Published in final edited form as:
Environ Microbiol. 2010 February ; 12(2): 378–392. doi:10.1111/j.1462-2920.2009.02075.x.

Diversity and population structure of sewage derived microorganisms in wastewater treatment plant influent

S.L. McLellan[*,1], **S.M. Huse**[2], **S.R. Mueller-Spitz**[1], **E.N. Andreishcheva**[2], and **M.L. Sogin**[2]

[1]Great Lakes Water Institute, University of Wisconsin- Milwaukee, 600 E. Greenfield Ave, Milwaukee, WI 53204, USA.

[2]Josephine Bay Paul Center, Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543, USA.

Abstract

The release of untreated sewage introduces non-indigenous microbial populations of uncertain composition into surface waters. We used massively parallel 454 sequencing of hypervariable regions in rRNA genes to profile microbial communities from eight untreated sewage influent samples of two wastewater treatment plants (WWTP) in metropolitan Milwaukee. The sewage profiles included a discernable human fecal signature made up of several taxonomic groups including multiple *Bifidobacteriaceae*, *Coriobacteriaceae*, *Bacteroidaceae*, *Lachnospiraceae*, and *Ruminococcaceae* genera. The fecal signature made up a small fraction of the taxa present in sewage but the relative abundance of these sequence tags mirrored the population structures of human fecal samples. These genera were much more prevalent in the sewage influent than standard indicators species. High-abundance sequences from taxonomic groups within the *Beta*- and *Gammaproteobacteria* dominated the sewage samples but occurred at very low levels in fecal and surface water samples, suggesting that these organisms proliferate within the sewer system. Samples from Jones Island (JI – servicing residential plus a combined sewer system) and South Shore (SS – servicing a residential area) WWTPs had very consistent community profiles, with greater similarity between WWTPs on a given collection day than the same plant collected on different days. Rainfall increased influent flows at SS and JI WWTPs, and this corresponded to greater diversity in the community at both plants. Overall, the sewer system appears to be a defined environment with both infiltration of rainwater and stormwater inputs modulating community composition. Microbial sewage communities represent a combination of inputs from human fecal microbes and enrichment of specific microbes from the environment to form a unique population structure.

Introduction

Sewage overflows release human pathogens into surface waters, creating a serious public health risk and environmental concern. Older cities in the US, particularly in the northeast and Great Lakes region, have combined sewer systems that collect both wastewater and stormwater. These systems can exceed capacity following heavy rainfall events and subsequently overflow to surface waters. The USEPA estimates that communities with combined sewers annually release more than 3 trillion liters of sewage into the nation's waterways (USEPA, 2004). Human exposure to sewage contaminated water can occur through drinking water or recreational activities. Surveillance for fecal pollution relies upon the detection of indicator organisms such as *Clostridium perfringens*, enterococci, fecal coliforms, or *Escherichia coli* (USEPA, 2000; Fujioka, 2002; NRC, 2004). Selective

[*]Corresponding Author: Sandra L. McLellan, mclellan@uwm.edu, 414-382-1700.

culturing methods have made these assessments simple, relatively fast (1–2 days) and inexpensive (Fujioka, 2002). Epidemiological studies have shown a correlation between gastrointestinal illness and elevated fecal indicator organism levels, but sometimes the relationship between indicators and disease is poor or nonexistent (Dufour, 1984; Haile *et al.*, 1999; Wade *et al.*, 2006; Colford *et al.*, 2007). Fecal indicator bacteria can enter surface waters from agricultural runoff, urban stormwater, wildlife, and other non-human sources (Reeves *et al.*, 2004; Shanks *et al.*, 2006; Surbeck *et al.*, 2006; Vereen *et al.*, 2007; Lee *et al.*, 2008). Discerning the source of pollution is important for evaluating health risk and formulating mitigation strategies, and is an ongoing challenge facing water resource managers and public health officials (Field and Samadpour, 2007; Stewart *et al.*, 2008).

Alternative host-specific indicators might have the capacity to differentiate sewage from other sources of fecal pollution and provide a more accurate assessment of human health risk due to pathogens (Dick *et al.*, 2005; Savichtcheva and Okabe, 2006; Guzman *et al.*, 2007). Molecular-based methods have identified new microorganisms that can serve as host-specific indicators of fecal pollution including certain members of the order *Bacteroidales*, *Methanobrevibacter smithii*, and more recently *Faecalibacterium* sp. (Bernhard and Field, 2000a; Carson *et al.*, 2005; Dick *et al.*, 2005; Ufnar *et al.*, 2006; Lamendella *et al.*, 2007; Lee *et al.*, 2008; Layton *et al.*, 2009; Zheng *et al.*, 2009). Yet, these taxa represent only a small fraction of the diversity in human fecal microbial communities or in other host sources of fecal contamination. Published reports describing complex sewage communities have focused on microbial population structures of sludge and pilot scale bioreactors (Chouari *et al.*, 2005; Franklin and Mills, 2006) without performing detailed surveys of sewage influent entering treatment plants. Anticipated differences between raw sewage and activated sludge (Liu *et al.*, 2007) will likely reflect shifts in microbial community composition from original influent through aeration and anaerobic digestion processes. Detailed inventories of microbial populations in raw sewage have the potential to improve water management activities through the identification of novel indicator taxa for human fecal contamination.

Culture-independent, molecular surveys of PCR amplicons for ribosomal RNA (rRNA) genes have documented considerable microbial diversity in the human gut (Ley *et al.*, 2006a). Some of these microbes might serve as alternative indicator organisms. The most comprehensive studies of nearly full-length rRNA genes from human intestinal microbial flora report 13,000–18,000 sequences (Eckburg *et al.*, 2005; Ley *et al.*, 2006b), with one study noting great differences between microbial communities of mucosal surfaces and feces. Several smaller surveys of gut flora describe correlations with disease states (Hopkins *et al.*, 2001; Zoetendal *et al.*, 2002; Bibiloni *et al.*, 2006; Eckburg and Relman, 2007; Frank *et al.*, 2007), the establishment of gut flora (Favier *et al.*, 2003; Magne *et al.*, 2006), and effects of antibiotics (Shoemaker *et al.*, 2001). Person to person and temporal variability appear to be great (Eckburg *et al.*, 2005; Turnbaugh *et al.*, 2009). When these communities leave the gut, the members will vary in their ability to survive and persist because of differences in cellular physiology (e.g. ability to adapt to low nutrients, temperature, low iron, etc.) and physical cell structures (e.g. spore formers vs. gram negative organisms). The selective persistence of specific subpopulations from fecal-associated microbial communities may exert a significant influence on sewage microbial population structures including shifts in composition of high and low abundance taxa.

To our knowledge, there are no reports of large molecular surveys for microbial communities from untreated human sewage. Conventional DNA sequencing of a few thousand 16S rRNA gene PCR amplicons in an environmental DNA sample will generally fail to detect sequences from most of the low-abundance organisms in a sewage microbial community, some of which correspond to human fecal populations. In contrast, massively-parallel pyrosequencing (MPSS) technology (Margulies *et al.*, 2005) can generate hundreds

of thousands of short sequences (MPSS tags) from hypervariable regions in rRNA genes (Sogin *et al.*, 2006; Huber *et al.*, 2007) which increases the molecular sampling effort required to detect low abundance taxa. These tag sequences serve as proxies for individual operational taxonomic units (OTUs). The frequency of different rRNA gene tags provides a first-order description of the relative abundance (i.e. evenness) of phylotypes in a population. The variable nature of the MPSS tags and paucity of positions do not allow direct inference of phylogenetic frameworks. However, matches of tag sequences to a comprehensive reference database of hypervariable regions from full-length sequences of known phylotypes provide information about taxonomic identity and novel microbial diversity (i.e. richness) (Huse *et al.*, 2008). The technology has successfully documented shifts in human and mouse fecal associated microbial population structures in response to antibiotic treatment (Dethlefsen *et al.*, 2008; Antonopoulos *et al.*, 2009) and differences in microbial population structures between lean and obese twins and their mothers (Turnbaugh *et al.*, 2009). In this study we examined the microbial community at high resolution for complex sewage communities. Our study site represents a typical urban area. Two WWTPs service the Metropolitan Milwaukee area. One WWTP includes inputs from combined sewers, which collects both sanitary sewage and stormwater. The second WWTP primarily receives inputs from separated sewers that contain only sanitary sewage; stormwater discharges directly to rivers (untreated) in a separate system. Our primary objective is to gain insights about the composition of low-abundance and dominant populations in microbial communities released into the environment as a result of sewage overflows and to identify a suite of new alternative fecal indicators.

Results

Taxonomic composition of sewage, human fecal, and surface water microbial communities

We obtained WWTP influent samples of the two major facilities servicing Metropolitan Milwaukee, WI on multiple dates in 2005 and 2007 (Table 1). JI WWTP, located in downtown Milwaukee, receives sewage from the northern sector of the metropolitan area. The service area includes approximately 50 km^2 of combined sewers that consist of both sanitary sewage and stormwater from Milwaukee's densest urban area. SS WWTP services primarily residential areas of the metropolitan area. The surface water sample was taken from the junction of three major rivers leading to Lake Michigan (e.g. the estuary). These waters receive sewage overflows when they occur, but sample collection took place during base-flow conditions when minimal fecal pollution occurs.

Sequencing of WWTP influent samples yielded between 17,338 and 34,080 V6 region sequence tags for each sample (Table 1) with a total of 215,090 sequences (17,310 unique tags) recovered from all eight samples. The analysis of surface waters provided 38,368 sequences (4,603 unique tags). We compared these datasets to a total of 1,229,403 sequences (38,264 unique tags) collected from human fecal samples isolated from five individuals before, during and after treatment with antibiotics as well as samples from 11 families of lean and obese twins and their mothers (Dethlefsen *et al.*, 2008;Turnbaugh *et al.*, 2009). We used the GAST algorithm to assign taxonomy to individual tags (Huse *et al.*, 2008). The human and surface water samples exhibited very different taxonomic compositions (Fig. 1). *Proteobacteria*, specifically those that normally occur in the environment, dominated the sewage samples (59% of the total number of tags). *Gammaproteobacteria* accounted for 38% of the tags while 15% and 4% resolved to *Beta*- and *Epsilonproteobacteria*. Members of *Actinobacteria*, *Bacteroidetes*, and *Firmicutes* made up 37.5% of the sewage dataset, collectively (Fig.1), but accounted for 97% of the tags in the human fecal samples.

The sewage and human fecal samples contained many different *Firmicute* populations. Nearly 54% of the *Firmicute* tags in sewage samples mapped to *Clostridia* and 43% to *Bacilli*. In contrast, *Clostridia* tags comprised 98% of the *Firmicutes* in human fecal samples. The *Bacteroidia* accounted for the majority of the *Bacteroidetes* in sewage (54%), but we also detected *Flavobacteria*, *Sphingobacteria*, and several tags that did not resolve beyond the class level. *Bacteroidia* completely dominated the human fecal samples and accounted for 97% of the *Bacteroidetes*). The two dominant *Actinobacteria* families in the human fecal communities (*Bifidobacteriaceae* and *Coriobacteriaceae*) represented only 1.2% of the sewage tags.

The surface waters exhibited a taxonomic profile that differed from both human fecal and sewage samples. *Betaproteobacteria* (51%) and *Alphaproteobacteria* (19%) dominated the surface water microbial community. *Actinobacteria* of non-human origins and *Bacteroidetes*, mostly *Sphingobacteria*, accounted for 5.4% and 7.3% of the tags, respectively. Fewer than 0.5% of the surface water tags corresponded to *Firmicutes*.

Diversity and compositional overlaps of sequence tags

To estimate the diversity of the surface water, human fecal, and a pooled dataset of all sewage influent samples, we identified OTUs corresponding to 3 and 6 percent sequence divergence. The combined sewage and human fecal datasets were randomly subsampled to the size of the smallest dataset (i.e. surface water) to generate subsets of equivalent size (~38,000 tags). After using MUSCLE (Edgar, 2004) to align tags from these subsets, DOTUR (Schloss and Handelsman, 2005) identified clusters using the furthest neighbor algorithm that allowed 3% and 6% sequence variation and calculated a variety of diversity indices. We identified a total of 3045, 1950, and 2833 OTUs (3% DOTUR clusters) in sewage, human fecal, and surface water, respectively. Sewage had the highest richness, followed closely by surface water at both the 3% and 6% sequence variation criteria (Fig. 2). Human fecal samples displayed considerably lower richness, which is consistent with the assessments of overall taxonomic composition; sewage contains taxa common to humans in addition to many taxa of environmental origin. The ACE value at the 3% criteria for sewage was 5014 and 4946 for surface waters compared with 3081 for human fecal samples. Chao values were similar, with sewage, human fecal, and surface water estimated as 4911, 2908, and 4891, respectively.

The Venn diagram reports the number of shared OTUs among the different sample types (Fig. 3). The majority of OTUs (84%) represented community members that were specific to only one of the sewage, human fecal, or freshwater microbial communities. As expected, the surface water and human fecal communities had few OTUs in common (i.e. 27 shared OTUs). In contrast, the overlap for human fecal and sewage influent included 611 shared OTUs while sewage and surface water had 479 OTUs in common. The number of tags observed within the shared OTUs was disproportionate for the different microbial communities. For example, 89% of the human fecal tags occurred within the 611 OTUs shared with sewage, whereas only 31% of the sewage tags were within these shared OTUs (Table 2). Microbes of environmental origin account for the balance of the sewage tags. The sewage and human shared OTUs may include candidates for new indicator organisms.

We also examined species richness endemic to individual sewage samples. ACE richness estimates for individual samples ranged from 1878 to 3709 (Table 1), which is considerably lower than estimates for the composite dataset of 5014. Table 2 summarizes the OTUs shared among all samples, and among a subset of samples. The majority of the tag dataset (71%) occurred within the OTUs common to all eight samples. The sewage OTUs that overlapped with the human dataset generally appeared in seven or eight of the sewage

samples. OTUs that occurred in one sample at a very low abundance accounted for a large proportion of the richness (e.g., >50%), but a very minor portion of the tags.

Identifying human signature and candidate alternative indicators

Many of the tag sequences from sewage, fecal, and surface water samples can provide characteristic profiles describing the relative abundance for each distinct microbial population. By using the human fecal tag profile as a reference, the distribution of individual tags coupled to their taxonomic assignments can serve as a signature for sewage-derived fecal pollution. Comparisons of taxa assigned by GAST show there were 20 dominant human taxa that accounted for 94% of the human tags shared with sewage (Table 3), with other lower-abundance taxa accounting for the remaining overlap. The 20 most dominant taxa in sewage include tag sequences for the human-associated taxa *Bacteroides*, *Lachnospiraceae*, and *Faecalibacterium*. *Bifidobacterium* was also discernable in sewage profiles (Table 4). However, a greater proportion of the sewage tags belonged to taxonomic groups within *Gammaproteobacteria* (i.e. *Acinetobacter*, *Aeromonas*, and *Pseudomonas*) in addition to *Arcobacter* (*Epsilonproteobacteria*), and *Trichococcus* (*Bacilli*). Tag sequences for these taxonomic groups occurred at very low frequencies in surface waters.

There were a total of 57,020 unique V6 tags in the combined datasets from sewage, human and surface waters. We examined the taxonomic profiles for the 4000 most abundant V6 tags (e.g. at the sequence level) in all samples combined. Fig. 4 shows tag profiles for human taxa that are prominent in sewage. Each panel represents 401 distinct tag sequences initially sorted by taxonomic assignment and then by decreasing tag frequency. The examination of the 4000 most abundant tag sequences captured the similarities and differences in the tag profiles for *Bacteroidales* (Fig. 4a) and *Firmicutes* (Fig. 4b and c). For the entire dataset of 57,020 unique tags, 3172 resolved to *Bacteroides* with the majority of these representing low abundance taxa in sewage, humans or surface water. Eighty-eight unique *Bacteroides* sequences accounted for nearly 85% of the total *Bacteroides* tags (shown in Fig. 4a). The remaining 15% of the *Bacteroides* tags distributed evenly at low levels across ~3100 *Bacteroides* tag sequences (data not shown). *Lachnospiraceae* and *Ruminococcaceae*, two other families of interest, had similar numbers of unique tags and distributions.

There were similarities in the relative abundance of tag sequences for specific populations within *Bacteroidales* and the *Firmicutes* in the human fecal and all of sewage samples. Many of these tag sequences were absent from surface water and therefore may represent good candidates for alternative indicator organisms. The eight WWTP samples had very similar abundance profiles of human-associated taxa (e.g. *Lachnospiraceae* in Fig. 4c) as well as taxa of environmental origin (e.g. *Flavobacteriaceae* in Fig. 4a). Human fecal and sewage microbial communities shared many high-abundance *Bacteroides* tags (Fig. 4a), including some that resolved to the species level, e.g. *B. fragilis*, *B. massiliensis*, and *B. ovatus*. We also observed a number of low-abundance, shared sequences from fecal and sewage samples but many did not resolve beyond the order level. Within the family *Porphyromonadaceae*, the surface waters contained some low-abundance tags that mapped to *Paludibacter*, while a single high-abundance *Paludibacter* tag occurred only in sewage (Fig. 4a taxa designation 215). Not all *Bacteroidales* tags in sewage appeared in human fecal samples. For example, several high-abundance *Parabacteroides* tags (Fig. 4a taxa designations 241–247) appeared in sewage but only one also occurred in high abundance in humans (Fig. 4a taxa designation 243), and other *Parabacteroides* tags were absent or in low levels. Similarly, sewage had multiple tags that mapped to *Prevotella* (*Prevotellaceae*) whereas humans contained a single example. The presence of tags from several closely related *Bacteroidales* taxa in sewage and their absence in humans suggest sewage contains

non-human fecal sources or there may be geographic specificity of human derived *Bacteroidales* within these taxa.

The *Firmicutes* provided the richest profile of tags that were common to humans and sewage (Fig. 4b and c). Within *Clostridiales*, few tags belonged to *Clostridiaceae*. *Lachnospiraceae* tags were the most common; the majority of these resolved only to the family level. The genera *Dorea*, *Lachnospira*, and *Roseburia* all contained high abundance tags common to human and sewage. There were a few examples of low abundance tags (e.g. <200 in the normalized dataset) within *Dorea* that were absent in sewage. Tags within the family *Ruminococcaceae* followed a similar pattern; many tags resolved to the family level. *Faecalibacterium*, *Ruminococcus*, and *Subdoligranulum* tags were predominate features in humans and proportionally reflected in sewage.

V6 tags associated with standard fecal indicators in WWTP

In examining the V6 tags recovered from eight WWTP samples, we rarely encountered the standard fecal indicator organisms. Overall, only two tags out of the entire dataset of WWTP samples (215,090 tags) resolved to *E. coli*. Only three tags mapped to *Enterococcus faecalis* (n=1) or *E. faecium* (n=2) and no tags matched *C. perfringens*. A greater number of V6 tags matched at least one reference sequence from an indicator organism, but they resolved to higher taxonomic levels (see Methods on annotation criteria). Even when considering tags that resolved to more general taxonomic designations, standard indicators rarely occurred in the WWTP samples.

The relative abundance of standard indicators was also determined using cultivation plus direct enumeration of total cell numbers. The total cell counts for sewage were approximately six orders of magnitude greater than culturable *E. coli and Enterococcus* levels, and nearly seven orders of magnitude greater than *C. perfringens* levels (Table 5). Overall, determinations made by culture combined with direct counts suggested that indicators were even more rare than tag sequencing would suggest. Several factors could contribute to this variance. *E. coli* and enterococci (*E. faecalis* and *E. faecium*) have seven and four copies of the 16S rRNA gene respectively, which is comparable to the 4.15 average for the 904 Bacterial genomes (Lee *et al.*, 2009) and would not account for the higher estimation of relative abundance using the tag sequencing compared with the direct counts plus culture methods. Enumeration of indicator organisms by cultivation will not detect the presence of non-viable cells, but direct staining and tag sequencing will detect these cells, which most likely accounts for the estimates of relatively low abundance of indicator organisms using direct staining plus culture methods. In addition, total cell counts will include Archaea that molecular surveys of Bacteria may not detect, which would further lower the ratio of cultured indicator cells compared to direct counts. Overall, these results demonstrate that standard fecal indicators are very minor populations in sewage.

Stability of sewage signal and comparison of two different WWTPs

Analysis of eight WWTP samples over a two and a half year period revealed that the microbial community composition is very stable. The same human taxa dominated each of the eight samples (Fig. 3). The profiles of individual tag sequences for the eight samples were also very similar (Fig. 4). Comparisons of microbial profiles show greater similarity amongst samples collected at the two treatment plants on the same day than samples from the same WWTP collected on different days (Fig. 5). The relative proportion of human-associated taxa in individual samples ranged from 13% to 28% of the total tags. There was no significant difference between SS and JI WWTPs, despite the fact that SS receives primarily sanitary sewage and JI receives both sanitary sewage and stormwater (e.g. combined sewage). Further, rain events did not correlate with changes in the proportion of

human-associated taxa at either WWTP. However, on one date (21 Aug 2007) with 3.38 cm of rainfall, influent flow increased at both WWTPs, and this corresponded to higher diversity in the JI and SS communities (Table 1). This suggests that rainfall modulates the community throughout the sewer system and may have an equal or even greater influence on composition than direct stormwater inputs into the combined sewer system.

Discussion

Sewage overflows release a complex array of microorganisms into the environment, including pathogenic bacteria, viruses, and protozoa. The presence of fecal bacteria is a major cause of water quality impairment in the nation's waterways and coastal regions (USEPA, 2006; Dorfman and Rosselot, 2008; Stewart *et al.*, 2008; USEPA, 2009); however, the source of fecal pollution is unknown in the majority of cases. Fecal pollution may be derived from humans, agricultural or domestic animals, or wildlife. It is necessary to differentiate between human and other sources of fecal contamination for accurate assessments of risk to public health and prioritization of mitigation efforts.

Standard water quality measurements rely upon culture of bacteria and are not specific to a host source. These culturing methods rely upon biochemical reactions on media such as modified membrane-Thermotolerant *Escherichia coli* (modified m-TEC) agar (USEPA, 2002), membrane-*Enterococcus* Indoxyl-β-D-Glucoside (mEI) agar (USEPA, 2000), and membrane-*Clostridium perfringens* (mCP) agar (Emerson and Cabelli, 1982). Culture methods limit the range of organisms that can serve as indicators, and this study illustrates that fecal indicators represent minor populations in water quality assessments that have major regulatory and economic consequences. More recently, molecular techniques have expanded the range of organisms that can be used for fecal indicators (Field et al., 2003), however, these methods are based upon the presence or absence of a single target, and have been developed without a comprehensive view of the complex microbial community associated with different fecal pollution sources.

Monitoring for multiple organisms that are in higher abundance and specific to human fecal contamination would greatly enhance water quality assessment capabilities. MPSS allowed the sequencing of many thousands of PCR amplicons for homologous hypervariable rRNA coding regions from a wide range of taxa in sewage and surface water (~230,000 tag sequences), which were compared to human fecal samples (~1.2M tag sequences). We were able to describe radically different microbial communities in humans, sewage and surface water and identify specific taxa shared among sewage and humans. Within named taxa, we observed further resolution by cataloguing the collection of sequences that mapped to the same taxonomic levels. For example, sewage contained between 2000–3000 unique V6 sequences for the taxonomic groups *Bacteroides*, *Lachnospiraceae*, and *Faecalibacterium* (Table 3); however, fewer than twenty unique V6 sequences accounted for the majority of tags for each of these taxonomic groups. This illustrates the high resolution that can be achieved with MPSS. The occurrence of shared sequences, their taxonomic identity, and their relative abundance in sewage from a large metropolitan area and in human fecal samples provides a complex signature of human fecal pollution.

The complex signatures determined by MPSS contained few tags that corresponded to traditional indicator organisms. The reference database used for V6 tag annotation contained several distinct sequences each for *E. coli*, *E. faecalis*, *E. faecium*, and *C. perfringens*, as well as other sequences that do not differentiate between genera within a family. For example, one V6 tag sequence matched reference sequences for various genera of *Enterobacteriaceae* (primarily *Klebsiella* and *Shigella*) as well as of *E. coli*. Even if we included these less specific V6 sequences, indicator organisms were extremely rare in the

sewage samples (<0.7%) of the total tags). This compares to cultivation experiments that detected ~ 2 indicator organisms/million total cells.

MPSS sequence tags as proxies for alternative indicator organisms

The *Bacteroidales* and *Clostridiales* groups exhibited high levels of diversity in both the sewage and human fecal communities. Previous reports have proposed *Bacteroides* as an alternative indicator (Kreader, 1995) and characterization of the broader group of *Bacteroidales* in multiple hosts has identified phylotypes that demonstrate host specificity for humans or other animals (Dick *et al.*, 2005; Layton *et al.*, 2006; Okabe *et al.*, 2007; Lamendella *et al.*, 2007). Bernhard and Field first described a human-specific *Bacteroides* marker (Bernhard and Field, 2000b), which has been used in numerous field studies to detect human sources of fecal pollution (Bernhard and Field, 2000b; Bower *et al.*, 2005; Okabe *et al.*, 2007; Santoro and Boehm, 2007). However, there is no clear host-specific structure across the *Bacteroidales* group as a whole; *Bacteroidales* from different host animals are found within the same clade (Dick *et al.*, 2005; Lamendella *et al.*, 2007; Jeter *et al.*, 2009). These assessments of *Bacteroidales* populations have relied upon comparisons between a relatively small number (97–1200) of rRNA sequences (Dick *et al.*, 2005; Lamendella *et al.*, 2007; Jeter *et al.*, 2009). Other studies have identified a *Faecalibacterium* phylotype using subtractive hybridization that appears to be specific to humans (Zheng *et al.*, 2009). All of these studies have focused on identification of a marker that can be used to detect the presence of fecal pollution from a human or animal source.

Our approach is more comprehensive than identifying a single marker. We detected thousands of *Bacteroidales*; however, a relatively small subset occurred at a high frequency (e.g. 84% of the *Bacteroides* were represented in the 88 *Bacteroides* tags shown in Fig. 4a). We found relative abundance patterns of these tags in sewage mirrored that of human fecal communities. Taxonomic group profiles within different hosts may contain unique signatures, and these might serve as complex indicators of fecal pollution versus reliance on the presence or absence of a single phylotype. Evaluating host sources by detecting signatures based on multiple taxa (which may co-occur because of their presence in the host) and their relative abundance would not be influenced by transient colonization, or low levels of cross colonization to other hosts. Utilization of a single marker requires a high level of host specificity, and any cross over between hosts confounds assessments. A similar approach to discriminating fecal pollution from humans, pigs, and cattle has utilized qPCR to detect the relative abundance of dominant bacterial groups: *Bacteroides*/*Prevotella*, *Clostridium coccoides*, and *Bifidobacterium* (Furet *et al.*, 2009), however the qPCR experimental design did not target specific populations within these groups. Employing community profiling by 454 sequencing enables high resolution within each of these taxa of interest (Fig. 4).

MPSS tags of environmental origin in sewage and comparison with river water

A large majority of the sewage tags were not shared with human fecal material and belonged to taxonomic groups that are generally not considered human comensual organisms. This finding is not surprising considering the nature of the sewer environment, which contains high nutrient levels in an aqueous state. Previous studies have shown that this environment is rich in biofilms (Chen *et al.*, 2003; Leung *et al.*, 2005; Ort and Gujer, 2008), which may act as a reservoir for resident organisms. Several taxa, including *Acinetobacter*, *Aeromonas*, *Pseudomonas*, and *Trichococcus*, were highly abundant in sewage, but not present in surface water, or only present at very low levels, suggesting these organisms may be specific to the sewer environment. Only one or two specific V6 sequences represented each of these high abundance taxa, which further suggests that these organisms comprise a relatively narrow population that could be propagating within the sewer system itself. The two WWTPs are

relatively isolated from one another, so the observation of the same high abundance environmental organisms in JI and SS WWTP samples on the same day and over time would indicate that the sewer system behaves as a defined environment and the sewage community is subjected to a common ecological forcing throughout the system.

Alternative indicators may also include taxa that may be resident organisms growing in the sewer system, but not associated with human fecal material. *Acinetobacter*, *Aeromonas*, and *Trichococcus* might be useful adjuncts to the fecal specific organisms since they occur in very high abundance. Both *Aeromonas* and *Arcobacter* have been associated with sewage-contaminated water (Collado *et al.*, 2008; Khan *et al.*, 2009). Concurrent detection of human signature organisms with the sewage specific organisms would provide an additional level of confidence for assessing sources of fecal pollution. Further work is necessary to determine if these high abundance taxa are specific to the sewer system, or if they propagate in other environments, particularly those associated with fecal pollution such as contaminated soils, sediments, or beach sand. Some of these environments also appear to support the growth of indicator bacteria (Alm *et al.*, 2003; Whitman and Nevers, 2003; Byappanahalli and Fujioka, 2004; Ishii *et al.*, 2006).

There is a high potential to be able to track signature taxa from sewage in surface water during contaminant events. The surface water we analyzed was from the junction the three main rivers (freshwater estuary) that lead to Lake Michigan, which serves as the receiving waters for sewage overflows in the Milwaukee area. The freshwater estuary system is highly diverse and contains microorganisms originating from the rivers, as well as Lake Michigan (Mueller-Spitz *et al.*, 2009). We found very little overlap between sewage and the estuary surface water community profiles. Taxa found in humans were not detected. The taxa that appeared to be derived from the sewage environment were nearly absent in surface water. Sewage and surface water both contained *Betaproteobacteria*, primarily members of *Comamonadaceae*, but the genera composition was completely different between the two sample types: sewage contained *Acidovorax*, *Simplicispira*, and *Polaromonas* while surface water contained *Curvibacter*, *Malikia*, and *Variovorax*.

Temporal factors are the major influence on sewage communities

The JI WWTP services both separated and combined sewer systems, therefore, sewage influent contains human fecal microbial populations from households and non-human fecal pollution from stormwater runoff during rain events. The SS WWTP services primarily separated sewer systems and its influent contains fecal populations mainly from human sources. We hypothesized there would be significant differences between the JI and SS microbial communities and that the variance would increase during rain events when non-human fecal material might enter the JI system. Surprisingly, WWTP samples were most similar on the same collection date, and rainfall did not affect the similarity. This is surprising given that very large amounts of stormwater entered the combined sewer system on 21 Aug 2007; stormwater from 50 km^2 of an urbanized area is conveyed to JI through this system. However, SS WWTP also was found to have high flows into the plant. Significant infiltration of rainwater through leaking or cracked sewer pipes occurs in the Milwaukee system as well as some direct inflow through openings in sewer manhole covers. The high diversity in both JI and SS samples suggests that additional community members enter the system during this time, such as soil bacteria from direct runoff or through infiltration. Further sampling would be necessary to assess if there are seasonal patterns in sewage communities, and discern specific taxa that correlate with sewer system configuration or environmental parameters. Such information will be important to establish which taxa provide a consistent signature that can be used to track sewage contaminated water.

Conclusions

The 454 sequencing platform allows us to collect sequence information for thousands of microbial organisms in sewage, including human commensals and closely related variants. This approach would completely redefine how we identify new indicators and track pollution. Instead of looking for a single indicator organism, suites of alternative indicators, and their relative abundance patterns could be determined in sewage and subsequently tracked in contaminated water. Using DNA-based detection methods, it is possible to test for a broader range of organisms, including fecal anaerobes, and to target multiple phylotypes within these groups that may be associated with a specific source of fecal pollution. These multi-taxa signatures could also be used to differentiate between human and non-human fecal sources. We identified tags shared between human feces and sewage, which are potentially new alternative indicators. New data generated as part of the human microbiome project, along with comparisons of microbial communities from other animals (Ley *et al.*, 2008) will increase and refine the database of human fecal pollution signatures. This approach provides new insight into the composition and complexity of sewage-derived microorganisms.

Experimental Procedures

Sewage treatment plant influent sample collection

We obtained WWTP influent samples of the two major facilities servicing Metropolitan Milwaukee, WI on multiple dates in 2005 and 2007 (Table 1). The entire service area is 1065 km^2. JI WWTP is located in downtown Milwaukee and receives sewage from 122 km of metropolitan interceptor sewer (MIS), the northern sector of the metropolitan area. A portion of the city serviced by JI WWTP (50 km^2) is comprised of combined sewers that consist of both sanitary sewage and stormwater from Milwaukee's most densely urban area. SS WWTP services primarily residential areas of the metropolitan area and receives sewage from 201 km of MIS. An additional 143 km of MIS can be diverted to either SS or JI. Samples consisted of 1 L of 24-hour flow-weighted samples collected from 6 am on the preceding day until 6 am on the stated collection day. Flow into the WWTP was averaged between measurements taken at the 6 am time points each day. Sewage samples were filtered through a 47 mm, 0.45 μm nitrocellulose membrane (Millipore, Billerica, MA) at 100 ml or until the filter clogged. Surface water was collected at the Milwaukee estuary on 19 June 2007 using a 5L grab sampler. Three samples were pooled prior to filtering 200 ml as described above. All samples were stored at −80°C until processing for DNA extraction. DNA extractions were carried out using the Fast DNA spin kit for Soil (MP Biomedicals, Solon, OH).

Detection of fecal indicators and direct counts for WWTP influent samples

Sewage samples were examined for culturable *E. coli*, enterococci, and *C. perfringens* levels to compare with total cell counts. Sewage influent samples were diluted to a range of 10^{-3} to 10^{-5} and 1 ml was filtered through a 0.45 μm pore size 47 mm nitrocellulose filter. Filters were placed on modified m-TEC (Difco, Sparkes, MD) for *E. coli* or mEI (Difco, Sparkes, MD) agar according to the US Environmental Protection Agency (USEPA) methods for *E. coli* and enteroocci enumeration (USEPA, 2000). Levels of *C. perfringens* were enumerated using membrane filtration and mCP media was prepared according to previously published methods (Armon and Payment, 1988). After anaerobic incubation, mCP plates were exposed to vapors of ammonium hydroxide for 20−30 sec and the straw colored colonies that turned pink to red were counted as *C. perfringens*. The direct counts were preformed on sewage following the acridine orange method presented by Jjemba *et al.* (2006). Both 10^{-3} to 10^{-4} dilutions were made from sewage influent. Twenty fields per slide were counted.

454 pyrosequencing

We used a Roche Genome Sequencer FLX System (GS-FLX) to characterize populations of PCR amplicons from hypervariable regions in rRNA genes. A cocktail of five fused primers at the 5' end of the V6 region (*E. coli* positions 967–985) and four primers at the 3' end (*E. coli* positions 1046-1028) captured the full diversity of rRNA sequences represented in molecular databases (Sogin *et al.*, 2006; Huber *et al.*, 2007). The 5' fused primer includes a 5 nt "key" inserted between the Life Sciences primer A and the 967F rRNA primer. The 5 nt key is permuted for each sample and permits the bioinformatics binning of tags according to sample source. We prepared amplicon libraries from at least three independent PCR cocktails for each sample to minimize the impact of potential early-round PCR errors. After passage through a Qiagen MinElute PCR purification kit (Qiagen, Valencia, CA) and monitoring on an Agilent Bioanalyzer 2100, the Life Science A primer sequence on each fragment anneals to complementary oligonucleotides that are tethered onto micron-scale beads. The annealing conditions favor one fragment per bead. The beads were emulsified in a PCR mixture-in-oil and PCR amplification occurs in micro-reactors generating ~ten million copies of a unique DNA template. After breaking the emulsion, the DNA strands are denatured and beads carrying single-stranded DNA clones are deposited into wells on one of the two regions of the PicoTiterPlate™ and sequenced on the GS-FLX, providing approximately 400,000 reads per run.

Quality trimming and taxonomic assignments

We trimmed the resulting sequences of their 5 nt keys and primers then filtered out as low-quality all reads without an exact match to the forward primer, without a recognizable reverse primer, with fewer than 50 bases, or with any ambiguous base calls (Ns) (Huse *et al.*, 2007). Reads were binned to samples using the 5 nt keys. We used the rRNA indexing algorithm Global Assignment of Sequence Taxonomy (GAST) (Huse *et al.*, 2008) for comparing tag sequences to known rRNA genes that have already been placed in a phylogenetic framework of more than 500,000 nearly full length reference SSU rRNA genes (RefSSU) based on the SILVA database (Pruesse *et al.*, 2007). GAST compares each tag against a database of V6 sequences excised from the RefSSU, selects the nearest reference V6 match or matches, and assigns taxonomy to the tag using a consensus of the taxonomy assigned to all RefSSU sequences containing that V6 region. GAST is more computationally intensive than relying upon BLAST score ratios (Rasko *et al.*, 2005) but results in a better match in 5–10% of comparisons (Huse *et al.*, 2008).

Creating OTU clusters and calculating richness

To create OTUs across the three environments (sewage, human, surface water), we randomly selected 38,368 sequences from each (the number of sequences in the smallest set, surface water). These sequences were combined, duplicates removed, and aligned with MUSCLE (with the –diags and –maxiters 2 options) (Edgar, 2004) and all pairwise sequence distances were calculated with quickdist (Sogin *et al.*, 2006). We used DOTUR (Schloss and Handelsman, 2005) to create furthest neighbor clusters at 3%, 6%, and 10% pairwise distances, all duplicate sequences were reassigned to their corresponding clusters. DOTUR output also calculates richness estimates (ACE and Chao) and rarefaction. For richness estimates of individual datasets, we likewise subsampled to the smallest set (17300 from JI 20 April 2005), created a multiple sequence alignment with MUSCLE on the unique sequences, clustered with DOTUR, reassigned all duplicate sequences to the clusters and used DOTUR to estimate the richness.

The OTU overlap among samples was plotted using The Venn Diagram Plotter provided by Pacific Northwest National Laboratory, Department of Energy, which is made available OMICS.PNL.GOV

We calculated the Bray-Curtis distances between WWTP samples based on the GAST taxonomic assignments made to each sample using the VAMPS website (http://vamps.mbl.edu). A similarity matrix of these distances was used to construct a UPGMA tree using the R package (Paradis *et al.*, 2004). Differences in human taxa abundance in WWTP samples were evaluated using a student's t-test with a significance level of $p \leq 0.05$.

Sequences were deposited in GenBank under the study accession ID SRP000905.

Acknowledgments

This work was funded by NIH grant 1 R21 AI076970-01A1, and by NSF/OCE grant 0430724.

References

Alm EW, Burke J, Spain A. Fecal indicator bacteria are abundant in wet sand at freshwater beaches. Wat Res. 2003; 37:3978–3982.

Antonopoulos DA, Huse SM, Morrison HG, Schmidt TM, Sogin ML, Young VB. Reproducible community dynamics of the gastrointestinal microbiota following antibiotic perturbation. Infect Immun. 2009; 77:2367–2375. [PubMed: 19307217]

Armon R, Payment P. A modified m-CP medium for enumerating *Clostridium perfringens* from water samples. Can J Microbiol. 1988; 34:78–79. [PubMed: 2897874]

Bernhard AE, Field KG. 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. 2000a; 66:1587–1594. [PubMed: 10742246]

Bernhard AE, Field KG. 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. 2000b; 66:4571–4574. [PubMed: 11010920]

Bibiloni R, Mangold M, Madsen KL, Fedorak RN, Tannock GW. The bacteriology of biopsies differs between newly diagnosed, untreated, Crohn's disease and ulcerative colitis patients. J Med Microbiol. 2006; 55:1141–1149. [PubMed: 16849736]

Bower PA, Scopel CO, Jensen ET, Depas MM, McLellan SL. Detection of genetic markers of fecal indicator bacteria in Lake Michigan and determination of their relationship to *Escherichia coli* densities using standard microbiological methods. Appl Environ Microbiol. 2005; 71:8305–8313. [PubMed: 16332817]

Byappanahalli M, Fujioka R. Indigenous soil bacteria and low moisture may limit but allow faecal bacteria to multiply and become a minor population in tropical soils. Water Sci Technol. 2004; 50:27–32. [PubMed: 15318482]

Carson CA, Christiansen JM, Yampara-Iquise H, Benson VW, Baffaut C, Davis JV, et al. Specificity of a *Bacteroides thetaiotaomicron* marker for human feces. Appl Environ Microbiol. 2005; 71:4945–4949. [PubMed: 16085903]

Chen GH, Leung DH, Hung JC. Biofilm in the sediment phase of a sanitary gravity sewer. Water Res. 2003; 37:2784–2788. [PubMed: 12753857]

Chouari R, Le Paslier D, Daegelen P, Ginestet P, Weissenbach J, Sghir A. Novel predominant archaeal and bacterial groups revealed by molecular analysis of an anaerobic sludge digester. Environ Microbiol. 2005; 7:1104–1115. [PubMed: 16011748]

Colford JM Jr, Wade TJ, Schiff KC, Wright CC, Griffith JF, Sandhu SK, et al. Water quality indicators and the risk of illness at beaches with nonpoint sources of fecal contamination. Epidemiology. 2007; 18:27–35. [PubMed: 17149140]

Collado L, Inza I, Guarro J, Figueras MJ. Presence of *Arcobacter* spp. in environmental waters correlates with high levels of fecal pollution. Environ Microbiol. 2008; 10:1635–1640. [PubMed: 18215159]

Dethlefsen L, Huse S, Sogin ML, Relman DA. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. PLoS Biol. 2008; 6:e280. [PubMed: 19018661]

Dick LK, Bernhard AE, Brodeur TJ, Santo Domingo JW, Simpson JM, Walters SP, Field KG. Host distributions of uncultivated fecal *Bacteroidales* bacteria reveal genetic markers for fecal source identification. Appl Environ Microbiol. 2005; 71:3184–3191. [PubMed: 15933020]

Dorfman, M.; Rosselot, KS. A Guide to Water Quality at Vacation Beaches. Natural Resources Defense Council; 2008. Testing the Waters.

Dufour AP. Bacterial indicators of recreational water quality. Can J Public Health. 1984; 75:49–56. [PubMed: 6367923]

Eckburg PB, Relman DA. The role of microbes in Crohn's disease. Clin Infect Dis. 2007; 44:256–262. [PubMed: 17173227]

Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, et al. Diversity of the human intestinal microbial flora. Science. 2005; 308:1635–1638. [PubMed: 15831718]

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32:1792–1797. [PubMed: 15034147]

Emerson DJ, Cabelli VJ. Extraction of *Clostridium perfringens* spores from bottom sediment sample. Appl Environ Microbiol. 1982; 44:1144–1149. [PubMed: 6295278]

Favier CF, de Vos WM, Akkermans AD. Development of bacterial and bifidobacterial communities in feces of newborn babies. Anaerobe. 2003; 9:219–229. [PubMed: 16887708]

Field KG, Samadpour M. Fecal source tracking, the indicator paradigm, and managing water quality. Water Res. 2007; 41:3517–3538. [PubMed: 17643471]

Field KG, Bernhard AE, Brodeur TJ. Molecular approaches to microbiological monitoring: fecal source detection. Environ Monit Assess. 2003; 81:313–326. [PubMed: 12620024]

Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci U S A. 2007; 104:13780–13785. [PubMed: 17699621]

Franklin RB, Mills AL. Structural and functional responses of a sewage microbial community to dilution-induced reductions in diversity. Microbiol Ecol. 2006; 52:280–288.

Fujioka, RS. Microbial indicators of water quality. In: Hurst, CJ.; Crawfod, RL.; Knudsen, GR.; McInerney, MJ.; Stetzenbach, LD., editors. Manual of Environmental Microbiology. Washington, D. C.: ASM Press; 2002. p. 234-243.

Furet JP, Firmesse O, Gourmelon M, Bridonneau C, Tap J, Mondot S, et al. Comparative assessment of human and farm animal faecal microbiota using real-time quantitative PCR. FEMS Microbiol Ecol. 2009; 68:353–362.

Guzman C, Jofre J, Montemayor M, Lucena F. Occurrence and levels of indicators and selected pathogens in different sludges and biosolids. J Appl Microbiol. 2007; 103:2420–2429. [PubMed: 18045427]

Haile RW, Witte JS, Gold M, Cressey R, McGee C, Millikan RC, et al. The health effects of swimming in ocean water contaminated by storm drain runoff. Epidemiology. 1999; 10:355–363. [PubMed: 10401868]

Hopkins MJ, Sharp R, Macfarlane GT. Age and disease related changes in intestinal bacterial populations assessed by cell culture, 16S rRNA abundance, and community cellular fatty acid profiles. Gut. 2001; 48:198–205. [PubMed: 11156640]

Huber JA, Welch DB, Morrison HG, Huse SM, Neal PR, Butterfield DA, Sogin ML. Microbial population structures in the deep marine biosphere. Science. 2007; 318:97–100. [PubMed: 17916733]

Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol. 2007; 8:R143. [PubMed: 17659080]

Huse SM, Dethlefsen L, Huber JA, Welch DM, Relman DA, Sogin ML. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. PLoS Genet. 2008; 4 e1000255.

Ishii S, Ksoll WB, Hicks RE, Sadowsky MJ. Presence and growth of naturalized *Escherichia coli* in temperate soils from Lake Superior watersheds. Appl Environ Microbiol. 2006; 72:612–621. [PubMed: 16391098]

Jeter SN, McDermott CM, Bower PA, Kinzelman JL, Bootsma MJ, Goetz GW, McLellan SL. Bacteroidales diversity in ring-billed gulls (*Laurus delawarensis*) residing at Lake Michigan beaches. Appl Environ Microbiol. 2009; 75:1525–1533. [PubMed: 19151182]

Jjemba PK, Kinkle BK, Shann JR. In-situ enumeration and probing of pyrene-degrading soil bacteria. FEMS Microbiol Ecol. 2006; 55:287–298. [PubMed: 16420636]

Khan IU, Loughborough A, Edge TA. DNA-based real-time detection and quantification of aeromonads from fresh water beaches on Lake Ontario. J Water Health. 2009; 7:312–323. [PubMed: 19240357]

Kreader CA. Design and evaluation of *Bacteroides* DNA probes for the specific detection of human fecal pollution. Appl Environ Microbiol. 1995; 61:1171–1179. [PubMed: 7538270]

Lamendella R, Domingo JW, Oerther DB, Vogel JR, Stoeckel DM. Assessment of fecal pollution sources in a small northern-plains watershed using PCR and phylogenetic analyses of *Bacteroidetes* 16S rRNA gene. FEMS Microbiol Ecol. 2007; 59:651–660. [PubMed: 17069624]

Layton A, McKay L, Williams D, Garrett V, Gentry R, Sayler G. 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. 2006; 72:4214–4224. [PubMed: 16751534]

Layton BA, Walters SP, Boehm AB. Distribution and diversity of the enterococcal surface protein (esp) gene in animal hosts and the Pacific coast environment. J Appl Microbiol. 2009

Lee YJ, Molina M, Santo Domingo JW, Willis JD, Cyterski M, Endale DM, Shanks OC. Temporal assessment of the impact of exposure to cow feces in two watersheds by multiple host-specific PCR assays. Appl Environ Microbiol. 2008; 74:6839–6847. [PubMed: 18806002]

Lee ZM, Bussema C 3rd, Schmidt TM. rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. Nucleic Acids Res. 2009; 37:D489–D493. [PubMed: 18948294]

Leung HD, Chen G, Sharma K. Effect of detached/re-suspended solids from sewer sediment on the sewage phase bacterial activity. Water Sci Technol. 2005; 52:147–152. [PubMed: 16206854]

Ley RE, Peterson DA, Gordon JI. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell. 2006a; 124:837–848. [PubMed: 16497592]

Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: human gut microbes associated with obesity. Nature. 2006b; 444:1022–1023. [PubMed: 17183309]

Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, et al. Evolution of mammals and their gut microbes. Science. 2008; 320:1647–1651. [PubMed: 18497261]

Liu XC, Zhang Y, Yang M, Wang ZY, Lv WZ. Analysis of bacterial community structures in two sewage treatment plants with different sludge properties and treatment performance by nested PCR-DGGE method. J Environ Sci (China). 2007; 19:60–66. [PubMed: 17913155]

Magne F, Abely M, Boyer F, Morville P, Pochart P, Suau A. Low species diversity and high interindividual variability in faeces of preterm infants as revealed by sequences of 16S rRNA genes and PCR-temporal temperature gradient gel electrophoresis profiles. FEMS Microbiol Ecol. 2006; 57:128–138. [PubMed: 16819956]

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005; 437:376–380. [PubMed: 16056220]

Mueller-Spitz SR, Goetz GW, McLellan SL. Temporal and spatial variability in nearshore bacterioplankton communities of Lake Michigan. FEMS Microbiol Ecol. 2009; 67:511–522. [PubMed: 19220863]

NRC. Indicators for Waterborne Pathogens. Washington DC: National Research Council of the National Academies; 2004.

Okabe S, Okayama N, Savichtcheva O, Ito T. Quantification of host-specific *Bacteroides-Prevotella* 16S rRNA genetic markers for assessment of fecal pollution in freshwater. Appl Microbiol Biotechnol. 2007; 74:890–901. [PubMed: 17139508]

Ort C, Gujer W. Sorption and high dynamics of micropollutants in sewers. Water Sci Technol. 2008; 57:1791–1797. [PubMed: 18547932]

Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004; 20:289–290. [PubMed: 14734327]

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glockner FO. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 2007; 35:7188–7196. [PubMed: 17947321]

Rasko DA, Myers GS, Ravel J. Visualization of comparative genomic analyses by BLAST score ratio. BMC Bioinformatics. 2005; 6:2. [PubMed: 15634352]

Reeves RL, Grant SB, Mrse RD, Copil Oancea CM, Sanders BF, Boehm AB. Scaling and management of fecal indicator bacteria in runoff from a coastal urban watershed in southern California. Environ Sci Technol. 2004; 38:2637–2648. [PubMed: 15180060]

Santoro AE, Boehm AB. Frequent occurrence of the human-specific *Bacteroides* fecal marker at an open coast marine beach: relationship to waves, tides and traditional indicators. Environ Microbiol. 2007; 9:2038–2049. [PubMed: 17635548]

Savichtcheva O, Okabe S. Alternative indicators of fecal pollution: relations with pathogens and conventional indicators, current methodologies for direct pathogen monitoring and future application perspectives. Water Res. 2006; 40:2463–2476. [PubMed: 16808958]

Schloss PD, Handelsman J. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl Environ Microbiol. 2005; 71:1501–1506. [PubMed: 15746353]

Shanks OC, Nietch C, Simonich M, Younger M, Reynolds D, Field KG. Basin-wide analysis of the dynamics of fecal contamination and fecal source identification in Tillamook Bay, Oregon. Appl Environ Microbiol. 2006; 72:5537–5546. [PubMed: 16885307]

Shoemaker NB, Vlamakis H, Hayes K, Salyers AA. Evidence for extensive resistance gene transfer among *Bacteroides* spp. and among *Bacteroides* and other genera in the human colon. Appl Environ Microbiol. 2001; 67:561–568. [PubMed: 11157217]

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, et al. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci U S A. 2006; 103:12115–12120. [PubMed: 16880384]

Stewart JR, Gast RJ, Fujioka RS, Solo-Gabriele HM, Meschke JS, Amaral-Zettler LA, et al. The coastal environment and human health: microbial indicators, pathogens, sentinels and reservoirs. Environ Health. 2008; 7 Suppl 2:S3. [PubMed: 19025674]

Surbeck CQ, Jiang SC, Ahn JH, Grant SB. Flow fingerprinting fecal pollution and suspended solids in stormwater runoff from an urban coastal watershed. Environ Sci Technol. 2006; 40:4435–4441. [PubMed: 16903282]

Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. Nature. 2009; 457:480–484. [PubMed: 19043404]

Ufnar JA, Wang SY, Christiansen JM, Yampara-Iquise H, Carson CA, Ellender RD. Detection of the nifH gene of *Methanobrevibacter smithii*: a potential tool to identify sewage pollution in recreational waters. J Appl Microbiol. 2006; 101:44–52. [PubMed: 16834590]

USEPA. Improved enumeration methods for recreational water quality indicators: enterococci and *Escherichia coli*. Washington DC: U.S. EPA Office of Water, Office of Science and Technology; 2000.

USEPA. Method 1603: *Escherichia coli* (*E. coli*) in water by membrane filtration using modified membrane-thermotolerant *Escherichia coli* agar (modified mTEC). EPA 821-R-02-023. Washington, DC: U.S. EPA Office of Water; 2002.

USEPA. Report to Congress: Impacts and Control of CSOs and SSOs. Washington, DC: U.S. EPA Office of Water; 2004.

USEPA. EPA's BEACH Report: 2005 Swimming season. Washington, DC: U.S. EPA Office of Water, Office of Science and Technology; 2006.

USEPA. The National Water Quality Inventory: Report to Congress, 2004 Reporting Cycle. Washington, DC: U.S. EPA Office of Water; 2009.

Vereen E Jr, Lowrance RR, Cole DJ, Lipp EK. Distribution and ecology of campylobacters in coastal plain streams (Georgia, United States of America). Appl Environ Microbiol. 2007; 73:1395–1403. [PubMed: 17172457]

Wade TJ, Calderon RL, Sams E, Beach M, Brenner KP, Williams AH, Dufour AP. Rapidly measured indicators of recreational water quality are predictive of swimming-associated gastrointestinal illness. Environ Health Perspect. 2006; 114:24–28. [PubMed: 16393653]

Whitman RL, Nevers MB. Foreshore sand as a source of *Escherichia coli* in nearshore water of a Lake Michigan beach. Appl Environ Microbiol. 2003; 69:5555–5562. [PubMed: 12957945]

Zheng G, Yampara-Iquise H, Jones JE, Andrew Carson C. Development of *Faecalibacterium* 16S rRNA gene marker for identification of human faeces. J Appl Microbiol. 2009; 106:634–641. [PubMed: 19200327]

Zoetendal EG, von Wright A, Vilpponen-Salmela T, Ben-Amor K, Akkermans AD, de Vos WM. Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. Appl Environ Microbiol. 2002; 68:3401–3407. [PubMed: 12089021]

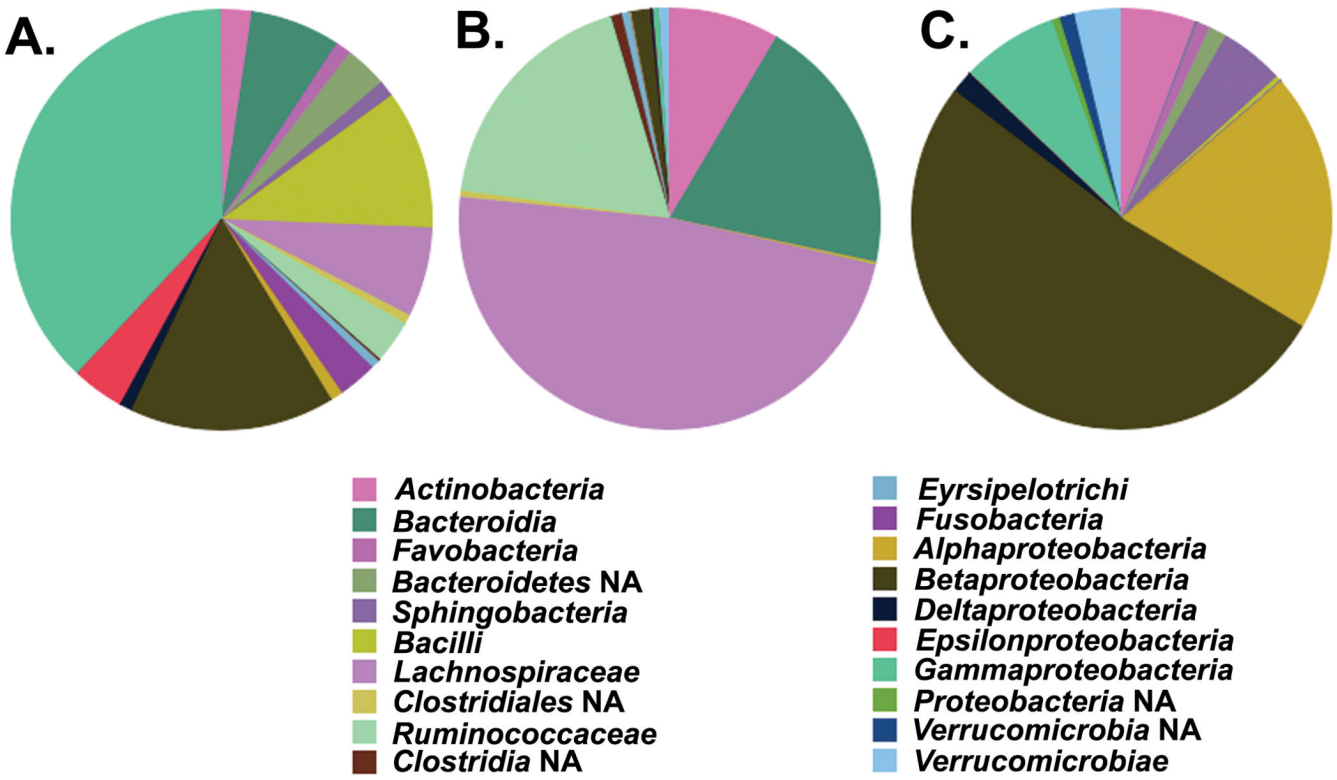

Fig. 1.

(A) Sewage, (B) human, and (C) surface water taxonomic composition by class. *Clostridia* is expanded to the order level, with *Clostridiales* further expanded to the family level. Taxa represented occur at ≥ 1% abundance in at least one sample.

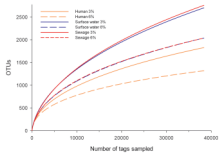

Figure 2.
Rarefaction curves of OTUs defined by 3% and 6% sequence variation in sewage, human, and surface water samples. The 6% curves of sewage and surface water overlap.

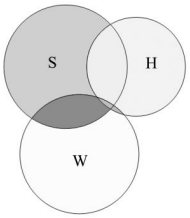

Figure 3.
Venn diagram of the overlap of OTUs from sewage, human, and surface water. OTUs represent clusters with 3% sequence variation. Within a random sub-sampling of 38,368 tag sequences from the total datasets, sewage (S) contained 3045 OTUs, human (H) contained 1950 OTUs, and surface water (W) contained 2833 OTUs. The numbers of overlapping OTUs were as follows: S vs. H = 611, S vs. W = 479, H vs. W = 27 and S vs. H vs. W = 20.

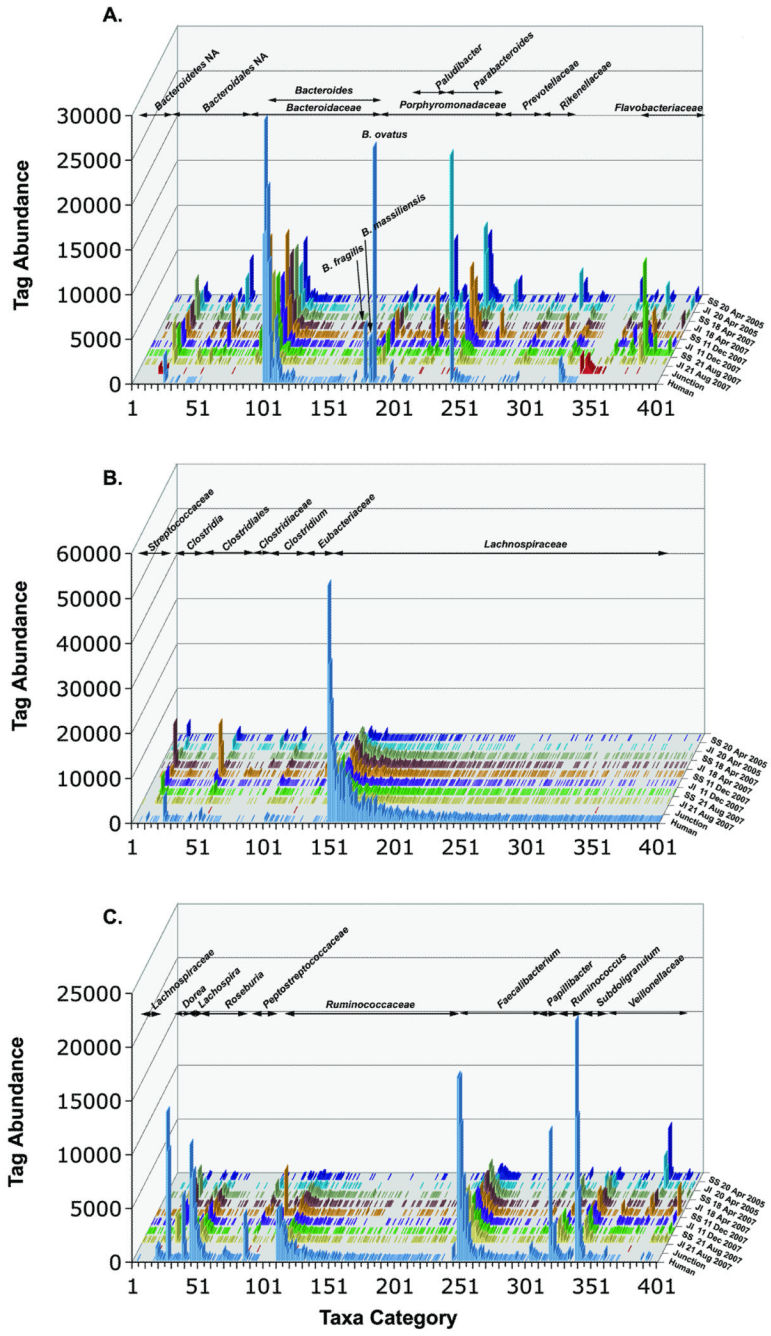

Fig. 4.

Histogram of *Bacteroidetes* (A), and *Firmicutes* (B and C) tags in humans compared with surface water (designated Junction) and eight WWTP samples. The x-axis represents the ordination of sequence tags: sorted first alphabetically by taxomonic name then by decreasing tag frequency. The major taxonomic designations are labeled along the human sample.

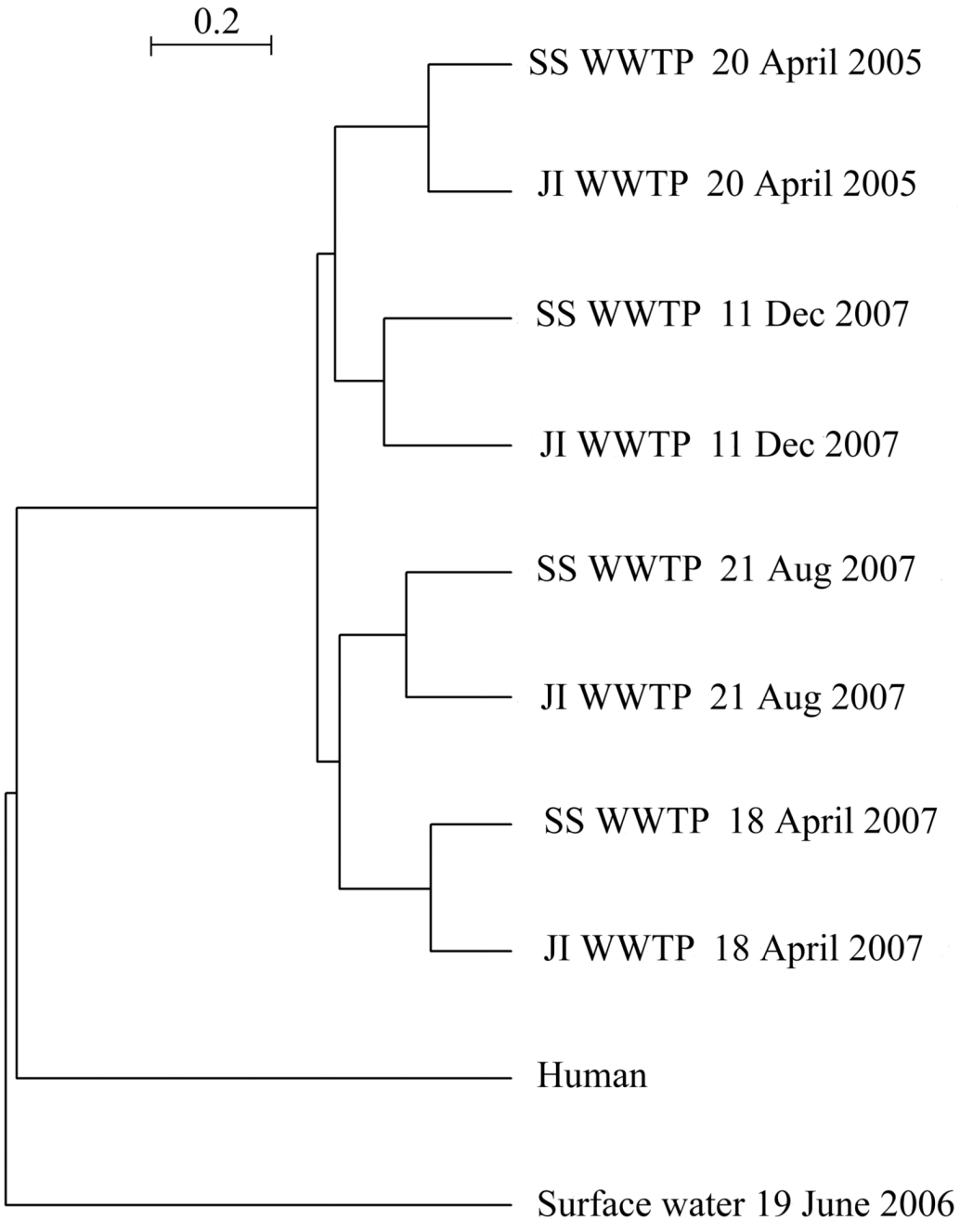

Fig. 5.
UPGMA tree based on Bray-Curtis distances of eight sewage samples, surface water, and human datasets (Dethlefsen *et al.*, 2008 and Turnbaugh *et al.*, 2009). Sewage samples clustered by collection date and were distinct from surface water and human fecal samples.

Table 1

WWTP samples analyzed for community composition.

Date	WWTP	Flow	Rainfall (cm) past		No. of tags	Richness measures[1]		Bray Curtis
	M L day^{-1}		24 hr	72 hr		ACE	Chao	distance[2]
20 April 2005	JI	318	trace*	0.13	17338	1878	1819	0.135
	SS	432			21352	2011	1955	
18 April 2007	JI	356	0.23	0.23	31877	2821	2742	0.133
	SS	545			26503	2320	2240	
21 Aug 2007	JI	1295	3.38	9.07	28684	3709	3695	0.174
	SS	840			34080	3428	3325	
11 Dec 2007	JI	231	trace	0.05	24463	2440	2374	0.213
	SS	281			30793	2567	2496	

[1] Datasets were subsampled to 17,300 to provide equal dataset size for calculations

[2] Calculated for JI and SS on same date.

** Trace rain corresponds to less than 0.025 cm of rain.

Table 2

OTUs (defined by 3% pairwise distance) shared among sewage samples and overlap with human and surface water.

Richness endemic to sewage samples		% of sewage tag dataset shared among sewage	No. of sewage OTUs shared with human[1] datasets	% of sewage tag dataset shared with humans	No. of sewage OTUs shared with surface water	% of sewage tag dataset shared with surface water
No. of samples	**No. of shared OTUs**					
8	118	70.2%	44	22.8%	53	54.0%
7	85	7.2%	42	3.5%	24	2.5%
6	77	4.0%	31	1.3%	17	1.1%
5	100	3.3%	38	1.0%	15	0.7%
4	158	3.1%	52	0.9%	31	0.7%
3	256	3.1%	66	0.7%	50	0.8%
2	494	3.7%	90	0.6%	88	1.0%
1	1757	5.4%	248	0.7%	201	0.7%
Total	**3045**	**100 %**	**611**	**31.5%**	**479**	**61.4%**

[1] Human datasets from Dethlefsen *et al.*, 2008 and Turnbaugh *et al.*, 2009

Table 3

The 20 most dominant taxa based upon V6 sequence tag abundance in human datasets compared with abundance in sewage and surface water.

Taxonomy		Percentage of sequence tags*		
Order/Family	**Genus (species)**	**Human**	**Sewage**	**Surface water**
Lachnospiraceae †		40.78	5.49	0.03
Bacteroidaceae	*Bacteroides*	11.17	1.98	<0.01
Ruminococcaceae	*Faecalibacterium*	7.28	1.35	ND**
Bifidobacteriaceae	*Bifidobacterium*	7.28	0.74	ND
Ruminococcaceae		4.87	0.91	0.02
Ruminococcaceae	*Subdoligranulum*	4.13	0.38	ND
Lachnospiraceae	*Roseburia*	4.09	0.37	ND
Bacteroidaceae	*Bacteroides ovatus*	2.25	0.17	ND
Lachnospiraceae	*Dorea*	2.17	0.30	ND
Ruminococcaceae	*Ruminococcus*	2.16	0.31	ND
Coriobacteriaceae	*Collinsella*	1.07	0.24	ND
Alcaligenaceae	*Sutterella*	1.00	0.05	0.01
Porphyromonadaceae	*Parabacteroides*	0.99	1.33	<0.01
Clostridia †		0.95	0.17	<0.01
Akkermansiaceae	*Akkermansia*	0.93	0.08	ND
Peptostreptococcaceae †		0.76	0.20	<0.01
Lachnospiraceae	*Lachnospira*	0.71	0.07	ND
Bacteroidaceae	*Bacteroides fragilis*	0.58	0.06	ND
Bacteroidaceae	*Bacteroides massiliensis*	0.57	0.12	ND
Ruminococcaceae	*Papillibacter*	0.48	0.06	<0.01
Total		**95.24**	**14.35**	**0.07**

* Based upon percentage of all sequence tags associated with taxonomic classification.

** ND, Taxa not detected.

† Unable to resolve tags to family and/or genus level.

Table 4

The 25 most dominant taxa in sewage based on V6 sequence tag abundance datasets compared with abundance in human datasets and surface water.

Taxonomy of sequence tag		Percentage of sequence tags[a]		
Order/Family		**Human**	**Sewage**	**Surface Water**
Moraxellaceae	*Acinetobacter*	<0.01[b]	15.46	0.26
Aeromonadaceae	*Aeromonas*	<0.01	11.34	0.52
Carnobacteriaceae	*Trichococcus*	<0.01	6.97	0.02
Lachnospiraceae [c]		40.78	5.49	0.03
Pseudomonadaceae	*Pseudomonas*	<0.01	4.54	0.73
Comamonadaceae [c]		<0.01	4.44	23.84
Campylobacteraceae	*Arcobacter*	<0.01	3.75	ND
Bacteroidetes [c]		<0.01	3.23	1.32
Fusobacteriales [c]		ND	2.45	<0.01
Bacteroidaceae	*Bacteroides*	11.17	1.98	<0.01
Comamonadaceae	*Acidovorax*	<0.01	1.86	0.28
Streptococcaceae	*Lactococcus*	0.02	1.67	<0.01
Comamonadaceae	*Simplicispira*	<0.01	1.35	0.09
Ruminococcaceae	*Faecalibacterium*	7.28	1.35	ND
Porphyromonadaceae	*Parabacteroides*	0.99	1.33	<0.01
Enterobacteriaceae [c]		0.24	1.23	1.13
Cytophagaceae	*Sporocytophaga*	ND	1.08	0.37
Betaproteobacteria [c]		<0.01	1.00	1.78
Bacteroidales[c]		0.39	1.00	0.07
Aeromonadaceae	*Tolumonas auensis*	ND	0.95	0.18
Ruminococcaceae [c]		4.87	0.91	0.02
Neisseriaceae [c]		ND	0.84	0.11
Rhodocyclaceae	*Propionivibrio*	ND	0.82	0.05
Porphyromonadaceae	*Paludibacter*	<0.01	0.81	0.04
Bifidobacteriaceae	*Bifidobacterium*	7.28	0.74	ND
Total percentage of sequence tags		**73.05**	**76.56**	**30.84**

[a] Based upon percentage of all sequence tags associated with taxonomic classification.

[b] Taxa detect at < 0.01%.

ND, Taxa not detected.

[c] Unable to resolve tags to family and/or genus level.

Table 5

Mean total cell counts by direct staining and culturable fecal indicator levels in WWTP influent samples from the Milwaukee metropolitan area.

WWTP	Total counts cells ml^{-1} (± 1 SD)	*E. coli*	enterococci CFU ml^{-1} (± 1 SD)	*C. perfringens*
Jones Island	3.96×10^9	1.08×10^4	5.40×10^3	522
	(± 1.91×10^9)	(± 1.91×10^3)	(± 1.58×10^3)	(± 218)
South Shore	4.69×10^9	1.24×10^4	5.27×10^3	396
	(± 3.54×10^9)	(± 2.64×10^3)	(± 1.75×10^3)	(± 255)