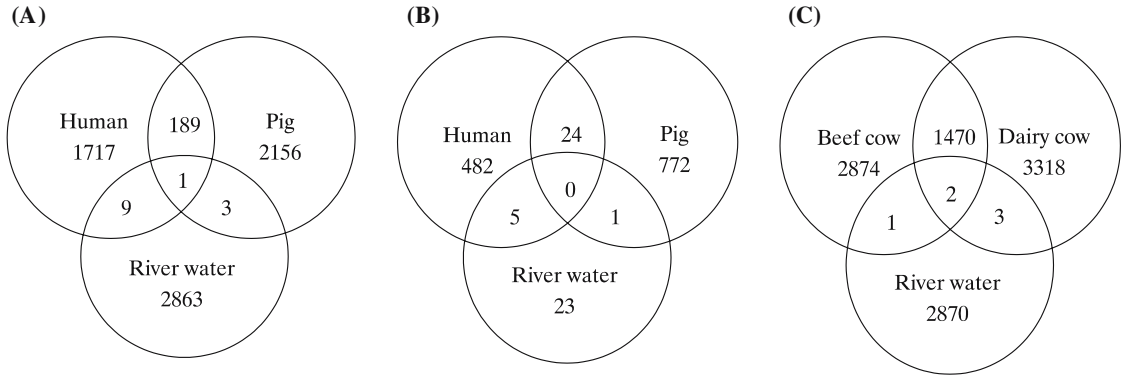

Fig. 5. Venn diagram comparing the OTUs found in the human, pig, beef cow, dairy cow, and river water samples, (A, C) for total bacteria, (B) for Bacteroidales group.

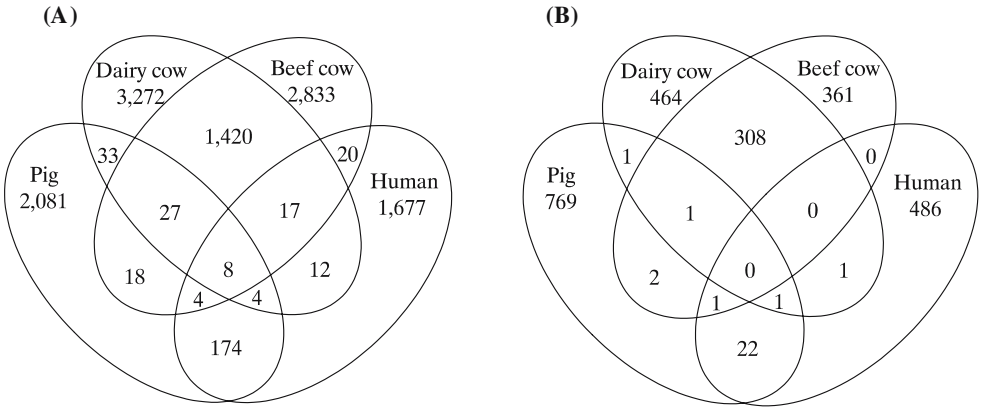

Fig. 6. Venn diagrams comparing the OTU memberships found in the beef cow, dairy cow, human, and pig samples: (A) for bacteria and (B) for Bacteroidales group, OTUs based on a sequence difference of 3%.

in the fecal samples. This was not surprising because these organisms are strict anaerobes that cannot multiply or survive in aerobic waters. In addition, Sphingobacteria and Flavobacteria were detected at very low rates in the fecal samples, accounting for 33% and 49% of the Bacteroidetes in the river water, respectively, while merely 0.4% of the bacterial sequences corresponded to the Firmicutes. In contrast to the Bacteroidetes and Firmicutes, Proteobacteria was predominant (59%) in the river water, while accounting for only 0.7-4.1% in the fecal samples. The phylum Actinobacteria, previously known as one of the main taxa in human feces (Claesson *et al.*, 2009), was not found at significant levels in the fecal samples; it accounted for 4.4% of bacterial sequences in the river water. Most Actinobacteria in the river water belonged to Actinomycetales, with no Bifidobacteriales sequences. When we compared the Bacteroidales group according to the structure-based Θ similarity coefficient, the river water was clustered with human feces at a cutoff of 3% (Fig. 4B).

A Venn diagram analysis showed that the river water shared ten and five bacteria and Bacteroidales OTUs at a 3% cutoff with the human sample, respectively. The shared OTUs were *Enterobacter cancerogenus*, *Enterobacter kobei*, *Dialister invisus*, *Dialister succinatiphilus*, *Alitipes putredinis*, *Prevotella* sp., and *Bacteroides vulgatus*. In addition, the pig sample shared four and one OTUs with the river water, all of which were either *Flavobacterium* sp. or *Prevotella* sp. Lastly, the river water shared three and five bacterial OTUs (*Rhodobacter* sp., *Acinetobacter johnsonii*, and *Paracoccus pantotrophus*) with the beef cow and dairy cow samples, respectively, although there were no shared OTUs for the Bacteroidales group at the 3% cutoff (Fig. 5).

Identification of host-specific alternative indicators

A Venn diagram analysis was used to identify the bacterial 16S rRNA gene sequences that were thought to be host-specific; 2,833, 3,272, 1,677, and 2,081 OTUs were obtained for the beef cow, dairy cow, human and pig based on sequence differences of 3%, respectively (Fig. 6A). These potential host-specific OTU sequences were extracted by removing the shared OTUs among samples using the Mothur software. We identified the potential alternative host-specific fecal indicators based on the sequences assigned to specific genus levels, excluding rarely-occurring genera, i.e., <0.02% (Table 4).

When we compared the host-specific sequences of each fecal sample, 11 and 4 potential host-specific indicator organisms were identified for the human and pig samples at the genus level, respectively. No potential indicator was found for the beef cow or dairy cow samples since many OTUs were not able to be assigned to a specific genus using the RDP classifier. However, a previous study showed that significant numbers (i.e., beef cow, 6.26%; dairy cow, 5.02%) of the sequences detected only in the beef cow and dairy cow samples were shown to be close to an uncultured bacterial sequence (i.e., EU470222 assigned by EzTaxon 2.1) (Ley *et al*., 2008). In the Bacteroidales group, the human-specific potential indicators were *Bacteroides vulgatus*, *Barnesiella intestinihominis*, *Alistipes putredinis*, *Alistipes shahii*, and *Alistipes finegoldii*. In addition to the Bacteroidales group, *Parasutterella excrementihominis*, *Veillonella* sp., *Dialister invisus*, *Megamonas funiformis*, and *Ruminococcus lactaris* can also be used as human-specific indicators. For the pig-specific potential indicators, *Lactobacillus amylovorus*, *Mitsuokella jalaludinii*, *Selenomonas bovis*, and *Atopostipes* sp. were identified, with no members of the Bacteroidales group indicated.

We also analyzed the occurrence of traditional fecal indicators. Interestingly, *Bifidobacterium*, known as one of major taxa in the human gut, was identified in the river water as only one sequence. The relative abundances of traditional fecal indicators, *E. coli*, *C. perfringens*, and enterococci were also determined. *Enterobacteriaceae* were found in human feces (1.20%) and river water (0.17%) but at very low rates in the other samples. *E. coli* was identified in the human sample (0.46%) and river water (0.02%), but *C. perfringens* and enterococci were not detected in any samples

Discussion

In this study, barcoded pyrosequencing targeting the V1-V3 hypervariable regions of the 16S rRNA gene was used to investigate microbial communities in major fecal pollution sources (beef cow, dairy cow, pig, and humans) and river water. Compared to traditional Sanger sequencing, pyrosequencing

can generate hundreds of thousands of sequences; however, the shorter reads require the choice of highly variable regions of the 16S rRNA gene and an appropriate algorithm to assign the sequences (Quince *et al.*, 2009). The V6 variable region has been selected by several researchers (Andersson *et al.*, 2008), partly because it is sufficiently short to be covered with 100 bp reads of the existing pyrosequencing technology. Because the average read length of pyrosequencing machines has increased to ~250 bp, increased selectivity in primer design is possible, allowing the possibility of targeting the longer variable region, V3 (Miller *et al*., 2009). In the present study, the median read length produced by pyrosequencing was about 480 bp, which may make it possible to more accurately assign 16S rRNA gene sequences to their phylogenetic affiliations.

The diversity of bacteria found in human feces was relatively low compared to those in the other hosts and river water. The human feces were dominated by a relatively small number of bacterial taxa, Bacteroidetes, Firmicutes, and Proteobacteria. It was not surprising that we found genus *Bifidobacterium* belonging to the *Actinobacteria* at a low rate only in the river water and not at all in the other samples, as it has been underestimated even in the metagenomic studies of microbial communities with universal primers targeting the 16S rRNA gene (Hill *et al.*, 2010). In addition, *Escherichia coli*, a traditional fecal indicator composing less than 1% of the intestinal microbial population (Drasar and Barrow, 1985), was detected in the human and river samples at very low rates, 0.46% and 0.02%, respectively. However, it was not found in the pig, beef cow or dairy cow feces, as was the case in a previous study (Dowd *et al*., 2008a).

Table 2 shows the relative genus distribution of the Bacteroidales group in the fecal and river water samples. The human feces also had a relatively low Bacteroidales group diversity and high Good's coverage compared to those of the other hosts. The beef cow, dairy cow, and pig feces had higher diversities compared to the human feces and most Bacteroidales sequences in the beef cow and dairy cow samples could not be assigned to a specific family. This could be explained by the fact that the information on the microbial community within the human gut has increased in recent years as a result of the development of culture-independent methods. In addition, the intestines of livestock are complex and diverse ecosystems for which data are insufficient (Dowd *et al*., 2008a). Interestingly, the river water had a somewhat different Bacteroidales group profile than did the fecal samples, in particular, for the genus *Paludibacter* (belonging to the family Porphyromonadaceae), which was absent from the human and pig samples but had the highest occurrence in the river water, as presented in Table 2.

Similarity tests among samples using community tree and Venn diagram analyses demonstrated that bacterial and Bacteroidales communities in the human were closer to those in the pig than to those in the other hosts. However, the Bacteroidales community of humans had a greater similarity to river water when compared according to the structure-based θ index (Yue and Clayton, 2005), placing more weight on species that are similar in abundance than for those with dissimilar abundance (Fig. 4B). In addition, a member of the Bacteroidales group, *Bacteroides vulgatus*, was detected only in human and river samples (Table 4), indicating that the river water was mainly contaminated from human sources.

Previous studies have proposed *Bacteroides* as a potential fecal indicator (Kreader, 1995). Culture-independent methods, such as T-RFLP (Terminal Restriction Fragment Length Polymorphism) and LH-PCR (Length Heterogeneity-PCR) have been used to identify 16S rRNA gene markers from fecal anaerobes (Bernhard and Field, 2000a). The *Bacteroides-Prevotella* 16S rRNA gene clone sequences of the fecal Bacteroidales group recovered from a variety of sources were compared to investigate the host-distribution pattern of the fecal Bacteroidales group (Dick *et al*., 2005). In addition, several host-specific primer sets were designed to identify and quantify fecal pollution sources (Layton *et al*., 2006; Kildare *et al*., 2007; Okabe *et al*., 2007). However, these studies were based on an incomplete understanding of the microbial community with regard to different fecal pollution sources. As a result, many previously developed Bacteroidales 16S rRNA gene markers did not have reliable specificities or sensitivities to discriminate fecal pollution sources (Ahmed *et al*., 2009; McLain *et al*., 2009). This could be attributed to the fact that Bacteroidales from different hosts were found in the same clade (Dick *et al*., 2005), and some real-time PCR assays may generate amplicons with the same primers and probe sequences but markedly different overall sequences (McLain *et al*., 2009). In addition, a relatively small number (<1,000) of sequences has previously been analyzed; in the present study, several thousands of Bacteroidales sequences were obtained and 500-800 OTUs were identified in the pyrosequencing analysis of fecal samples.

Culture-independent methods have identified several microbes that can serve as potential host-specific fecal indicators, including certain members of the order Bacteroidales, e.g., human-specific *Bacteroides thetaiotaomicron* (Carson *et al*., 2005), known as a predominant species in human feces (Kreader, 1995) and human-specific *Faecalibacterium* (Zheng *et al*., 2009). In this study, *B. thetaiotaomicron* was detected only in the human sample but at a very low rate (0.01%). Instead, *B. vulgatus* was a dominant species in the human feces (3.13%) and was also detected in the river water. In addition, the genus *Faecalibacterium* was detected in the human sample (10.89%) at a much higher rate than that in the pig (1.38%), or river water (0.04%) samples. However, the human-specific *Faecalibacterium* was not identified in this study. In fact, *Faecalibacterium prausnitzii* was detected both in the human and pig samples.

In addition to the Bacteroidales group (i.e., *B. vulgates*), we also found some host-specific microorganisms, including *Parasutterella excrementihominis*, *Veillonella* sp., *Dialister invisus*, *Megamonas funiformis*, and *Ruminococcus lactaris* in human feces and *Lactobacillus amylovorus* and *Atopostipes* sp. in pig feces. The most promising host-specific indicators, with the exceptions of *B. vulgatus* and *Dialister invisus*, were not detected in the river water in this study, but these could be used as host-specific fecal indicators in a PCR-based assay. For example, *Lactobacillus amylovorus* was not found in cow, sheep, or human samples, although it was previously tested as a pig-specific marker with a real-time PCR primer set developed by Konstantinov *et al*. (2005) (Marti *et al*., 2010). In addition, despite the substantial difference of the pig Bacteroidales group composition from those of the other hosts (Table 2), no pig-specific Bacteroidales marker was found in previous

studies (Jeong *et al*., 2008, 2010) and the previously developed assays lacked reliable specificities (Marti *et al*., 2010), with only one exception, Pig-2-Bac (Mieszkin *et al*., 2009).

The candidate organisms have not been suggested as host-specific indicators except for *Lactobacillus amylovorus*, but they have been generally detected or isolated in previous studies. These organisms include *Parasutterella excrementihominis* (Eckburg *et al*., 2005; Nagai *et al*., 2009), *Veillonella* sp. (Eckburg *et al*., 2005; Khoruts *et al*., 2010), *Dialister invisus* (Eckburg *et al*., 2005; Finegold *et al*., 2010), *Megamonas funiformis* (Sakon *et al*., 2008), *Ruminococcus lactaris* (Ballard *et al*., 2005) from human feces and *Atopostipes* sp. (Cotta *et al*., 2004) from swine manure. The criteria for fecal indicators include (i) sufficient density for detection, (ii) easy to enumerate, (iii) present in a constant ratio to pathogens, (iv) have a survival characteristic similar to the survival profiles of the pathogens whose presence it indicates, (v) unable to multiply in natural environments and (vi) released into the environment solely in the feces of warm-blooded animals. In addition, when using bacterial indicators to assess the magnitude of sewage input to a body of water or the distance of the sampling site from the source, the survival characteristics of the bacterial indicator must be known (Resnick and Levin, 1981). The suggested host-specific indicators in this study are expected to satisfy criteria (i), (v), and (vi), being similar to a conventional indicator because they are anaerobes and therefore cannot grow in the environment. However, further study is needed before these organisms can be used as host-specific indicators.

It may be possible to evaluate the specificity and coverage of existing host-specific primers using the pyrosequencing data in this study. For example, three markers, HF134, HF183 (Bernhard and Field, 2000b), and BacH (Reischer *et al*., 2007) were found to be human-specific and interestingly matched the sequence of *Bacteroides dorei* (96.8% similarity) in this study. In addition, two markers, PF163 (Dick *et al*., 2005) and Pig-2-Bac (Mieszkin *et al*., 2009), were found to be pig-specific, while Pig-1-Bac (Mieszkin *et al*., 2009) and Pig-Bac2 (Okabe *et al*., 2007) were shown to be matched to the sequences of non-specific hosts. Finally, the cow-specific markers CF128 and CF193 (Bernhard and Field, 2000b) did not correspond to any other sequence and BacR (Reischer *et al.*, 2006) was matched to the pig sequences in this study. Therefore, the large amount of sequence data produced by pyrosequencing could help to not only identify alternative potential indicators but also to evaluate existing host-specific markers. In addition, the suggested microorganisms could be used as multiple fecal indicators that are not dependent upon the absence or presence of a single indicator (i.e., the *Bacteroidales* 16S rRNA gene marker). Since the taxonomic profiles within different hosts may include specific signatures, these also could serve as composite fecal indicators. In addition, the monitoring of multiple indicators that are highly abundant and host-specific would greatly enhance fecal pollution source tracking (McLellan *et al*., 2009).

In this study, we collected sequence data for several thousands of microorganisms in fecal and river water samples. A better picture of the microbial community was obtained by comparing these bacterial and Bacteroidales groups in the fecal and river water samples. This approach allowed us to discover prospective indicators and molecular fingerprints for the fecal sources and river water instead of depending on a single indicator microorganism. Future research will be needed to verify the specificity and usefulness of alternative host-specific indicators identified in this study. Although the pyrosequencing technique is not yet quantitative, the lower cost, avoidance of cloning biases and high throughput appear to indicate the technology as the definitive method for fecal pollution source tracking

Acknowledgements

This study was supported by the Gyeonggi-do Institute of Health and Environment.

References

Ahmed, W., A. Goonetilleke, D. Powell, and T. Gardner. 2009. Evaluation of multiple sewage-associated bacteroides PCR markers for sewage pollution tracking. *Water Res.* 43, 4872-4877.

Andersson, A.F., M. Lindberg, H. Jakobsson, F. Backhed, P. Nyren, and L. Engstrand. 2008. Comparative analysis of human gut microbiota by barcoded pyrosequencing. *PLoS One* 3, e2836.

Ballard, S.A., E.A. Grabsch, P.D. Johnson, and M.L. Grayson. 2005. Comparison of three PCR primer sets for identification of *vanB* gene carriage in feces and correlation with carriage of vancomycin-resistant Enterococci: Interference by *vanB*-containing anaerobic Bacilli. *Antimicrob. Agents Chemother.* 49, 77-81.

Bernhard, A.E. and K.G. Field. 2000a. Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. *Appl. Environ. Microbiol.* 66, 1587-1594.

Bernhard, A.E. and K.G. Field. 2000b. A PCR assay to discriminate human and ruminant feces on the basis of host differences in bacteroides-prevotella genes encoding 16S rRNA. *Appl. Environ. Microbiol.* 66, 4571-4574.

Carson, C.A., J.M. Christiansen, H. Yampara-Iquise, V.W. Benson, C. Baffaut, J.V. Davis, R.R. Broz, W.B. Kurtz, W.M. Rogers, and W.H. Fales. 2005. Specificity of a bacteroides thetaiotaomicron marker for human feces. *Appl. Environ. Microbiol.* 71, 4945-4949.

Chun, J., J.H. Lee, Y. Jung, M. Kim, S. Kim, B.K. Kim, and Y.W. Lim. 2007. Eztaxon: A web-based tool for the identification of prokaryotes based on 16S ribosomal rRNA gene sequences. *Int. J. Syst. Evol. Microbiol.* 57, 2259-2261.

Claesson, M.J., O. O'Sullivan, Q. Wang, J. Nikkila, J.R. Marchesi, H. Smidt, W.M. de Vos, R.P. Ross, and P.W. O'Toole. 2009. Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PLoS One* 4, e6669.

Cotta, M.A., T.R. Whitehead, M.D. Collins, and P.A. Lawson. 2004. *Atopostipes suicloacale* gen. nov., sp. nov., isolated from an underground swine manure storage pit. *Anaerobe* 10, 191-195.

Dethlefsen, L., S. Huse, M.L. Sogin, and D.A. Relman. 2008. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol.* 6, e280.

Dick, L.K., A.E. Bernhard, T.J. Brodeur, J.W. Santo Domingo, J.M. Simpson, S.P. Walters, and K.G. Field. 2005. Host distributions of uncultivated fecal bacteroidales bacteria reveal genetic markers for fecal source identification. *Appl. Environ. Microbiol.* 71, 3184-3191.

Dowd, S.E., T.R. Callaway, R.D. Wolcott, Y. Sun, T. McKeehan, R.G. Hagevoort, and T.S. Edrington. 2008a. Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded flx amplicon pyrosequencing (btefap). *BMC Microbiol.* 8, 125.

Dowd, S.E., Y. Sun, R.D. Wolcott, A. Domingo, and J.A. Carroll. 2008b. Bacterial tag-encoded FLX amplicon pyrosequencing

(btefap) for microbiome studies: Bacterial diversity in the ileum of newly weaned salmonella-infected pigs. *Foodborne Pathog. Dis.* 5, 459-472.

Drasar, B.S. and P.A. Barrow. 1985. Intestinal microbiology wokingham, Van Nostrand Reinhold, UK.

Eckburg, P.B., E.M. Bik, C.N. Bernstein, E. Purdom, L. Dethlefsen, M. Sargent, S.R. Gill, K.E. Nelson, and D.A. Relman. 2005. Diversity of the human intestinal microbial flora. *Science* 308, 1635-1638.

Finegold, S.M., S.E. Dowd, V. Gontcharova, C. Liu, K.E. Henley, R.D. Wolcott, E. Youn, and *et al*. 2010. Pyrosequencing study of fecal microflora of autistic and control children. *Anaerobe* 16, 444-453.

Hill, J.E., W.M.U. Fernando, G.A. Zello, R.T. Tyler, W.J. Dahl, and A.G.V. Kessel. 2010. Improvement of the representation of bifidobacteria in fecal microbiota metagenomic libraries by application of the *cpn60* universal primer cocktail. *Appl. Environ. Microbiol.* 76, 4550-4552.

Jeong, J.Y., K.I. Gil, K.H. Lee, and J.O. Ka. 2008. Molecular identification of fecal pollution sources in water supplies by host-specific fecal DNA markers and terminal restriction fragment length polymorphism profiles of 16S rRNA gene. *J. Microbiol.* 46, 599-607.

Jeong, J.Y., H.D. Park, K.H. Lee, J.H. Hwang, and J.O. Ka. 2010. Quantitative analysis of human and cow-specific 16S rRNA gene markers for assessment of fecal pollution in river waters by real-time PCR. *J. Microbiol. Biotechnol.* 20, 245-253.

Khoruts, A., J. Dicksved, J.K. Jansson, and M.J. Sadowsky. 2010. Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent clostridium difficile-associated diarrhea. *J. Clin. Gastroenterol.* 44, 354-360.

Kildare, B.J., C.M. Leutenegger, B.S. McSwain, D.G. Bambic, V.B. Rajal, and S. Wuertz. 2007. 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal bacteroidales: A bayesian approach. *Water Res.* 41, 3701-3715.

Konstantinov, S.R., H. Smidt, and W.M. de Vos. 2005. Representational difference analysis and real-time PCR for strain-specific quantification of *Lactobacillus sobrius* sp. Nov. *Appl. Environ. Microbiol.* 71, 7578-7581.

Kreader, C.A. 1995. Design and evaluation of bacteroides DNA probes for the specific detection of human fecal pollution. *Appl. Environ. Microbiol.* 61, 1171-1179.

Layton, A., L. McKay, D. Williams, V. Garrett, R. Gentry, and G. Sayler. 2006. Development of bacteroides 16S rRNA gene taqman-based real-time PCR assays for estimation of total, human, and bovine fecal pollution in water. *Appl. Environ. Microbiol.* 72, 4214-4224.

Ley, R.E., M. Hamady, C. Lozupone, P.J. Turnbaugh, R.R. Ramey, J.S. Bircher, M.L. Schlegel, T.A. Tucker, M.D. Schrenzel, R. Knight, and J.I. Gordon. 2008. Evolution of mammals and their gut microbes. *Science* 320, 1647-1651.

Marti, R., P. Dabert, C. Ziebal, and A.M. Pourcher. 2010. Evaluation of *Lactobacillus sobrius*/*L. amylovorus* as a new microbial marker of pig manure. *Appl. Environ. Microbiol.* 76, 1456-1461.

McLain, J.E., H. Ryu, L. Kabiri-Badr, C.M. Rock, and M. Abbaszadegan. 2009. Lack of specificity for PCR assays targeting human bacteroides 16S rRNA gene: Cross-amplification with fish feces. *FEMS Microbiol. Lett.* 299, 38-43.

McLellan, S.L., S.M. Huse, S.R. Mueller-Spitz, E.N. Andreishcheva, and M.L. Sogin. 2009. Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. *Environ. Microbiol.* 12, 378-392.

Mieszkin, S., J.P. Furet, G. Corthier, and M. Gourmelon. 2009. Estimation of pig fecal contamination in a river catchment by real-time PCR using two pig-specific bacteroidales 16S rRNA genetic markers. *Appl. Environ. Microbiol.* 75, 3045-3054.

Miller, S.R., A.L. Strong, K.L. Jones, and M.C. Ungerer. 2009. Barcoded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in yellowstone national park. *Appl. Environ. Microbiol.* 75, 4565-4572.

Nagai, F., M. Morotomi, H. Sakon, and R. Tanaka. 2009. *Parasutterella excrementihominis* gen. Nov., sp. Nov., a member of the family alcaligenaceae isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* 59, 1793-1797.

Okabe, S., N. Okayama, O. Savichtcheva, and T. Ito. 2007. Quantification of host-specific bacteroides-prevotella 16S rRNA genetic markers for assessment of fecal pollution in freshwater. *Appl. Microbiol. Biotechnol.* 74, 890-901.

Quince, C., A. Lanzen, T.P. Curtis, R.J. Davenport, N. Hall, J.M. Head, L.F. Read, and W.T. Sloan. 2009. Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat. Methods* 6, 639-641.

Reischer, G.H., D.C. Kasper, R. Steinborn, A.H. Farnleitner, and R.L. Mach. 2007. A quantitative real-time PCR assay for the highly sensitive and specific detection of human faecal influence in spring water from a large alpine catchment area. *Lett. Appl. Microbiol.* 44, 351-356.

Reischer, G.H., D.C. Kasper, R. Steinborn, R.L. Mach, and A.H. Farnleitner. 2006. Quantitative PCR method for sensitive detection of ruminant fecal pollution in freshwater and evaluation of this method in alpine karstic regions. *Appl. Environ. Microbiol.* 72, 5610-5614.

Resnick, I.G. and M.A. Levin. 1981. Assessment of bifidobacteria as indicators of human fecal pollution. *Appl. Environ. Microbiol.* 42, 433-438.

Sakon, H., F. Nagai, M. Morotomi, and R. Tanaka. 2008. *Sutterella parvirubra* sp. nov. and *Megamonas funiformis* sp. nov., isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* 58, 970-975.

Sanapareddy, N., T.J. Hamp, L.C. Gonzalez, H.A. Hilger, A.A. Fodor, and S.M. Clinton. 2009. Molecular diversity of a north carolina wastewater treatment plant as revealed by pyrosequencing. *Appl. Environ. Microbiol.* 75, 1688-1696.

Schloss, P.D., S.L. Westcott, T. Ryabin, J.R. Hall, M. Hartmann, E.B. Hollister, R.A. Lesniewski, and *et al*. 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537-7541.

Tillett, D., D.L. Parker, and B.A. Neilan. 2001. Detection of toxigenicity by a probe for the microcystin synthetase a gene (*mcyA*) of the cyanobacterial genus microcystis: Comparison of toxicities with 16S rRNA and phycocyanin operon (phycocyanin intergenic spacer) phylogenies. *Appl. Environ. Microbiol.* 67, 2810-2818.

Walk, S.T., E.W. Alm, L.M. Calhoun, J.M. Mladonicky, and T.S. Whittam. 2007. Genetic diversity and population structure of *Escherichia coli* isolated from freshwater beaches. *Environ. Microbiol.* 9, 2274-2288.

Wang, Q., G.M. Garrity, J.M. Tiedje, and J.R. Cole. 2007. Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261-5267.

Yue, J.C. and M.K. Clayton. 2005. A similarity measure based on species proportions. *Comm. Statist. Theory Methods* 34, 2123-2131.

Zheng, G., H. Yampara-Iquise, J.E. Jones, and C. Andrew Carson. 2009. Development of *Faecalibacterium* 16S rRNA gene marker for identification of human faeces. *J. Appl. Microbiol.* 106, 634-641.