

Assessing hog lagoon waste contamination in the Cape Fear Watershed using *Bacteroidetes* 16S rRNA gene pyrosequencing

Ann M. Arfken^{1,2} · Bongkeun Song^{1,2,3,4} · Michael A. Mallin^{1,2}

Received: 21 April 2015 / Revised: 15 June 2015 / Accepted: 18 June 2015 / Published online: 19 July 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Hog lagoons can be major sources of waste and nutrient contamination to watersheds adjacent to pig farms. Fecal source tracking methods targeting *Bacteroidetes* 16S rRNA genes in pig fecal matter may underestimate or fail to detect hog lagoon contamination in riverine environments. In order to detect hog lagoon wastewater contamination in the Cape Fear Watershed, where a large number of hog farms are present, we conducted pyrosequencing analyses of *Bacteroidetes* 16S rRNA genes in hog lagoon waste and identified new hog lagoon-specific marker sequences. Additional pyrosequencing analyses of *Bacteroidetes* 16S rRNA genes were conducted with surface water samples collected at 4 sites during 5 months in the Cape Fear Watershed. Using an operational taxonomic unit (OTU) identity cutoff value of 97 %, these newly identified hog lagoon markers were found in 3 of the river samples, while only 1 sample contained the pig fecal marker. In the sample containing the pig fecal marker, there was a relatively high percentage (14.1 %) of the hog lagoon

markers and a low pig fecal marker relative abundance of 0.4 % in the *Bacteroidetes* 16S rRNA gene sequences. This suggests that hog lagoon contamination must be somewhat significant in order for pig fecal markers to be detected, and low levels of hog lagoon contamination cannot be detected targeting only pig-specific fecal markers. Thus, new hog lagoon markers have a better detection capacity for lagoon waste contamination, and in conjunction with a pig fecal marker, provide a more comprehensive and accurate detection of hog lagoon waste contamination in susceptible watersheds.

Keywords Microbial source tracking · *Bacteroidetes* · Fecal Pollution

Introduction

Fecal contamination of water bodies imposes serious risks to human health and aquatic ecosystems. Fecal matter may harbor pathogenic microbes such as *Citrobacter freundii*, *Enterocytozoon bieneusi*, and some strains of *Escherichia coli* that may contaminate drinking water supplies (Meays et al. 2004). Fecal waste is also rich in nitrogen and phosphorus, which may lead to eutrophication of nearby lakes and rivers (Mallin and Cahoon 2003). Current standard practices of the Environmental Protection Agency (EPA) rely on enumeration and cultivation of fecal indicators to monitor contamination in recreational and public water sources (EPA 2003). Fecal indicator bacteria include total fecal coliforms, *E. coli*, and fecal enterococci (Okabe et al. 2007). Unfortunately, these enumeration methods do not allow for identification of the contamination source (Field et al. 2003; Scott et al. 2002). In addition, fecal contaminants such as enterococci and *E. coli* are able to survive and proliferate outside of a host (Desmarais et al.

Electronic supplementary material The online version of this article (doi:10.1007/s00253-015-6784-x) contains supplementary material, which is available to authorized users.

✉ Bongkeun Song
songb@vims.edu

¹ University of North Carolina Wilmington, 601 S. College Rd, Wilmington, NC, USA

² Center for Marine Sciences, University of North Carolina Wilmington, Wilmington, NC, USA

³ Department of Life Science, Dongguk University-Seoul, Seoul 100-715, South Korea

⁴ Present address: Department of Biological Sciences, Virginia Institute of Marine Science, College of William & Mary, Gloucester Point, VA, USA

2001), creating a dispute as to their accuracy as a pollution indicator (Scott et al. 2002; Simpson et al. 2002).

Over the last several years, rapid advancements in microbial source tracking (MST) have provided researchers with more discriminatory methods than traditional fecal indicator bacteria for determining host fecal contamination sources in aquatic environments (Roslev and Bukh 2011). Many of these methods rely on bacteria from phylum *Bacteroidetes* as alternative fecal indicators of water contamination (Hold et al. 2002; Leser et al. 2002; Krentz et al. 2013). *Bacteroidetes* are abundant in the intestines of warm-blooded animals, with a 1000-fold greater concentration than coliform bacteria (Ficksdal et al. 1985) and have shown to exhibit host specificity (Bernhard and Field 2000a, 2000b; Layton et al. 2006; Haugland et al. 2010). These bacteria are also obligate anaerobes that will not survive long outside of a host's intestinal tract (Ficksdal et al. 1985; Allsop and Stickler 1985). Their strict anaerobic and nutrient requirements make them a good indicator species to detect recent fecal contamination in water.

While *Bacteroidetes* have been shown to have a great deal of potential for MST, many methods have fallen short due to issues dealing with geographical differences in *Bacteroidetes* populations, multiple host or non-specificity of probes, and limited detection in the environment (Okabe et al. 2007; Mieszkin et al. 2009; Marti et al. 2011; Lamendella et al. 2013). The majority of *Bacteroidetes* 16S rRNA gene probes and primers used in the environment are designed from sequences obtained directly from fecal samples (Okabe et al. 2007; Dick et al. 2005) but not from the actual waste source contaminating the environment (Lamendella et al. 2013). The most significant contributors of wastewater to the environment are generally from treatment plants and septic tanks for human fecal waste, and manure storage and waste lagoons for animal waste (Lamendella et al. 2013). Sanapareddy et al. (2009) found that human-specific fecal bacteria made up only a small percentage of the total bacteria present in a municipal wastewater treatment plant, treating mostly human wastewater. Furthermore, Mieszkin et al. (2009) showed that the composition of *Bacteroidetes* 16S rRNA genes from stored lagoon waste can vary in composition from pig feces, while Lamendella et al. (2009) found that the detection limits for pig fecal markers were much lower in lagoon and manure pits than feces. MST methods that depend solely on indicators based on fecal origins may therefore underestimate or fail to detect contamination from sources such as stored fecal or manure waste in the riverine environment.

In North Carolina, hog lagoon waste, not directly deposited pig fecal waste, is the major contributor of waste contamination to local watersheds (Mallin et al. 1997). Hog waste lagoons are open-air basins comprised of swine fecal matter, urine, wash water, and rain. When the lagoons reach a certain level, the liquid supernatant is often sprayed onto surrounding fields. These lagoons and spray fields are subject to flooding,

stormwater runoff, and soil leaching resulting in lagoon and microbial contamination of downstream water bodies and nearby watersheds (Mallin and Cahoon 2003; Mallin 2000). Because lagoon waste mixtures sit for extended periods of time and undergo various stages of anaerobic decomposition, what is sprayed onto fields or leaked from lagoons is chemically and biologically altered from directly deposited feces.

To date, however, no studies have examined or compared potential *Bacteroidetes* hog lagoon indicators to watersheds in order to assess hog lagoon waste contamination. Lamendella et al. (2009, 2013) showed that *Bacteroidales* communities in hog lagoons were different from pig fecal communities and introduced the idea that multiple markers, using 16S rRNA gene next-generation sequencing, may be necessary to monitor and assess water bodies impacted by swine lagoons. As a result, pyrosequencing analysis of *Bacteroidetes* 16S rRNA genes may be a valuable alternative MST approach to detect different types of waste contamination in various water bodies.

In this present study, we examined both temporal and spatial variation of *Bacteroidetes* communities in the North East Cape Fear River and Black River Watersheds and compared them to local NC hog lagoon communities using 454 pyrosequencing techniques. Employing a strict 97 % identity cutoff for operational taxonomic unit (OTU) determination and clustering, we identified two potential hog lagoon markers. Using hog lagoon markers in conjunction with a pig fecal indicator provided a more comprehensive and accurate method for assessing hog lagoon waste contamination in North Carolina watersheds.

Materials and methods

Water quality measurements and sample locations Water quality measurements were conducted bi-monthly at 5 monitoring sites in the Northeast Cape Fear River Watershed and 4 monitoring sites in the Black River Watershed over a 1-year period from May 2009 to May 2010. The 5 sites in the Black River Watershed include Great Coharie Creek, Little Coharie Creek, Six Runs Creek, Colly Creek, and Black River at Highway 210. The 4 sites in the Northeast Cape Fear River Watershed include Panther Branch, Goshen Swamp, Sarecta, and Burgaw Creek 117 (Fig. 1 and Table S1).

Water samples were collected approximately 0.1 m below the surface in sterile plastic bottles provided by the contract laboratory and placed on ice for no more than 6 h before lab analysis (Mallin et al. 2010). Total nitrogen, total phosphorus, ammonium, and nitrate/nitrite were analyzed using APHA (1995) techniques. Fecal coliform bacteria were enumerated using membrane filtration (APHA 1995).

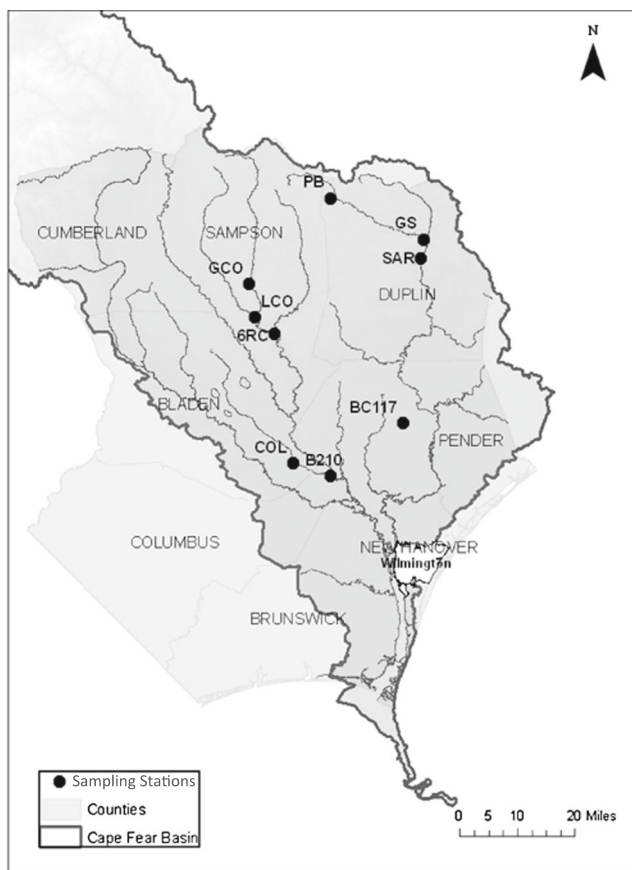


Fig. 1 Locations of river sample collection sites in the Cape Fear River Watershed, NC (PB, Panther Branch; GS, Goshen Swamp; SAR, Sarecta; BC117, Burgaw Creek 117; GCO, Great Coharie Creek; LCO, Little Coharie Creek; 6RC, Six Runs Creek; COL, Colly Creek; B210, Black River at Highway 210)

Sample selection and collection Among the 9 sampling and monitoring sites, 2 sites in the Northeast Cape Fear River (Panther Branch and Burgaw Creek 117) and 2 sites in the Black River (Six Runs Creek and Black River at Hwy 210) for the months of May, July, and September 2009 and March and May 2010 were selected for MST analyses. Sample site and month selections were based on site proximity to hog farms and concentrated animal feeding operations (CAFOs) and months with high numbers of fecal coliform numbers (Table S2A and S2B). Panther Branch is a stream station located downstream of approximately a dozen swine CAFOs, with one facility less than a km away. Six Runs Creek is on a 4th order stream that drains a watershed containing approximately 150 CAFOs, with none in the immediate vicinity, however. Burgaw Creek 117 is a site on a 2nd order stream with 4 CAFOs upstream of it, and site Black River at Hwy 210 is located on a 5th order Black River. This river has >300 swine CAFOs located in watersheds tributary to the Black River, but none immediately adjacent to site B210. These watersheds also contain a large but unknown number of poultry CAFOs that may influence the waterways.

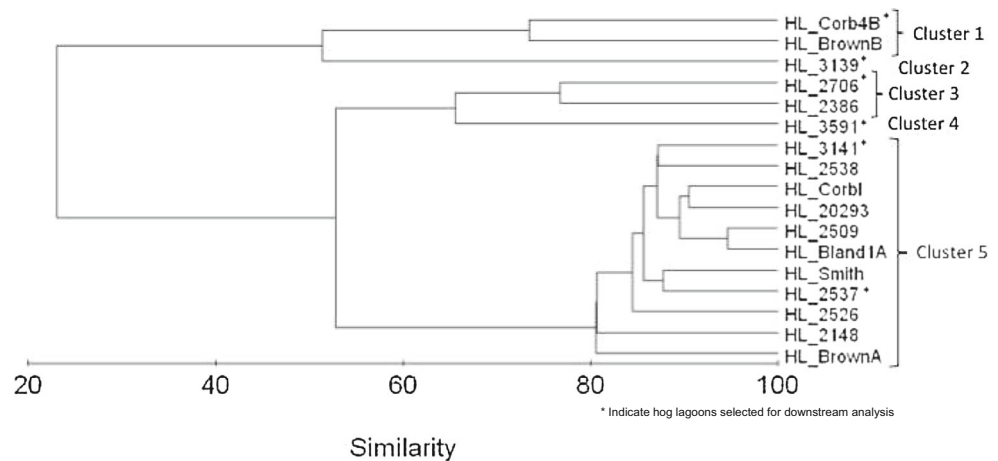
Water samples were collected in sterile 500-mL Pyrex glass bottles and stored on ice for transportation. Each sample was individually filtered on a sterile filter (Whatman GF/F 47 mm with pore size of 0.7 μm). The filters were wrapped in aluminum foil and stored at $-20\text{ }^{\circ}\text{C}$ until DNA was extracted. In addition, wastewater samples from 17 different hog lagoons were kindly provided by Murphy-Brown, LLC, a division of Smithfield Foods. Hog lagoon slurries (1.2 mL) were centrifuged at 10,000g for 5 min to concentrate biomass for DNA extraction.

DNA extraction Prior to DNA extraction, filters were cut in half using a sterilized razorblade. Environmental DNA was extracted from filters or hog lagoon pellets using the PowerSoil DNA Kit (Mo-bio Laboratories, Inc., Carlsbad, CA) following the manufacturer's protocol. All samples were disrupted using a Thermo Savant Fast Prep FP 120 Cell Disrupter (Qbiogene Inc., Carlsbad CA).

T-RFLP analysis *Bacteroidetes* 16S rRNA genes in hog lagoon waste were amplified using 6-FAM-labeled Bac32F and Bac708R primers as described by Bernhard and Field (2000a). PCR products were run on 1.0 % agarose using gel electrophoresis. Samples positive for the *Bacteroidetes* 16S rRNA gene displayed an amplified band at approximately 676 bp. Positive PCR products were gene cleaned using the UltraClean GelSpin DNA Purification Kit (Mo-Bio, Carlsbad, CA). DNA was quantified using the Quant-iT dsDNA Assay Kit, High Sensitivity (Invitrogen, Carlsbad, CA). A total of 20 ng of PCR product was digested overnight at 37 $^{\circ}\text{C}$ with 5 units of *Acil* restriction endonuclease (New England Biolabs, Ipswich MA). The restriction enzyme *Acil* was selected based on a previous study (Savichtcheva and Okabe 2009). Digested products were precipitated with isopropanol and run on a 3130 \times /Genetic Analyzer (Applied Biosystems, Carlsbad CA). Fragment analysis was conducted using the Gene Mapper 4.0 software (Applied Biosystems, Carlsbad CA). Based on a similarity dendrogram constructed from T-RFLP fingerprints (Fig. 2), five clusters were identified, with a cluster being defined as greater than 75 % similarity. From each cluster, one representative hog lagoon sequence was selected, with the exception of cluster 5, in which two representative sequences were selected. All six representative hog lagoon wastewater samples were then used for downstream pyrosequencing analysis of the *Bacteroidetes* 16S rRNA gene.

Quantitative PCR Quantitative PCR of the general *Bacteroidetes* 16S rRNA gene was conducted for the river samples using the 7500 Fast Real-Time PCR system (Applied Biosystems, Carlsbad, CA) with the primers Bac32F (5'-AACGCTAGCTACAGGCTTAACA-3') and Bac404R (5'-CAATATTCCTCACTGCTGCCTCCCGTA-

Fig. 2 Hog lagoon similarity dendrogram and clusters I, II, and III based on T-RFLP fingerprints using *Bacteroidetes* 16S rRNA genes cut with *AciI*



3') described by (Dick and Field 2004). Standards were generated from clone libraries constructed from amplification of the *Bacteroidetes* 16S rRNA fragment using the primers described above (Perfect Prep Cloning Kit, 5 Prime, Gaithersburg, MD). Plasmid extraction was conducted on a positively identified *Bacteroidetes*-specific 16S rRNA gene clone using the Zyppy Plasmid Miniprep Kit (Zymo, Irvine, CA) and purified using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI). A tenfold serial dilution of the purified plasmid was used for quantification standards. Quantitative PCR (qPCR) efficiency was 91.7 %, calculated from the slope (-3.537) of the liner standard curve ($r^2 = 0.998$). Assays were carried out in a volume of 20 μ L containing 1.0 ng of template DNA and SYBR green using GoTaq qPCR Master Mix (Promega Corporation, Madison, WI). Reactions were carried out in triplicates, and PCR specificity was monitored by analysis of dissociation curves. Amplification calculations based on the standard curve slopes were performed using ABI PRISM 7000 sequence detection software (Applied Biosystems, Carlsbad, CA).

16S rRNA gene pyrosequencing Hypervariable V1 and V2 regions on the *Bacteroidetes* 16S rRNA gene were selectively amplified from river and hog lagoon DNA using GoTaq Master mix (Promega, Fitchburg WI) using modified *Bacteroidetes* qPCR primers Bac32F (Bernhard and Field 2000a) and Bac404R (Dick and Field 2004). The Bac32F primers fused with the 454 A-adaptor sequence, and 8-nucleotide barcodes were used to multiplex individual samples. Triplicate reactions for each sample were amplified following the PCR protocol: initial denaturation at 95 $^{\circ}$ C for 5 min, followed by 40 cycles of 94 $^{\circ}$ C for 30 s, 55 $^{\circ}$ C for 30 s, and 72 $^{\circ}$ C for 2 min. To ensure complete amplification, an extension time of 10 min at 72 $^{\circ}$ C was added. The amplified products were gene cleaned using the UltraClean GelSpin DNA Purification Kit (Mo-Bio, Carlsbad, CA). The resulting amplicon libraries were then amplified following Roche's

emulsion PCR protocol (454 Life Sciences, Branford, CT, USA). Samples were processed using the Roche 454 GS Junior Titanium Series platform.

Quality control and OTU assignment Raw pyrosequences from both hog lagoon waste and river water were initially denoised and filtered using Acacia (Bragg et al. 2012) with a minimum quality score of 30, maximum k-mer distance between reads of 13 and maximum of 2 standard deviations from mean length. Trimming of barcodes, primers, and sequence length was conducted using the RDP pipeline initial process (<http://rdp.cme.msu.edu>) with a minimum quality score of 30 and minimum length of 200 bases (Cole et al. 2014). Mothur (Schloss et al. 2009) was used to combine the processed sequences, which were aligned to the SILVA template, pre-clustered, and screened for chimeric sequences using uchime (Edgar et al. 2011). The resulting high-quality sequences were clustered into OTUs based on 97 % identity. A representative sequence for each OTU was assigned a taxonomic identity using Silva taxonomy and the Wang classification method (Wang et al. 2007) (<http://rdp.cme.msu.edu>) with an 80 % minimum identity.

Hog lagoon marker and reference sequences analyses A general *Bacteroidetes* 16S rRNA gene reference database determining host specificity was constructed from hog lagoon markers determined from hog lagoon sequences in this study and from *Bacteroidetes* host-specific fecal derived markers described in Boehm et al. (2013): Rum2Bac (ruminant) (Mieszkin et al. 2010), CF193 (ruminant) (Bernhard and Field 2000b), BacCan (dog) (Kildare et al. 2007), HoF597 (horse) (Dick et al. 2005), BacH (human) (Reischer et al. 2007) and HF183 (human) (Bernhard and Field 2000b), and PF163 (pig) (Dick et al. 2005). To construct a 16S rRNA gene database for hog lagoon waste specificity in this study, 8379 high-quality, trimmed, and processed 16S rRNA gene pyrosequences from 6 hog lagoon samples were clustered into 335 distinct OTUs based on 97 % sequence identity using the

Mothur package (Schloss et al. 2009). A representative sequence for each OTU was then compared to reference sequences obtained from the BLAST database. Two hog lagoon-specific markers were determined from reference BLAST sequence sources based on dominance in hog lagoon samples (OTU comprised >10 % of total hog lagoon sequences and was found in at least five out of six hog lagoon samples) and a ≥ 99 % identity to hog lagoon waste reference sequences only, with e values $\leq 8 \times 10^{-176}$. Representative OTU hog lagoon sequences that matched with hog lagoon waste and other non-hog lagoon waste reference sequences were not considered specific. The reference database containing *Bacteroidetes* host-specific markers was then compared with river sequences using the Mothur pipeline (Schloss et al. 2009). To create a neighbor-joining tree of the reference database containing *Bacteroidetes* host-specific sequences, sequences were first aligned within MEGA (www.megasoftware.net) using muscle alignment, and trees were constructed using the bootstrapped kimora-2-parameter method (Tamura et al. 2011).

Diversity and statistical analyses T-RFLP fingerprints were analyzed using T-RFLP Analysis Expedited (T-Rex) software (Culman et al. 2009, <http://trex.biohpc.org/>). Variations in *Bacteroidetes* 16S rRNA genes were assessed using a Bray-Curtis similarity matrix, which was then used to perform a similarity dendrogram analysis in Primer-5 software package (Primer-E Ltd., Luton, UK) to visualize variations in the communities. Pyrosequencing statistical analyses were performed using Mothur and Mac StatPlus software. OTU clustering and identification for diversity analyses were conducted based on 97 % sequence identity. Diversity, species richness, and coverage of *Bacteroidetes* communities in river samples were estimated using Ace, Chao1, Shannon, Simpson, and Observed Species metrics. Nutrients and fecal coliform data were tested for normality and subsequently log-transformed. Parametric Pearson's correlation coefficients were calculated to assess significant relationships between the abundance of *Bacteroidetes* 16S rRNA genes and fecal coliform data, diversity, and environmental parameters.

Nucleotide sequence accession numbers Sequences for each pyrosequencing library obtained in this study have been deposited in the NCBI Sequence Read Archive (Table S3).

Results

Fecal coliform enumeration and water quality measurements Fecal coliform counts and water quality/nutrient parameters (TN, PO₄, NH₄, NO_x) were reported for sites in the Northeast Cape Fear River (Panther Branch and Burgaw Creek 117) and in the Black River (Six Runs Creek and

Black River at Hwy 210) during the selected months of May 2009, July 2009, September 2009, March 2010, and May 2010 (Table S2A and S2B). Fecal coliform counts varied widely among individual sampling sites and sampling months but were overall higher at the Northeast Cape Fear River sites than in the Black River. The highest count (10×10^3 CFU 100 mL⁻¹) was found at site Burgaw Creek 117 during May 2009. With the exception of samples from March 2010 at site Panther Branch and March and May 2010 at site Burgaw Creek 117, both Northeast Cape Fear River sites had coliform counts above the North Carolina water quality standard of 200 CFU 100 mL⁻¹. It should be noted that both Panther Branch and Burgaw Creek 117 receive inputs from industrial and municipal wastewater treatment facilities as well as non-point source fecal pollution. In comparison, none of the samples collected at the Black River sites exceeded the North Carolina water quality standard for fecal coliform counts. In addition, total nitrogen, phosphorus, ammonium, and nitrate/nitrite levels were consistently higher in the Northeast Cape Fear River compared to the Black River.

Variation of *Bacteroidetes* communities in hog lagoon samples

Based on T-RFLP fingerprint and T-Rex analysis, 16S rRNA gene similarity of *Bacteroidetes* communities among the hog lagoon samples was >50 % in all 17 samples (Fig. 2) and grouped into five distinct clusters, using a criteria of >75 % similarity. One hog lagoon sample from clusters 1 to 4 (Corb4B from cluster 1, 3139 from cluster 2, 2706 cluster 3, and 3591 from cluster 4) and 2 hog lagoon samples from cluster 5 (3141 and 2537) were selected for pyrosequencing analysis of *Bacteroidetes* 16S rRNA genes.

Abundance and diversity of *Bacteroidetes* in river samples

A total of 17,721 trimmed and high-quality *Bacteroidetes* 16S rRNA gene sequence reads were obtained from 20 river samples. Sequence depth ranged from 603 to 1344 reads with an average sequence length of 350 bp and an average number of 886 sequences per sample. Within the *Bacteroidetes* phylum, 7 families from the river samples were identified, in addition to an unidentified *Bacteroidetes* group (Table S4). The majority of sequence reads from the river samples belonged to either *Bacteroidaceae* (38.2 %) or *Prevotellaceae* (35.6 %). A percentage of 20.1 of river sample sequences fell into the unidentified *Bacteroidetes* group while the other families *Porphyromonadaceae*, *Chitinophagaceae*, *Flavobacteriaceae*, *Cryomorphaceae*, and *Bacteroides* S24-7 made up less than 6.2 % of sequences in the river samples.

Species richness at 97 % identity among the samples was demonstrated in rarefaction plots (Fig. 3). The Chao1 and Ace indices reported the richness, and the Simpson and Shannon indices reported diversity of *Bacteroidetes* communities in each sample (Table 1). Many of the river rarefaction curves are close to approaching horizontal asymptotes at 600 sequences, indicating that adequate coverage of the *Bacteroidetes* diversity was captured in these analyses. In addition, sequence coverage estimates for each sample ranged from 88.2 to 99.9 %, with an average coverage of 95.4 %. Based on Chao1, Simpson, Ace, and Shannon indices, site Panther Branch showed the most temporal variation in diversity and richness with relatively low diversity and richness during May and July 2009 and increased richness and diversity during March and May 2010. In comparison, site Six Runs Creek had the least temporal variation in richness diversity and consistently had the highest levels of richness and diversity.

Total number of *Bacteroidetes* 16S rRNA gene copy numbers L^{-1} ranged from 5.7×10^4 100 mL^{-1} at site Panther Branch July 2009 to 1.2×10^7 100 mL^{-1} at site Panther Branch May 2009 (Fig. 4). The highest average abundances and variation of *Bacteroidetes* were found in the Northeast Cape Fear River, while lower average abundances and variation were found in the Black River. Abundance of *Bacteroidetes* 16S rRNA genes correlated negatively and

significantly with Shannon Diversity ($r = -0.564$, $p < 0.05$) and positively and significantly correlated with the relative abundance of family *Prevotellaceae* ($r = 0.401$, $p < 0.05$).

Relative abundances of host-specific phylotypes in hog lagoons and rivers

Relative abundances of host-specific *Bacteroidetes* 16S rRNA markers were calculated by dividing the number of marker sequences by the total number of *Bacteroidetes* sequences per sample or samples. Based on sequence analyses, 44.9 % of hog lagoon pyrosequences were comprised of 2 major hog lagoon markers, hog lagoon marker 1 (34.0 %) and hog lagoon marker 2 (10.9 %) (Table 2 and Fig. 5); both markers belonged to an unidentified *Bacteroidetes* group. Hog lagoon markers were ≥ 99 % similar to the NCBI reference sequences obtained from hog lagoon or anaerobic digesters treating swine waste. The 2 hog lagoon markers accounted for 19.9 to 70.7 % of total pyrosequences in hog lagoon samples and were the dominant OTUs in all but 1 hog lagoon. In comparison, the pig fecal marker (PF163) encompassed a relatively small percentage (1.3 %) of total pyrosequences in the hog lagoon samples, ranging from undetected to 2.4 %. No other host-specific fecal markers (human, dog, ruminant, or horse) were detected in the hog lagoon samples.

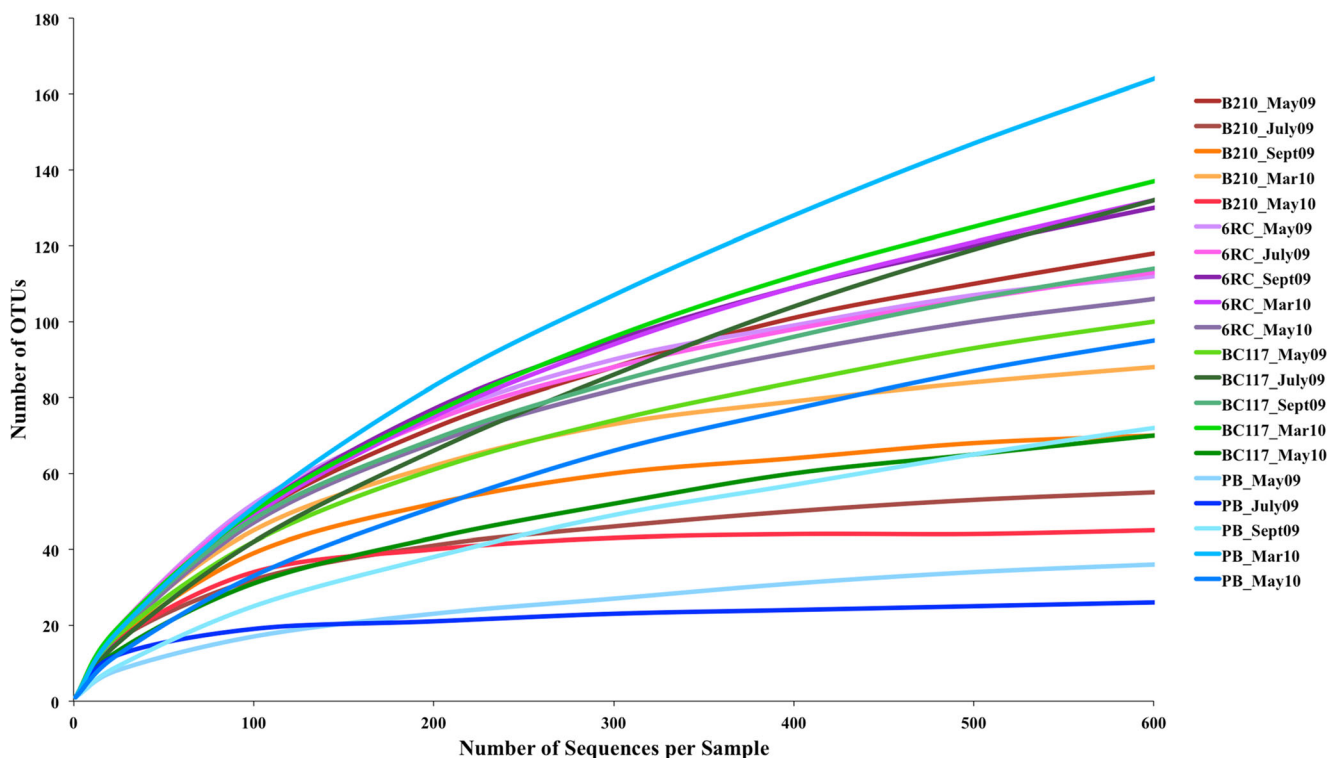


Fig. 3 *Bacteroidetes* rarefaction curve for river water samples based on 600 sequences. Rarefaction curves were generated using MOTHR software and an OTU identification of ≥ 97 % sequence identity (B210,

Black River at Hwy 210; 6RC, Six Runs Creek; BC117, Burgaw Creek 117, PB, Panther Branch)

Table 1 Diversity and richness metrics for river samples^a

		Total number of sequences	Total number of OTUs	Coverage (%)	Chao1 ^b	Simpson ^b	Ace ^b	Shannon ^b
Panther Branch	May 2009	940	42	98.6	63.2000	9.2310	67.4782	2.0261
	July 2009	691	26	99.4	25.7500	3.8826	35.9261	2.7453
	Sept 2009	1190	103	95.5	114.1538	11.9188	258.5712	2.2994
	Mar 2010	1061	227	88.2	366.0000	4.5830	576.1828	4.1337
	May 2010	961	119	94.4	169.0588	24.7316	204.3617	2.9779
Burgaw Creek 117	May 2009	800	111	95.1	148.0476	24.6908	138.8267	3.7600
	July 2009	784	154	89.0	243.0000	9.0030	357.0855	3.4386
	Sept 2009	979	136	95.1	200.0588	34.0987	166.3734	3.9936
	Mar 2010	1079	179	93.0	220.0385	35.1320	298.7611	4.1110
	May 2010	729	75	96.4	81.8333	9.2310	147.4652	3.0673
Six Runs Creek	May 2009	880	125	96.1	132.6250	36.3765	127.6622	4.0970
	July 2009	603	113	93.9	152.1765	46.8701	146.2581	4.1627
	Sept 2009	986	157	94.3	160.3846	20.5442	181.6527	3.9229
	Mar 2010	897	158	92.3	196.2143	27.5825	302.8024	4.0190
	May 2010	892	119	96.5	119.1818	16.8085	133.0007	3.8128
Black River Hwy 210	May 2009	1344	150	97.1	162.4000	36.8005	155.5283	4.0547
	July 2009	770	59	97.9	68.0000	23.3690	83.4241	3.3826
	Sept 2009	701	72	98.1	78.8000	21.0869	76.5794	3.5672
	Mar 2010	708	91	97.5	96.0526	34.3948	101.3181	3.9001
	May 2010	726	45	99.9	45.3330	24.7797	44.5141	3.4214

^aOTU identification based on $\geq 97\%$ sequence identity

^bBased on 600 sequences

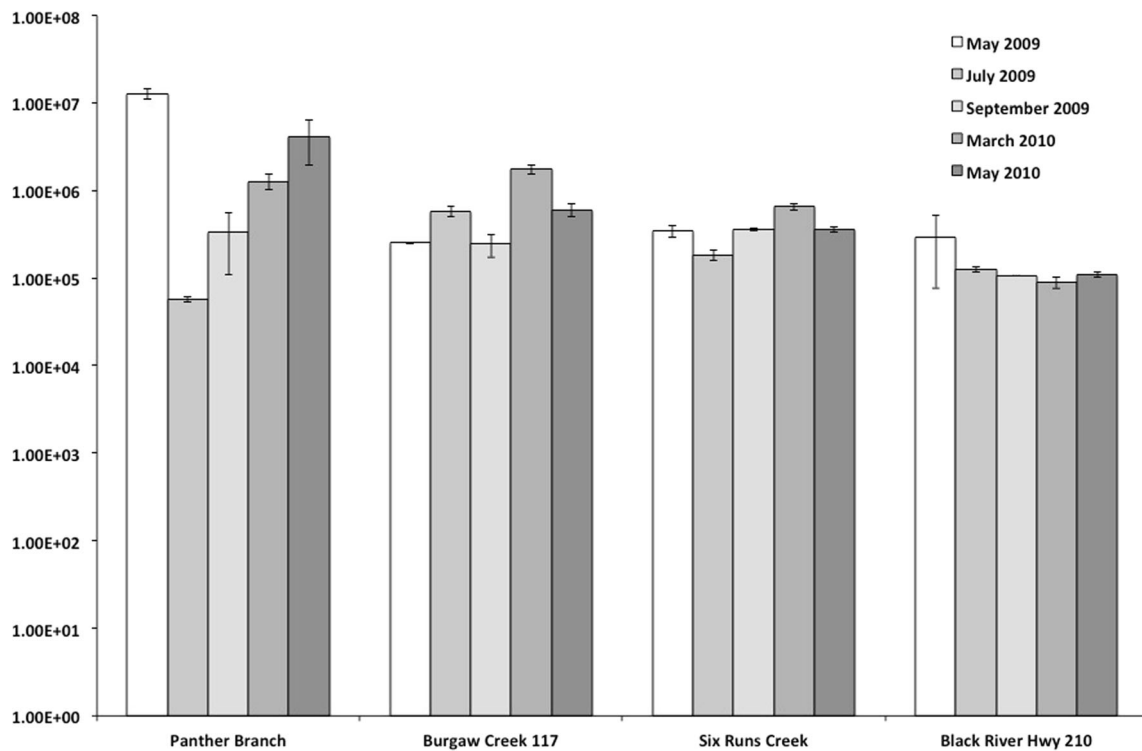


Fig. 4 General *Bacteroidetes* 16S rRNA gene copy numbers per liter of river water. Error bars represent ± 1 standard deviation

Table 2 Relative percent abundances of *Bacteroidetes* markers in hog lagoon samples

	Hog lagoon 1 (HLW1)	Hog lagoon 2 (HLW2)	Pig (PF163)	Human (HF183)	Dog (BacCan)	Ruminant (Rum2Bac/CF193)	Horse (HoF597)
HL2537	31.3	0	1.1	0	0	0	0
HL2706	17.1	2.8	0.1	0	0	0	0
HL3139	22.9	47.8	2.4	0	0	0	0
HL3141	41.4	19.7	2.3	0	0	0	0
HL3591	57.8	8.4	0	0	0	0	0
HLCORB4B	35.2	0.1	2.2	0	0	0	0

In the river samples, host-specific markers were found in 7 out of 20 samples showing contamination (Table 3). Hog lagoon markers were present in 3 of the river samples (Panther Branch March 2010, Burgaw Creek 117 September 2009, and Black River at Hwy 210). Black River at Hwy 210 March 2010 was the only sample in this study that showed a relatively high percentage (14.1 %) of the hog lagoon markers, as well as the only sample that contained the pig fecal marker (PF163) with an abundance of 0.4 %. Host-specific fecal markers for human (HF183) were found only at site Burgaw Creek 117 during the months of May 2009 (4.0 %) and March 2010 (0.2 %). Dog-specific fecal markers (BacCan) were detected at site Panther Branch May 2010 and at site Burgaw Creek 117 May 2009 for all months except September 2009. No host-specific fecal markers were found at site Six Runs Creek for all months sampled.

Discussion

Pyrosequencing of the *Bacteroidetes* 16S rRNA gene allows for a more comprehensive analysis regarding the monitoring of potential occurrences of hog lagoon and host-specific contamination in river systems. As shown in this study, pig-specific sequences are a minor component of stored hog lagoon waste and when measured in isolation, may not necessarily be indicative of hog lagoon waste. By analyzing pyrosequencing results and relative abundances of a well-established pig-specific marker and hog lagoon *Bacteroidetes* markers in this study, we were able to better examine the overall profile of waste contamination in rivers and make more accurate determinations of true hog lagoon contamination. Our criterion to use a ≥ 97 % sequence identity cutoff for pig, hog lagoon and other host-specific fecal

Fig. 5 Phylogenetic tree of identified hog lagoon markers and host-specific fecal markers based on *Bacteroidetes* 16S rDNA sequences obtained from GenBank and corresponding clades: Rum2Bac (ruminant), CF193 (ruminant), BacCan (dog), HoF597 (horse), BacH and HF183 (human), and PF163 (pig). Bootstrap values are given based on 1000 replicates

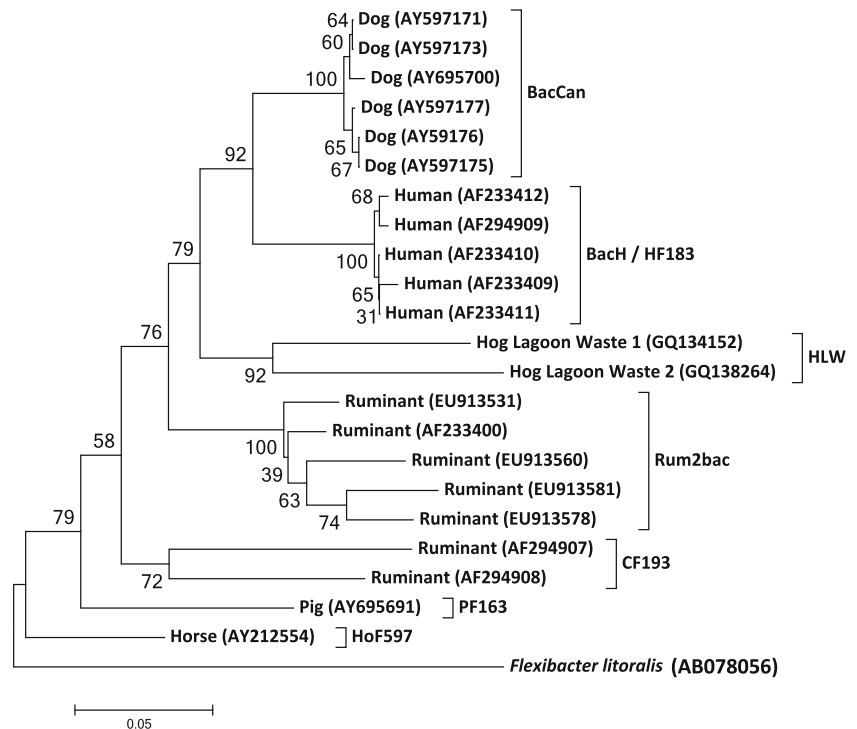


Table 3 Relative percent abundances of *Bacteroidetes* markers in river samples

		Hog lagoon 1 (HLW1)	Hog lagoon 2 (HLW2)	Pig (PF163)	Human (HF183)	Dog (BacCan)	Ruminant (Rum2Bac/CF193)	Horse (HoF597)
Panther Branch	May 2009	0	0	0	0	0	0	0
	July 2009	0	0	0	0	0	0	0
	Sept 2009	0	0	0	0	0	0	0
	Mar 2010	0.6	0	0	0	0	0	0
	May 2010	0	0	0	0	0.1	0	0
Burgaw Creek 117	May 2009	0	0	0	4.0	11.9	0	0
	July 2009	0	0	0	0	0.8	0	0
	Sept 2009	0.4	0	0	0	0	0	0
	Mar 2010	0	0	0	0.2	0.1	0	0
	May 2010	0	0	0	0	0	0	0
Six Runs Creek	May 2009	0	0	0	0	0	0	0
	July 2009	0	0	0	0	0	0	0
	Sept 2009	0	0	0	0	0	0	0
	Mar 2010	0	0	0	0	0	0	0
	May 2010	0	0	0	0	0	0	0
Black River at HWY 210	May 2009	0	0	0	0	0	0	0
	July 2009	0	0	0	0	0	0	0
	Sept 2009	0	0	0	0	0	0	0
	Mar 2010	9.6	4.5	0.4	0	0	0	0
	May 2010	0	0	0	0	0	0	0

markers provided for a more detailed and critical method of determining host specificity and reduced the likelihood of erroneously assessing hog lagoon contamination.

Bacteroidetes 16S rRNA genes from 6 local North Carolina hog lagoons were initially pyrosequenced to determine the *Bacteroides* communities in hog lagoon waste. Hog lagoon waste was selected over pig fecal samples because the vast majority of swine waste in the Cape Fear River Watershed is generated within CAFOs (Mallin et al. 1997). Lamendella et al. (2013) showed that the composition of *Bacteroidetes* 16S rRNA gene sequences in pig feces was different than hog lagoon waste; *Bacteroidetes* from hog lagoons clustered more closely with “swine-impacted water” than did those from pig fecal or manure pit.

In our study, only a small percentage (<1.3 %) of pyrosequences from hog lagoon samples was identified with the pig fecal marker (PF163), while a large percentage of pyrosequences (44.9 %) identified with anaerobic digesters treating swine waste (Fig. 5 and Table 2). The low numbers of pig fecal markers in hog lagoon waste support the findings from Lamendella et al. (2009), which found hog lagoons and manure pits to have higher detection limit requirements than raw feces. The low abundance of pig-specific fecal sequences provides further evidence that *Bacteroidetes* communities undergo a dramatic shift from raw fecal matter to stored hog lagoon waste and suggests that pig-specific markers of fecal

origin may have poor survival rates in lagoon conditions (Lamendella et al. 2009). In addition, our study of North Carolina hog lagoons calculated similar levels of *Bacteroidetes* diversity and species richness for North Carolina hog lagoons as those studied in Lamendella et al. (2013). *Bacteroidetes* communities in hog lagoons may experience the same pressures or environmental factors, which affect community structure (Lamendella et al. 2013). Interestingly, however, none of the dominant hog lagoon markers found in this study was ≥ 97 % identical to those found in Lamendella et al. (2013). Instead, the hog lagoon markers in the study were more than >99 % similar to sequences obtained from the NCBI database related to anaerobic sequencing batch reactors treating swine waste. This suggests that while the hog lagoon environment may similarly affect the diversity of lagoon *Bacteroidetes* communities, the community members themselves may remain regionally and geographically distinct.

As expected, the percentages of hog lagoon markers were generally low in the river water samples due to the natural dilution of hog lagoon waste in the watershed and mixing of other contamination sources (Table 3). Exposure to various environmental factors such as predation, temperature, and sunlight may also affect the decay rate of markers (Krentz et al. 2013). Different species of *Bacteroides* have been shown to have variable rates of decay from exposure to different

environmental parameters and seasonality (Ballesté and Blanch 2010). Additionally, river water in this study was filtered through 0.7- μm pore size of glass-fiber filters, which may have allowed some *Bacteroidetes* populations to pass through filtration.

Overall, however, the hog lagoon markers were detected more frequently than pig-specific fecal markers and at a much higher concentration when both markers were present. Out of the 20 river samples, only 3 (Panther Branch March 2010, Burgaw Creek 117 September 2009, and Black River at Hwy 210 March 2010) had evidence of hog lagoon markers (Fig. 5 and Table 3). Of these samples, only Black River Hwy 210 March 2010, showed evidence of both hog lagoon markers 1 and 2, as well as relatively high levels of the combined hog lagoon markers (14.1 %). In addition, it was the only sample that had evidence of the pig-specific fecal marker (PF163), albeit at very low concentrations (0.4 %). This suggests that hog lagoon contamination must be somewhat significant in order for pig fecal markers to be detected, whereas low levels of hog lagoon contamination may still be present and detectable when pig-specific fecal markers are not.

In terms of host-specific fecal contamination, site Burgaw Creek 117 appears to be the most contaminated of the sample sites (Fig. 2b). Dog-specific fecal markers were detected for all months sampled except September 2009, while human-specific fecal markers were detected in both May 2009 and March 2010. In addition, Burgaw Creek 117 had the highest overall levels of fecal coliform counts and nutrient levels of all the sites sampled. Contamination at this site is most likely a result of site Burgaw Creek 117 being located downstream from the Town of Burgaw wastewater treatment plant, which has historically suffered from numerous failures and incomplete treatment of its sewage. The occasional presence of human fecal contamination at Burgaw Creek 117 (Table 3) is likely a result of those failures, and the very high nitrogen and phosphorus concentrations may be a result of improper nutrient treatment of the human sewage. This treatment plant was subsequently taken off-line in late 2012, piping the town's discharge to a treatment plant in another town. In addition, the frequent presence of dog fecal contamination at this site is likely a result of stormwater runoff carrying domestic canine waste into open runoff ditches that also drain to the sampling site.

The overall examination of the relative abundance and composition of the *Bacteroidetes* 16S rRNA gene in hog lagoon and river samples using pyrosequencing techniques provided a comprehensive insight into the *Bacteroidetes* community structure of these environments. From these analyses we were able to identify 2 distinct hog lagoon markers abundant in hog lagoon samples. Subsequently this allowed for a more detailed analysis of the *Bacteroidetes* community composition in river systems. With the steady decrease in the cost of next-generation sequencing and the amount of diversity present in

wastewater and feces, these types of analyses for MST may prove more comprehensive assessment of fecal and wastewater contamination in aquatic ecosystems.

Acknowledgments This work was supported by the North Carolina Water Resources Research Institute (#2006-1467-04) and the Lower Cape Fear River Program. We would also like to acknowledge M.R. McIver and M.I. Haltom for field and laboratory assistance, L.B. Cahoon for advice, and S. Barrett for assistance with statistical analyses.

Compliance with ethical standards None of the authors has any financial or personal relationships that could inappropriately influence or bias the content of the paper. There is no involvement of human participants and animals in this study.

References

- Allsop K, Stickler DJ (1985) An assessment of *Bacteroides fragilis* group organisms as indicators of human faecal pollution. *J Appl Bacteriol* 58:95–99
- APHA (1995) Standard methods for the examination of water and wastewater, 19th edn. American Public Health Association, Washington, DC
- Ballesté E, Blanch AR (2010) Persistence of *Bacteroides* species populations in a river as measured by molecular and culture techniques. *Appl Environ Microbiol* 76:7608–7616
- Bernhard AE, Field KG (2000a) Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16s rDNA markers from fecal anaerobes. *Appl Environ Microbiol* 66:1587–1594
- Bernhard AE, Field KG (2000b) A PCR assay to discriminate human and ruminant feces on the basis of host differences in *Bacteroides-Prevotella* gene encoding 16S ribosomal RNA. *Appl Environ Microbiol* 66:4571–4574
- Boehm AB, Van de Werfhorst LC, Griffith JF, Holden PA, Jay JA, Shanks OC, Wang D, Weisberg SB (2013) Performance of forty-one microbial source tracking methods: a twenty-seven lab evaluation study. *Water Res* 47:6812–6828
- Bragg L, Stone G, Imelfort M, Hugenholtz P, Tyson GW (2012) Fast, accurate error-correction of amplicon pyrosequences using Acacia. *Nat Methods* 9:425–426
- Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM (2014) Ribosomal database project: data and tools for high throughput rRNA analysis. *Nucl Acids Res* 42:D633–D642
- Culman SW, Bukowski R, Gauch HG, Cadillo-Quiroz H, Buckley DH (2009) T-REX: software for the processing and analysis of T-RFLP data. *BMC Bioinf* 10:171
- Desmarais TR, Solo-Gabriel HM, Palmer CJ (2001) Influence of soil on fecal indicator organisms in a tidally influenced subtropical environment. *Appl Environ Microbiol* 68:1165–1172
- Dick LK, Field KG (2004) Rapid estimation of numbers of fecal *Bacteroidetes* by use of a quantitative PCR assay for 16S rRNA genes. *Appl Environ Microbiol* 70:5695–5697
- 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* 73:4857–4866
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27:2194–2200

- Environmental Protection Agency (EPA) (2003) Bacterial water quality standards for recreational waters (freshwater and marine waters). United States Environmental Protection Agency Office of Water. EPA/823/R-03/008
- Ficksdal L, Maki JS, LaCroix SJ, Staley JT (1985) Survival and detection of *Bacteroides* spp., prospective indicator bacteria. Appl Environ Microbiol 49:148–150
- Field KG, Bernhard AE, Brodeur TJ (2003) Molecular approaches to microbiological monitoring: fecal source detection. Environ Monit Assess 81:313–326
- Haugland RA, Varma M, Sivaganesan M, Kelty C, Peed L, Shanks OC (2010) Evaluation of genetic markers from the 16S rRNA gene V2 region for use in quantitative detection of selected *Bacteroidales* species and human fecal waste by QPCR. Syst Appl Microbiol 33:348–357
- Hold GL, Pryde SE, Russell VJ, Furrrie E, Flint HJ (2002) Assessment of microbial diversity in human colonic samples by 16S rDNA sequence analysis. FEMS Microbiol Ecol 39:33–39
- Kildare BJ, Leutenegger CM, McSwain BS, Bambi DG, 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
- Krentz CA, Prystajecy N, Isaac-Renton J (2013) Identification of fecal contamination sources in water using host-associated markers. Can J Microbiol 59:210–220
- Lamendella R, Santo Domingo JW, Yannarell AC, Ghosh S, Giovani GD, 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
- Lamendella R, Li KC, Oerther D, Santo Domingo JW (2013) Molecular diversity of *Bacteroidales* in fecal and environmental samples and swine-associated subpopulations. Appl Environ Microbiol 79:816–824
- Layton A, McKay L, Williams D, Garrett V, Gentry R, Saylor 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
- Leser TD, Amenuvor JA, Jensen TK, Lindecrone RH, Boye M, Moller K (2002) Culture-independent analysis of gut bacteria: the pig gastrointestinal tract microbiota revisited. Appl Environ Microbiol 68:673–690
- Mallin MA (2000) Impacts of industrial animal production on rivers and estuaries. Am Sci 88:2–13
- Mallin MA, Cahoon LB (2003) Industrialized animal production—a major source of nutrient and microbial pollution to aquatic ecosystems. Popul Environ 24:369–385
- Mallin MA, Burkholder JM, McIver MR, Shank GC, Glasgow HB, Touchette BW, Springer J (1997) Comparative effects of poultry and swine waste lagoon spills on the quality of receiving streamwaters. J Environ Qual 26:1622–1631
- Mallin MA, McIver MR, Merritt JF (2010) Environmental assessment of the Lower Cape Fear River System, 2009. CMS Report No. 10-04. Center for Marine Science, University of North Carolina at Wilmington, Wilmington
- Marti R, Mieszkin S, Solecki O, Pourcher A-M, Hervio-Heath D, Gourmelon M (2011) Effect of oxygen and temperature on the dynamic of the dominant bacterial populations of pig manure and on the persistence of pig-associated genetic markers, assessed in river water microcosms. Appl Microbiol 111:1159–1175
- Meays CL, Broersma K, Nordin R, Mazumder A (2004) Source tracking fecal bacteria in water: a critical review of current methods. J Environ Manag 73:71–79
- 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
- Mieszkin S, Yala JF, Joubrel R, Gourmelon M (2010) Phylogenetic analysis of *Bacteroidales* 16S rRNA gene sequences from human and animal effluents and assessment of ruminant faecal pollution by real-time PCR. J Appl Microbiol 108:974–984
- 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 73:890–901
- Reischer G, Kasper D, Steinborn R, Farnleitner A, Mach R (2007) A quantitative real-time PCR assay for the highly sensitive and specific detection of human faecal influence in spring water from a large alpine catchment area. Lett Appl Microbiol 72:351–356
- Roslev P, Bukh AS (2011) State of the art molecular markers for fecal pollution source tracking in water. Appl Microbiol Biotechnol 89:1341–1355
- Sanapareddy N, Hamp TJ, Gonzalez LC, Hilger HA, Fodor AA, Clinton SM (2009) Molecular diversity of a North Carolina wastewater treatment plant as revealed by pyrosequencing. Appl Environ Microbiol 75:1688–1696
- Savichtcheva O, Okabe S (2009) Qualitative and quantitative estimation of host-specific fecal pollution using *Bacteroides-Prevotella* 16S rRNA genetic markers by T-RFLP and real-time PCR analyses. Water Sci Technol 59:1831–1840
- 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
- Scott TM, Rose JB, Jenkins TM, Farrah SR, Lukasik J (2002) Microbial source tracking: current methodology and future directions. Appl Environ Microbiol 68:5796–5803
- Simpson JM, Santo Domingo JW, Reasoner DJ (2002) Microbial source tracking: state of the science. Environ Sci Technol 45:5279–5288
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739
- Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73:5261–5267