View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Missouri University of Science and Technology (Missouri S&T): Scholars' Mine

Missouri University of Science and Technology

Scholars' Mine

Civil, Architectural and Environmental Engineering Faculty Research & Creative Works

Civil, Architectural and Environmental Engineering

01 Feb 2013

# Molecular Diversity of Bacteroidales in Fecal and Environmental Samples and Swine-Associated Subpopulations

Regina Lamendella

Kent C. Li

Daniel B. Oerther
*Missouri University of Science and Technology*, oertherd@mst.edu

Jorge W. Santo Domingo

Follow this and additional works at: https://scholarsmine.mst.edu/civarc_enveng_facwork

Part of the Civil and Environmental Engineering Commons

## Recommended Citation

R. Lamendella et al., "Molecular Diversity of Bacteroidales in Fecal and Environmental Samples and Swine-Associated Subpopulations," *Applied and Environmental Microbiology*, vol. 79, no. 3, pp. 816-824, American Society for Microbiology, Feb 2013.
The definitive version is available at https://doi.org/10.1128/AEM.02535-12

This Article - Journal is brought to you for free and open access by Scholars' Mine. It has been accepted for inclusion in Civil, Architectural and Environmental Engineering Faculty Research & Creative Works by an authorized administrator of Scholars' Mine. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact scholarsmine@mst.edu.

AEM
Journals.ASM.org

# Molecular Diversity of *Bacteroidales* in Fecal and Environmental Samples and Swine-Associated Subpopulations

**Regina Lamendella,[a,b]* Kent C. Li,[b] Daniel Oerther,[a]* Jorge W. Santo Domingo[c]**

University of Cincinnati, Department of Civil and Environmental Engineering, Cincinnati, Ohio, USA[a]; Lawrence Berkeley National Laboratory, Earth Sciences Division, Ecology Department, Berkeley, California, USA[b]; National Risk Management Research Laboratory, Environmental Protection Agency, Cincinnati, Ohio, USA[c]

**Several swine-specific microbial source tracking methods are based on PCR assays targeting *Bacteroidales* 16S rRNA gene sequences. The limited application of these assays can be explained by the poor understanding of their molecular diversity in fecal sources and environmental waters. In order to address this, we studied the diversity of 9,340 partial (>600 bp in length) *Bacteroidales* 16S rRNA gene sequences from 13 fecal sources and nine feces-contaminated watersheds. The compositions of major *Bacteroidales* populations were analyzed to determine which host and environmental sequences were contributing to each group. This information allowed us to identify populations which were both exclusive to swine fecal sources and detected in swine-contaminated waters. Phylogenetic and diversity analyses revealed that some markers previously believed to be highly specific to swine populations are shared by multiple hosts, potentially explaining the cross-amplification signals obtained with nontargeted hosts. These data suggest that while many *Bacteroidales* populations are cosmopolitan, others exhibit a preferential host distribution and may be able to survive different environmental conditions. This study further demonstrates the importance of elucidating the diversity patterns of targeted bacterial groups to develop more inclusive fecal source tracking applications.**

Microbial source tracking (MST) is a rapidly evolving area in applied microbiology that focuses on identifying the source(s) of fecal contamination impacting environmental waters (1). Dozens of MST assays have been proposed, but many recent field applications have focused on using PCR-based assays targeting *Bacteroidales* 16S rRNA gene sequences (2–4). As a group, *Bacteroidales* are diverse and numerically abundant within the animal distal gut and other human cavities (5). From a fecal pollution standpoint, some *Bacteroidales* populations have been suggested to exhibit a preferential host distribution (6–8), explaining why multiple methods have targeted this bacterial group. However, only a limited number of *Bacteroidales*-based methods have been successfully used in field applications, in part explained by the fact that most methods only partially comply with crucial criteria for source identification, such as host specificity, host distribution, and the temporal and geographic stability of the genetic markers (2, 9). Indeed, assays originally proposed to be specific to some of the most relevant fecal pollution source types, such as human, cattle, and swine, cross amplify with nonspecific targets in studies using a greater number of fecal specimens or feces collected from different geographic locations (8, 10, 11). Other problems relate to the development of assays based on sequences derived directly from human and animal fecal samples and not from fecal sources that are commonly present in the environment. For example, fecal sources such as manure pits and waste lagoons are often responsible for animal fecal loadings. In the case of human fecal sources, wastewater treatment plants and septic tanks are the most important contributors of human fecal pollution. Additionally, there is limited information available on the survival rates of fecal populations in impacted waters, a problem that is difficult to address, as current host-specific markers are likely to target multiple populations, some of which are associated with different survival rates. There is also the significant challenge of discriminating between fecal bacterial indicator strains (i.e., *Escherichia coli* and enterococci) associated with recent contamination events and those adapted to secondary habitats (12, 13).

While human sources are considered to carry higher public health risks (14), sources of fecal pollution from domesticated animals are a significant detriment to water quality and impose risks to human health and aquatic ecosystems. A recent study indicated that current farming practices are responsible for 70% of the pollution in U.S. rivers and streams (15). In particular, animal manure has been identified as a large contributor to water pollution due to its overabundance (16). The U.S. Environmental Protection Agency (USEPA) estimates that the volume of manure from confined animal feeding operations (CAFOs) is three times our nation's volume of human fecal waste (17). Specifically, the concentration of swine farming operations has increased significantly over the past 4 decades, resulting in the production of large amounts of more concentrated waste products. Not surprisingly, the swine industry worldwide has become an increasing environmental concern, due to microbial pollutants from these operations potentially impacting nearby water bodies through runoff or accidental spillage or groundwater contamination by infiltration. The marked increase in the amount of swine waste produced per

**Received** 16 August 2012 **Accepted** 13 November 2012
**Published ahead of print** 16 November 2012
Address correspondence to Jorge W. Santo Domingo, santodomingo.jorge@epa.gov.
* Present address: Regina Lamendella, Von Liebig Center for Science, Juniata College, Huntingdon, Pennsylvania, USA; Daniel Oerther, Environmental Research Center, Missouri University of Science and Technology, Rolla, Missouri, USA.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.02535-12.
Copyright © 2013, American Society for Microbiology. All Rights Reserved.
doi:10.1128/AEM.02535-12

Downloaded from http://aem.asm.org/ on January 27, 2016 by Missouri S&T

farming operation has raised concerns about swine waste storage and treatment processes. For example, the Environmental Integrity Project documented 329 manure spills in Iowa between 1992 and 2002, due to failure or overflow of manure storages, uncontrolled runoff from open feedlots, improper manure application on cropland, deliberate pumping of manure onto the ground, and intentional breeches in storage lagoons (18). When introduced into water, swine fecal waste can present risks to human health because this waste can harbor a variety of human pathogens (*Escherichia coli* O157:H7, *Citrobacter freundii*, *Enterocytozoon bieneusi*) and thus represents an important reservoir for zoonotic pathogens (19–21). Swine waste also can contain high concentrations of residual antibiotics (22), nutrients (23), and heavy metals (24). Clearly, control of swine waste from entering waters used for recreation, fishing, and public water supplies is relevant to assist in meeting water quality standards assigned to protect environmental and human health. If the origin of fecal pollution can be correctly and rapidly identified, best management practices and remediation efforts (such as improvement of waste storage structures, timing of processed fecal waste application in agricultural fields, changing of the pasture management approach, and addition of fences and runoff control structures) could be introduced in a timely and cost-effective manner.

Regardless of the targeted gene, with a few exceptions (25), most of the currently available PCR-based host-specific assays have been designed by comparing a relatively small number of sequences closely related to the bacterial taxa in question. Even for those targeting the 16S rRNA gene, development of host-specific assays has been conducted with less than a hundred sequences, although in some cases fingerprinting methods have been combined to screen for differences between fecal microbial communities (6, 19). In the case of swine assays, pig fecal sequences used for primer development were derived from specific local geographic areas. For example, Dick et al. (6) designed a pig-specific assay based on a host-specific phylogenetic clade containing less than 10 sequences from one pooled fecal sample collected in Oregon, while Okabe et al. (26) developed an assay that targeted two small *Prevotella* pig-specific clades, each containing a small number of sequences derived from two Japanese pig farms. Similarly, cattle, dog, horse, and wildlife *Bacteroidetes* assays have been based on small sequencing databases. Given the vast diversity of *Bacteroidales* populations, additional sequencing seems necessary in order to resolve the level of specificity of these 16S rRNA gene-based assays.

Unfortunately, the membership of *Bacteroidales* harbored within different animal types and environmental systems is still poorly characterized. In order to design comprehensive assays for accurately quantifying contributions of fecal pollution from different hosts, it is necessary to first characterize the diversity and distribution of targeted populations and their relative abundances in both fecal and environmental matrices. Thus, evaluating the community structure, membership, and abundance of *Bacteroidales* from geographically diverse host feces and environmental fecal sources can be used to reveal previously unknown host-specific populations. This information is also necessary to estimate their relative abundance in various hosts and to identify populations relevant to environmental fecal pollution. This study focused on evaluating the molecular diversity and distribution of *Bacteroidales* populations derived from several swine and nonswine fecal sources and polluted environmental water samples, as an approach to identify signature sequences specific to swine fecal sources that can be used to detect swine-fecal impacts in environmental waters.

## MATERIALS AND METHODS

Sequences for diversity and bioinformatics analyses were obtained from the ARB-Silva database and from clone libraries generated in the EPA-JSD laboratory using *Bacteroidales*-specific primers (Table 1). Sequences from the ARB-Silva database were selected on the basis of the following criteria: organism name *Bacteroidetes*, sequence quality of >95%, sequence length of >600 bp, and pintail quality of >90%. Approximately 70% (6,413 sequences) of the total sequences used in the analyses were generated in our laboratory. The sequences were generated from different samples: animal feces, manure pits, waste lagoons, septic tanks, wastewater, sediment, soil to which manure had been applied, and groundwater and surface water (Table 1). For most animal sources, several samples (i.e., 3 to 18 individual samples) were used to generate individual clone libraries. In a few libraries, DNA extracts from multiple samples were pooled in equimolar amounts to generate clone libraries. Sequences generated were then pooled for the diversity analyses following steps described elsewhere (8, 11). The samples were collected aseptically with sterile spatulas, transferred into sterile tubes, transported to the laboratory within 6 h of collection, and stored at −80°C until further processing. The primary goal was to include in the analysis as many different animal types and environmental sources as possible to check for host specificity, with emphasis on hosts and environmental sources considered important in the pollution of surface waters in the United States.

**DNA extraction, PCR assays, cloning, and sequence analyses.** Fecal and environmental DNA extractions were performed as previously described (11). In all cases, a general *Bacteroidales* 16S rRNA gene PCR assay (Bac32f/Bac708r) was used to amplify *Bacteroidales* from samples using the conditions previously described by Bernhard and Field (27), with the exception that a different DNA polymerase and thermal cycler were used to generate PCR products. Briefly, PCR assays were conducted using the following conditions: 94°C for 5 min, followed by 30 cycles of 94°C for 20 s, 55°C for 20 s, and 72°C for 30 s and a final extension step consisting of 72°C for 5 min. Fecal and water sample DNA template concentrations used in the PCR assays were adjusted on the basis of published detection limits (27). Final PCR solutions (25-μl total volume) contained 2.5 μl of TaKaRa *Ex Taq* 10× buffer (20 mM $Mg^{2+}$), 2 μl of deoxynucleoside triphosphate mixture (2.5 mM each), 1 μl of 25% acetamide, 17.5 μl of ultrapure water, 12.5 pmol of each forward and reverse primer, and 0.625 U of *Ex Taq* DNA polymerase (TaKaRa Mirus Bio, Madison, WI). Reactions were conducted on a DNA Engine 2 Tetrad thermal cycler (Bio-Rad Laboratories, Inc., Hercules, CA). Amplification products were visualized using 1% agarose gels and GelStar nucleic acid stain (Cambrex BioScience, East Rutherford, NJ). Cloning was performed using the pCR4.1 TOPO vector as previously described (8). Clones were submitted to Cincinnati Children's Medical Hospital DNA Core Facility (Cincinnati, OH) for sequencing using BigDye sequencing chemistry (Applied Biosystems, Foster City, CA) and an Applied Biosystems Prism 3730XL DNA analyzer.

**Sequence quality and phylogeny.** Sequences were manually verified and cleaned using Sequencher software (Gene Codes, Ann Arbor, MI). Sequences were sent to the Greengenes online pipeline (28) for alignment using the Nearest Alignment Space Termination (NAST) alignment tool and subsequently checked for chimeras using Bellerophon (version 3) software (29). Chimeras and putative chimeras were not included in further analyses. Nonchimeric sequences were submitted to Greengenes for alignment using the NAST algorithm (28, 30). Sequences were also submitted to BLAST homology search algorithms to assess sequence similarity to sequences in the Greengenes database (28, 31). The distance matrix and phylogenetic tree were generated using ARB software (32). Trees were inferred from 650 sequence positions using neighbor joining (using a Kimura correction) and maximum parsimony (using the Phylip DNAPARS tool) (32). To statistically evaluate branching confidence,

Downloaded from http://aem.asm.org/ on January 27, 2016 by Missouri S&T

TABLE 1 *Bacteroidales* 16S rRNA gene sequences from fecal source and environmental samples

| Sample | Matrix | Major contaminating source(s) | Total no. of sequences[a] |
|---|---|---|---|
| Wastewater | Water (influent and effluent) | Human | 1,245 (397) |
| Cattle feces | Feces | Cattle | 1,216 (982) |
| Human feces | Feces | Human | 844 (844) |
| Water, Canada | Surface water | Avian | 765 |
| Water, OH/IL, pig contaminated | Surface water | Swine | 668 |
| Water, MA | Surface water | Human, cattle | 618 |
| Pig lagoon | Feces-water slurry | Pig | 582 |
| Water, NY | Surface water | Unknown | 560 |
| Pig feces | Feces | Swine | 409 (150) |
| Pig manure pit | Feces | Swine | 377 |
| Gull feces | Feces | Seagull | 339 (339) |
| Water, TN | Surface water | Human, cattle, wildlife | 332 |
| Water, Singapore | Surface water | Swine, unknown | 196 |
| Water, SC | Surface water | Unknown | 172 |
| Groundwater, IL, pig contaminated | Groundwater | Swine | 169 |
| Water, NE | Surface water | Cattle | 168 |
| Horse feces | Feces | Horse | 140 (101) |
| Chicken feces | Feces | Chicken | 99 (99) |
| Sediment | Sediment | Cattle, unknown | 88 |
| Septic tank | Feces-water slurry | Human | 83 (15) |
| Soil to which pig manure had been applied | Soil | Swine | 79 |
| Wild pig feces | Feces | Swine | 71 |
| Water, OH/IL, pig contaminated, PF163 marker | Surface water | Swine | 67 |
| Fish gut | Luminal contents | Fish | 53 |
| Total | | | 9,340 (2,927) |

[a] The number of sequences that were retrieved from the ARB-Silva database is indicated in parentheses.

bootstrap values were obtained from a consensus of 100 parsimonious trees using MEGA software (http://www.megasoftware.net). A *Werenella* sp. 16S rRNA gene sequence (GenBank accession number AJ234059) was used as the outgroup, while cultured *Bacteroidales* species were included in the analyses as points of reference.

**Taxonomic assignment, diversity, and statistical analysis.** Molecular diversity analyses and assemblage comparison of clone libraries were performed using the QIIME (Quantitative Insights into Microbial Ecology) pipeline (33). Briefly, the16S rRNA gene sequences were clustered with uclust and assigned to operational taxonomic units (OTUs) with 97% similarity. Consensus sequences from each OTU were assigned a Greengenes taxonomy. As the number of sequence reads in each sample varied, the OTU table was rarified using 50 iterations prior to alpha-diversity calculations. The subsampled OTU tables were then used to calculate the observed number of OTUs and the abundance-based coverage estimate (ACE), Chao1, Dominance, and Heip evenness metrics. The OTU file containing the relative abundance of each OTU in each sample was used to construct a heat map, where species were clustered using the unweighted-pair group method using average linkages (UPGMA) clustering algorithm. Multivariate community analysis was performed within PCORD (version 5) software (34) using normalized OTU tables generated in QIIME. Nonmetric multidimensional scaling plots were calculated using the Sorensen distance measure. The Bray-Curtis distance measure was used to create the distance matrix. Nonmetric multidimensional scaling (nMDS) was created in PCORD (version 5) using the slow-and-thorough setting with 500 iterations, 0.0000001 instability, and 250 randomized runs. The best solution for each dimension was selected to create the lowest final stress. The OTU heat map was generated within QIIME, OTUs were clustered using the UPGMA hierarchical clustering algorithm, and abundance data were log transformed. QIIME was also used for network-based analysis to display OTU partitioning between samples. Sample nodes were clustered on the basis of the number of OTUs shared between the samples and weighted according to the number of sequences in that OTU. OTUs and samples were clustered using the spring-embedded algorithm implemented in Cytoscape (version 2.8).

Hypothesis testing comparing the *Bacteroidales* assemblages was performed using the mothur program (35). Briefly, the Yue and Clayton measure of similarity between the structures of any two *Bacteroidales* assemblages (OTU distance = 0.03) was used to generate pairwise similarities. The statistical significance of these pairwise similarities was tested using the libshuff method with 10,000 randomizations (35). This test describes whether two or more communities have the same structure using the Cramer-von Mises test statistic. The significance of the test statistic indicates the probability that the communities have the same structure by chance. Since each pairwise comparison requires two significance tests, a Bonferroni correction for multiple comparisons was applied. The experiment-wide false detection rate was 0.05, and thus, values less than 0.025 were considered statistically significant. If either of the pairwise significance values is statistically significant, one can conclude that the two communities are significantly different.

**Nucleotide sequence accession numbers.** Representative sequences generated in this study have been deposited in the GenBank database under accession numbers KC300359 to KC304787.

## RESULTS AND DISCUSSION

**Diversity of *Bacteroidales* 16S rRNA gene sequences.** A total of 9,340 nonchimeric *Bacteroidales* sequences were retrieved from 13 different fecal sources and nine different fecally contaminated watersheds (Table 1). When the diversity of each of the *Bacteroidales* libraries was assessed, the Chao1 and ACE indices revealed a high diversity of most fecal and environmental *Bacteroidales* assemblages (Table 2). ACE and Chao1 diversity estimates showed that the swine fecal and manure pit samples were the most diverse, with very low dominance and high evenness (Table 2). Sequencing several hundred *Bacteroidales* clones for many of the libraries did

Downloaded from http://aem.asm.org/ on January 27, 2016 by Missouri S&T

**TABLE 2** Alpha-diversity metrics for fecal source and environmental samples[a]

| Sample | ACE | Chao1 | Dominance | Heip evenness | No. of OTUs |
|---|---|---|---|---|---|
| Pig feces | 345.34 | 303.55 | 0.020 | 0.28 | 182 |
| Pig manure pit | 480.68 | 438.79 | 0.030 | 0.26 | 185 |
| Pig lagoon | 81.76 | 82.15 | 0.060 | 0.22 | 83 |
| Cattle feces | 123.38 | 123.81 | 0.10 | 0.16 | 147 |
| Gull feces | 44.19 | 39.22 | 0.73 | 0.050 | 27 |
| Human feces | 72.27 | 76.30 | 0.070 | 0.21 | 91 |
| Wastewater | 224.58 | 202.86 | 0.050 | 0.21 | 285 |
| Water, pig contaminated | 332.86 | 314.81 | 0.030 | 0.26 | 262 |
| Groundwater, pig contaminated | 27.34 | 28.35 | 0.20 | 0.17 | 20 |
| Water, MA | 119.96 | 129.58 | 0.070 | 0.20 | 118 |
| Water, NE | 88.040 | 87.67 | 0.070 | 0.20 | 49 |
| Water, NY | 65.18 | 57.82 | 0.16 | 0.14 | 53 |
| Water, TN | 341.88 | 305.91 | 0.020 | 0.29 | 165 |
| Water, Singapore | 263.23 | 239.90 | 0.020 | 0.27 | 103 |
| Water, SC | 144.86 | 136.98 | 0.060 | 0.20 | 63 |
| Water, CA | 211.21 | 182.62 | 0.050 | 0.21 | 179 |

[a] ACE, Chao, dominance, and Heip evenness were calculated on rarified OTU tables at an equivalent depth of 160 sequences per sample. Samples having less than 150 sequences were excluded from the alpha-diversity analysis.

not saturate the diversity of this bacterial group, as 409 and 377 sequences were retrieved from pig feces and pig manure pits, respectively, while covering only between 51 and 67% of the estimated *Bacteroidales* diversity. The rarefaction curves of pig fecal and manure pit samples are not approaching a horizontal asymptote, indicating that the current sequencing effort had not saturated diversity (Fig. 1). In contrast, the *Bacteroidales* diversity observed from pig groundwater and pig lagoons was less (Fig. 1 and Table 2). For example, the Chao1 estimates were 28 and 82 *Bacteroidales* OTUs for these environments, respectively (Table 2). This result suggests that environmental factors are playing an important role in the structural dynamics of *Bacteroidales* and as a result driving down the diversity of this bacterial group within these environments.

nMDS of *Bacteroidales* assemblages revealed specific clustering patterns among the fecal source and environmental samples (Fig. 2A to C). Pig fecal and manure pit samples clustered closely along with soils to which pig manure had been applied, indicating that these samples have a very similar *Bacteroidales* community structure. Interestingly, pig lagoon *Bacteroidales* clustered more closely with water environments impacted by swine fecal pollution, suggesting that a different *Bacteroidales* population structure exists within swine lagoons than manure pits or feces. This finding has important implications for source-tracking assay development, as different swine waste management practices (i.e., manure pit versus lagoon) may result in different *Bacteroidales* assemblage structures. In such cases, multiple markers may in fact be necessary to target these different fecal source populations. Moreover, *Bacteroidales* sequences from swine-impacted waters clustered more closely with swine lagoon and other surface water environments, suggesting that swine-fecal source *Bacteroidales* assemblages transported into the environment may undergo another population shift. Understanding the differential survivability of fecal source-specific populations and their dynamics is critical for the quantification of different fecal sources in environmental monitoring scenarios. Recent studies have indicated that human- and bovine-specific *Bacteroidales* markers have differential survivability under various environmental conditions (36, 37). Thus, studying the molecular diversity of source-specific populations from feces to processed fecal waste and, ultimately, transport of source-specific bacteria into the environment seems necessary for discovery of markers of these source-specific targets. As our study is limited to samples collected in a limited number of geographic locations, additional studies need to be conducted to determine if the relationships among pig waste samples noted here are significantly different when we increase the number of geographically different samples.

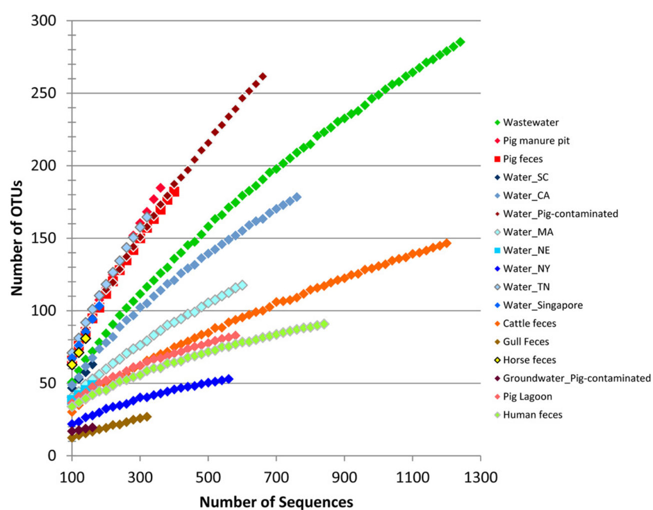


**FIG 1** *Bacteroidales* rarefaction curves for fecal and environmental libraries. Rarefaction curves were generated only for samples having more than 150 sequences. Rarified OTU tables were generated using 50 iterations. Observed OTUs were calculated for each rarified OTU table and averaged.

While nonmetric multidimensional scaling demonstrated clustering of swine fecal sources, *Bacteroidales* assemblages associated with swine waste and contaminated waters exhibited several differences as well. For example, hypothesis testing was used to compare any two *Bacteroidales* assemblages, pig feces, pig manure pit, pig lagoon, and soil to which pig manure had been applied, and revealed significantly different structures among these swine

Downloaded from http://aem.asm.org/ on January 27, 2016 by Missouri S&T

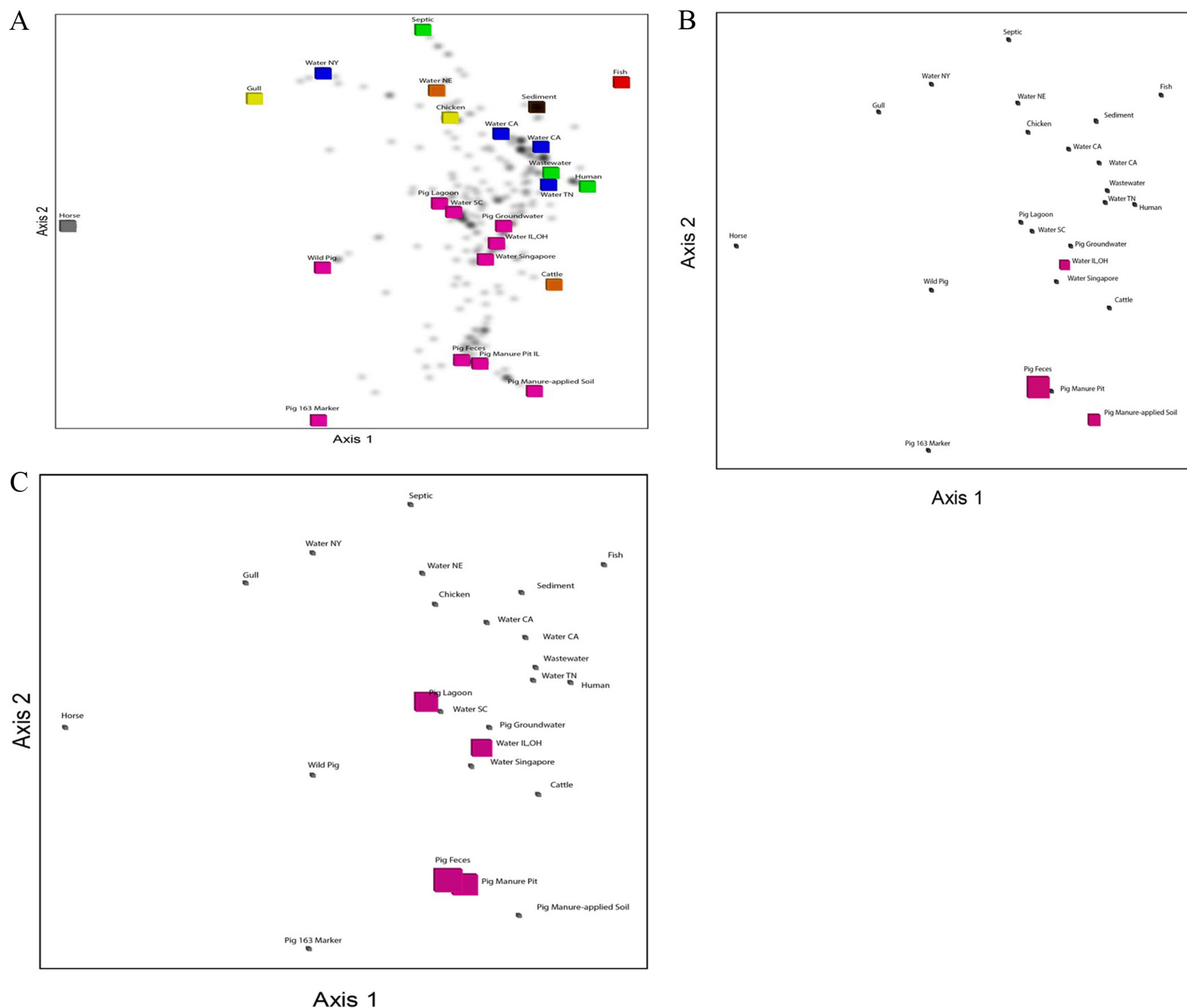


**FIG 2** (A) nMDS of fecal source and environmental samples. The final stress for the nMDS displayed here was 6.466. Black dots, OTUs. (B and C) nMDS with sample points scaled by the relative abundance of (B) OTU 960 and (C) OTU 1401. Black squares, samples containing none of these specific OTUs.

samples ($P < 0.025$) (see Table S1 in the supplemental material). These results further demonstrate the need for MST approaches which use a multiple-target/marker approach per source type to cover differentiated host-specific populations. *Bacteroidales* community structures between swine and cattle feces and swine lagoon and cattle feces were significantly different (see Table S1 in the supplemental material), although these environments may harbor similar *Bacteroidales* populations, as indicated by the nMDS analysis, in which the cattle fecal sample clustered most closely with swine samples (Fig. 2). The cosmopolitan nature of some *Bacteroidales* populations (i.e., they are present in multiple hosts) has previously been noted (11, 38), and finding truly host-specific *Bacteroidales* populations may prove to be difficult, given our poor understanding of the diversity and host distribution of this bacterial group. The limited phylogenetic resolution of the 16S rRNA gene may further complicate targeting of these smaller clusters of host-specific populations.

In order to better elucidate the distribution of *Bacteroidales* populations within the different fecal source and environmental samples, network analysis was performed (Fig. 3). This approach allowed visualization of the OTU partitioning between samples and clustering of the samples on the basis of the number of OTUs shared between the samples. As a result, we were able to further probe *Bacteroidales* diversity, revealing what may be ecologically relevant patterns within fecal and environmental matrices. For example, this analysis unveiled several OTUs that were shared by multiple hosts/environmental sources, such as OTUs shared between a swine fecal source and cattle feces and between municipal wastewater and swine lagoons. While many populations appeared to have shared distributions within several fecal and environmental samples, this analysis also led to the discovery of swine-specific populations that could also be identified in environmental samples known to be contaminated with swine feces (Fig. 3) and OTUs shared exclusively by at least two swine-related samples. We



Downloaded from http://aem.asm.org/ on January 27, 2016 by Missouri S&T

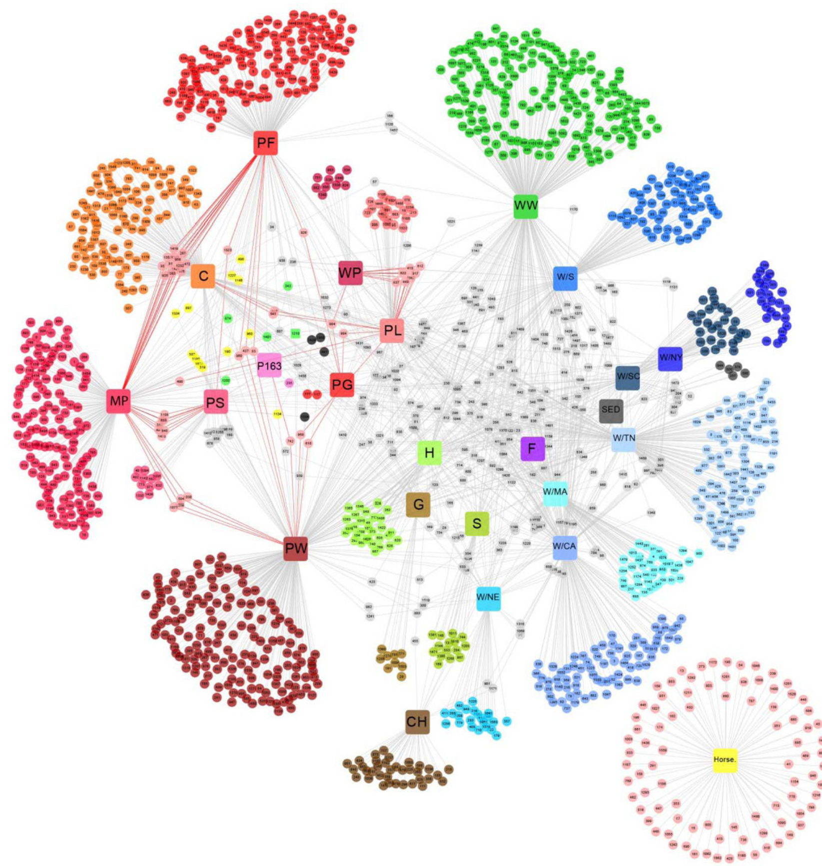


**FIG 3** Network of OTUs from fecal source and environmental samples. Large nodes are samples (MP, pig manure pit; PS, soil to which pig manure had been applied; WP, wild pig; PW, water from IL/OH, pig contaminated; PF, pig feces; PG, groundwater from IL, pig contaminated; P163, water, pig contaminated, sequences amplified with pig-specific PF163 marker; PL, pig lagoon; WW, wastewater; H, human feces; S, septic tank; W/NE, water from NE; W/CA, water from Canada; W/MA, water from MA; W/TN, water from TN; W/NY, water from NY; W/SC, water from SC; W/S, water from Singapore; Horse, horse feces; CH, chicken feces; G, gull feces; F, fish gut; C, cattle; SED, sediment). Network-based analysis was used to display OTU partitioning between samples. Sample nodes are clustered on the basis of the number of OTUs shared between the samples and are weighted according to the number of sequences in that OTU. OTUs and samples were clustered using the spring-embedded algorithm implemented in Cytoscape. Red edges (i.e., lines) highlight connections shared by more than one pig-related sample. Pink, green, yellow, and black OTU nodes, OTUs shared by 2, 3, 4, and 5 pig-related samples, respectively; OTU nodes colored the same as each sample node, OTUs unique to that sample; gray nodes, OTUs shared by more than one type of host. A G test for independence was used to test whether sample nodes are more connected within a group than expected by chance. Each pair of samples is classified according to whether its members shared at least one OTU and whether they shared a category. Pairs were then tested for independence in these categories.

also identified OTUs that were shared exclusively by pig feces and environments contaminated by swine fecal pollution, including OTU 960 (Fig. 2B) and OTU 1401 (Fig. 2C). While the relative abundance of these OTUs diminished from the fecal source to the environmental samples, these OTUs were still detected in swine-contaminated environments, suggesting that they could be promising targets for swine-specific fecal pollution.

Most (>70%) of the swine-specific OTUs have high sequence homology to known *Prevotella* species. Recent studies have also reported a high abundance of *Prevotella* species in the swine gut (39, 40). Although the physiological reasons explaining the prevalence of *Prevotella* over other *Bacteroidetes* is unknown, diet composition might play an important role (41). Altogether, these data suggest that some specific *Prevotella* populations may share endemic or host-specific distributions with the swine host. Approximately 10% of the swine-specific OTUs (i.e., 6 of 57) could not be classified beyond the order-level taxonomy, highlighting potentially novel *Bacteroidales* populations with swine-specific host distributions. Additionally, in most cases a high percentage of the *Bacteroidales* OTUs were completely unique to a given library, showing that sample representation may have a large impact on the observed population structure and that further sequencing studies from an even more diverse array of fecal and environmental samples is necessary for a comprehensive coverage of *Bacteroidales* diversity (see Fig. S2 in the supplemental material).

***In silico*** **evaluation of currently available swine-targeted assays.** Several swine-specific fecal markers that target the 16S rRNA gene of *Bacteroidales* populations have recently been published (6, 26, 42). However, the host specificity and host distribution of these targets have not been adequately assessed, and the utility of these assays in identifying swine fecal contamination on a broad geographic scale is largely unknown. A few recent studies have assessed the utility of these markers in environmental-monitoring scenarios, with generally poor results. For example, while 16S rRNA gene-based assays targeting swine *Bacteroidales* populations exhibit moderate levels of host specificity (i.e., 70 to 95% in animal fecal samples), the occurrence of 16S rRNA genes in environmental samples downstream of suspected swine inputs was poor in



Downloaded from http://aem.asm.org/ on January 27, 2016 by Missouri S&T

studies conducted in Europe and the United States (11, 43). In a study conducted in Japan, pig-specific *Bacteroidales* quantitative PCR assays amplified DNA extracted from cattle feces (26). This cross amplification can confound quantification of fecal loads from various sources, limiting resolution of the relative contribution of fecal sources. While these environmental studies are important in evaluating the utility of swine-specific markers, they offer no solution for improving the utility of these markers. Their poor performance can be explained in part by the small size of the swine fecal sequence libraries used to generate host-specific primers and the fact that some of the presumed host-specific populations are rather cosmopolitan in nature, which could not be assessed unless sequence depth and sampling coverage were significantly increased.

In order to evaluate currently available swine-specific markers, *in silico* searches were performed using this large library of *Bacteroidales* sequences. *In silico* searches for currently available swine-targeted markers provided evidence for their lack of host specificity and environmental detection (see Table S2 in the supplemental material). Specifically, the Bac1f/Bac1r primer set (26) matches sequences from several nonspecific fecal sources, including human, cattle, and wildlife sequences (see Table S2 in the supplemental material). Moreover, while the Bac2f/Bac2R markers showed high host specificity, they hit only two sequences derived from swine fecal waste-impacted waters, which might explain their poor performance against environmental water samples. The PF163 marker (6) hybridized to a few nonswine fecal sources but also matched several sequences derived from swine-contaminated waters, which may explain why assays with this marker have been projected to work better than other assays in different environmental-monitoring scenarios (11). However, the latter marker has shown both *in silico* and experimental host specificity problems. For example, sequence analyses using the mothur and QIIME pipelines revealed sequences shared by different hosts, including sequences that the Bac1, PF163, and Bac2 markers anneal to. The fact that these sequences fall within OTUs shared by multiple host types suggests that these markers target cosmopolitan *Bacteroidales* populations and therefore are not truly host specific. On the other hand, sequences unique to swine fecal sources and environments known to be impacted by swine feces were retrieved (see Table S3 in the supplemental material). Thus, studying the diversity patterns and phylogeny of *Bacteroidales* populations provided some evidence as to why the currently available markers are performing poorly in watershed-based studies and helped us uncover novel host-specific targets for the environmental detection of swine fecal pollution (see Fig. S1 in the supplemental material).

In conclusion, in this study we evaluated the molecular diversity of *Bacteroidales* populations within fecal sources and environmental matrices. Rarefaction curves indicated that additional sequencing targeting *Bacteroidales* members is needed to completely describe the molecular diversity of this bacterial taxon associated with most fecal sources and environmental samples. It is difficult to determine the extent of the sequencing effort that is needed to better estimate the diversity associated with different hosts, particularly when the sequencing effort does not completely capture the diversity within one sample type and the sequences analyzed do not account for differences in *Bacteroidales* temporal and spatial diversity in either host type (e.g., swine versus human) or fecal source type (swine feces versus swine waste lagoon). On the other hand, the results demonstrated the importance of understanding the distribution and occurrence of fecal source tracking targets within feces, waste management processes, and environmental waters. Our analysis revealed the identity of shared and swine-specific *Bacteroidales* populations, which may prove useful for source-specific assay development. It should be noted that identifying strictly host-specific populations that are shared among different sources of the same fecal type (e.g., swine feces versus waste lagoons) could be quite challenging, once the sequencing depth is significantly increased. In such cases, methods will depend on targeting populations that exhibit a preferential source distribution rather than true host specificity. Alternatively, different assays will be needed to detect fecal pollution when the sources are likely to be from processed fecal waste, as it is the case for poultry litter (44). Similarly, this study revealed a high diversity of *Bacteroidales* populations within swine feces and processed swine fecal waste, such as manure pits and waste lagoons, and uncovered several distinct swine-specific populations, suggesting that multiple targets are necessary for accurately assessing swine fecal pollution in watershed-based applications.

Future studies should focus on deeper sequencing efforts that include studying molecular diversity from more geographically diverse fecal sample sets, particularly from processed fecal waste, as considerably less is known about these fecal sources, even though manure is often used as fertilizer in farming operations. In our opinion, knowing the diversity of different *Bacteroidales* populations and other bacterial targets from different fecal source types will lead to a comprehensive understanding of their distribution in the environment and their true utility in fecal source tracking studies. Next-generation sequencing technologies will be useful on these efforts, as bar-coding approaches can increase the number of samples tested in a more in-depth, economical, and less time-consuming fashion (45, 46). However, conventional cloning and sequencing approaches will continue to be relevant, in light of the high sequencing error rates from next-generation technologies (47) and due to the need for generating databases composed of nucleic acid fragments $>$600 bp in length. As other bacterial groups are emerging as potential targets for human and animal fecal sources (25, 48), molecular surveys and diversity analyses will be needed to further validate their value within the MST toolbox.

## ACKNOWLEDGMENTS

R.L. was the recipient of a USEPA-UC research fellowship. This research was funded in part by a New Start Award from the USEPA National Center for Computational Toxicology to J.W.S.D. The U.S. Environmental Protection Agency, through its Office of Research and Development, funded and managed or partially funded and collaborated in the research described herein.

The study has been subjected to USEPA's administrative review and has been approved for external publication. Any opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the Agency; therefore, no official endorsement should be inferred. Any mention of trade names or commercial products does not constitute endorsement or recommendation for use.

## REFERENCES

1. **Santo Domingo JW, Bambic DG, Edge TA, Wuertz S.** 2007. Quo vadis source tracking? Towards a strategic framework for environmental monitoring of fecal pollution. Water Res. **41**:3539–3552.
2. **Fremaux B, Gritzfeld J, Boa T, Yost CK.** 2009. Evaluation of host-specific Bacteroidales 16S rRNA gene markers as a complementary tool for detecting fecal pollution in a prairie watershed. Water Res. **43**:4838–4849.

Downloaded from http://aem.asm.org/ on January 27, 2016 by Missouri S&T

3. **Layton A, McKay L, Williams D, Garrett V, Gentry R, Sayler G.** 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.
4. **Stapleton CM, Kay D, Wyer MD, Davies C, Watkins J, Kay C, McDonald AT, Porter J, Gawler A.** 2009. Evaluating the operational utility of a *Bacteroidales* quantitative PCR-based MST approach in determining the source of faecal indicator organisms at a UK bathing water. Water Res. **43**:4888–4899.
5. **Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, Schlegel ML, Tucker TA, Schrenzel MD, Knight R, Gordon JI.** 2008. Evolution of mammals and their gut microbes. Science **320**:1647–1651.
6. **Dick LK, Bernhard AE, Brodeur TJ, Santo Domingo JW, Simpson JM, Walters SP, Field KG.** 2005. Host distributions of uncultivated fecal *Bacteroidales* bacteria reveal genetic markers for fecal source identification. Appl. Environ. Microbiol. **71**:3184–3191.
7. **Kildare BJ, Leutenegger CM, McSwain BS, Bambic D, Rajal VB, Wuertz S.** 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.
8. **Lamendella R, Santo Domingo JW, Oerther DB, Vogel JR, Stoeckel DM.** 2007. Assessment of fecal pollution sources in a small northern-plains watershed using PCR and phylogenetic analyses of Bacteroidetes 16S rRNA gene. FEMS Microbiol. Ecol. **59**:651–660.
9. **U. S. Environmental Protection Agency.** 2005. Microbial source tracking guide document. EPA-600/R-05/064. Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC.
10. **Harwood VJ, Brownell M, Wang S, Lepo J, Ellender Ajidahun RD, Hellein A, Kennedy KNE, Ye X, Flood C.** 2009. Validation and field testing of library-independent microbial source tracking methods in the Gulf of Mexico. Water Res. **43**:4812–4819.
11. **Lamendella R, Santo Domingo JW, Yannarell AC, Ghosh S, Di Giovanni G, Mackie RI, Oerther DB.** 2009. Evaluation of swine-specific PCR assays used for fecal source tracking and analysis of molecular diversity of swine-specific *Bacteroidales* populations. Appl. Environ. Microbiol. **75**: 5787–5796.
12. **Badgley BD, Thomas FI, Harwood VJ.** 2010. The effects of submerged aquatic vegetation on the persistence of environmental populations of *Enterococcus* spp. Environ. Microbiol. **12**:1271–1281.
13. **Walk ST, Alm EW, Gordon DM, Ram JL, Toranzos GA, Tiedje JM, Whittam TS.** 2009. Cryptic lineages of the genus *Escherichia.* Appl. Environ. Microbiol. **75**:6534–6544.
14. **Schoen ME, Soller JA, Ashbolt NJ.** 2011. What faecal source(s) contribute to bather illness in human-impacted waters? Water Res. **45**:2670–2680.
15. **Horrigan L, Lawrence R, Walker P.** 2002. How sustainable agriculture can address the environmental and human health harms of industrial agriculture. Environ. Health Perspect. **110**:445–456.
16. **Jongbloed AW, Lenis NP.** 1998. Environmental concerns about animal manure. J. Anim. Sci. **76**:2641–2648.
17. **U.S. Environmental Protection Agency.** 2003. National pollutant discharge elimination system permit regulation and effluent limitation guidelines and standards for concentrated animal feeding operations (CAFOs). Fed. Regist. **68**:7176–7274.
18. **Anonymous.** 2004. Threatening Iowa's future: Iowa's failure to implement and enforce the Clean Water Act for livestock operations. Environmental Integrity Project, Washington, DC. http://www.environmentalintegrity.org/pdf/publications/Report_Threatening_Iowa_Future.pdf.
19. **Buckholt MA, Lee JH, Tzipori S.** 2002. Prevalence of *Enterocytozoon bieneusi* in swine: an 18-month survey at a slaughterhouse in Massachusetts. Appl. Environ. Microbiol. **68**:2595–2599.
20. **Fratamico PM, Bagi LK, Bush EJ, Solow BT.** 2004. Prevalence and characterization of Shiga toxin-producing *Escherichia coli* in swine feces recovered in the National Animal Health Monitoring System's Swine 2000 study. Appl. Environ. Microbiol. **70**:7173–7178.
21. **Guan TY, Holley RA.** 2003. Pathogen survival in swine manure environments and transmission of human enteric illness—a review. J. Environ. Qual. **32**:383–392.
22. **Chee-Sanford JC, Aminov RI, Krapac IJ, Garrigues-Jeanjean N, Mackie RI.** 2001. Occurrence and diversity of tetracycline resistance genes in lagoons and groundwater underlying two swine production facilities. Appl. Environ. Microbiol. **67**:1494–1502.
23. **Kellog RL, Lander CH, Moffitt DC, Gollehon N.** 2000. Manure nutrients relative to the capacity of cropland and pastureland to assimilate nutrients: spatial and temporal trends for the United States. U.S. Department of Agriculture, Washington, DC.
24. **Petraitis E.** 2007. Research into heavy metal concentrations in agricultural soils. Ekologija **53**:64–69.
25. **Newton RJ, Vandewalle JL, Borchardt MA, Gorelick MH, McLellan SL.** 2011. *Lachnospiraceae* and *Bacteroidales* alternative fecal indicators reveal chronic human sewage contamination in an urban harbor. Appl. Environ. Microbiol. **77**:6972–6981.
26. **Okabe S, Okayama N, Savichtcheva O, Ito T.** 2007. Quantification of host-specific *Bacteroides-Prevotella* 16S rRNA genetic markers for assessment of fecal pollution in freshwater. Appl. Microbiol. Biotechnol. **74**: 890–901.
27. **Bernhard AE, Field KG.** 2000. Appl. Environ. Microbiol. **66**:4571–4574.
28. **DeSantis TZ, Hugenholtz P, Keller K, Brodie EL, Larsen N, Piceno YM, Phan R, Andersen GL.** 2006. NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. Nucleic Acids Res. **34**:394–399.
29. **Huber T, Faulkner G, Hugenholtz P.** 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics **20**:2317–2319.
30. **DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL.** 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. **72**:5069–5072.
31. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.
32. **Ludwig W, Strunk O, Westram R, Richter L, Meier H, Kumar Y, Buchner A, Lai T, Steppi S, Jobb G, Förster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, König A, Liss T, Lüßmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A, Schleifer K-H.** 2004. ARB: a software environment for sequence data. Nucleic Acids Res. **32**:1363–1371.
33. **Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R.** 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods **7**:335–336.
34. **McCune B, Grace JB, Urban DL.** 2002. PCORD v5 software. MjM Software Design, Glenden Beach, OR.
35. **Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF.** 2009. Introducing mothur: open source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. **75**:7537–7541.
36. **Walters SP, Field KG.** 2009. Survival and persistence of human and ruminant-specific faecal *Bacteroidales* in freshwater microcosms. Environ. Microbiol. **11**:1410–1421.
37. **Bell A, Layton AC, McKay L, Williams D, Gentry R, Sayler GS.** 2009. Factors influencing the persistence of fecal *Bacteroides* in stream water. J. Environ. Qual. **38**:1224–1232.
38. **Jeter SN, McDermott CM, Bower PA, Kinzelman JL, Bootsma MJ, Goetz GW, McLellan SL.** 2009. *Bacteroidales* diversity in ring-billed gulls (*Laurus delawarensis*) residing at Lake Michigan beaches. Appl. Environ. Microbiol. **75**:1525–1533.
39. **Lamendella R, Santo Domingo JW, Ghosh S, Martinson J, Oerther DB.** 2011. Comparative fecal metagenomics unveils unique functional capacity of the swine gut. BMC Microbiol. **11**:103. doi:10.1186/1471-2180-11-103.
40. **Looft T, Johnson TA, Allen HK, Bayles DO, Alt DP, Stedtfeld RD, Sul WJ, Stedtfeld TM, Chai B, Cole JR, Hashsham SA, Tiedje JM, Stanton TB.** 2012. In-feed antibiotic effects on the swine intestinal microbiome. Proc. Natl. Acad. Sci. U. S. A. **109**:1691–1696.
41. **Wu GD, Chen J, Hoffmann C, Bittinger K, Chen Y-Y, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, Sinha R, Gilroy E, Gupta K, Baldassano R, Nessel L, Li H, Bushman FD, Lewis JD.** 2011. Linking long-term dietary patterns with gut microbial enterotypes. Science **334**: 105–108.


Downloaded from http://aem.asm.org/ on January 27, 2016 by Missouri S&T

42. **Mieszkin S, Furet JP, Corthier G, Gourmelon M.** 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.
43. **Gourmelon M, Caprais MP, Segura R, Le Mennec C, Lozach S, Piriou JY, Rince A.** 2007. Evaluation of two library-independent microbial source tracking methods to identify sources of fecal contamination in French estuaries. Appl. Environ. Microbiol. **73**:4857–4866.
44. **Weidhaas JL, Macbeth TW, Olsen RL, Harwood VJ.** 2011. Correlation of quantitative PCR for a poultry-specific *Brevibacterium* marker gene with bacterial and chemical indicators of water pollution in a watershed impacted by land application of poultry litter. Appl. Environ. Microbiol. **77**:2094–2102.
45. **McLellan SL, Huse SM, Mueller-Spitz SR, Andreishcheva EN, Sogin ML.** 2010. Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. Environ. Microbiol. **12**:378–392.
46. **Andersson AF, Lindberg M, Jakobsson H, Bäckhed F, Nyrén P, Engstrand L.** 2008. Comparative analysis of human gut microbiota by barcoded pyrosequencing. PLoS One **3**:e2836. doi:10.1371/journal.pone.0002836.
47. **Kunin V, Engelbrektson A, Ochman H, Hugenholtz P.** 2010. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. Environ. Microbiol. **12**:118–123.
48. **Ryu H, Griffith JF, Khan IU, Hill S, Edge TA, Toledo-Hernandez C, Gonzalez-Nieves J, Santo Domingo J.** 2012. Comparison of gull feces-specific assays targeting the 16S rRNA genes of *Catellicoccus marimammalium* and *Streptococcus* spp. Appl. Environ. Microbiol. **78**:1909–1916.

Downloaded from http://aem.asm.org/ on January 27, 2016 by Missouri S&T