

Identification of Host-Specific Genetic Markers within 16S rDNA Intervening Sequences of 73 Genera of Fecal Bacteria

Zhenyu Shen^{1,2}, Ning Zhang³, Azlin Mustapha², Mengshi Lin², Dong Xu³, Daiyong Deng¹, Mary Reed¹ and Guolu Zheng^{1*}

¹Department of Agriculture and Environmental Sciences, Lincoln University in Missouri, 904 Chestnut Street, Jefferson City, MO 65101, USA

²Food Science Program, University of Missouri, Columbia, MO 65211, USA

³Department of Computer Science and C.S. Bond Life Sciences Center, University of Missouri, Columbia, MO 65211, USA

*Corresponding author: Guolu Zheng, Department of Agriculture and Environmental Sciences, Lincoln University in Missouri, 904 Chestnut Street, Jefferson City, MO 65101, USA, Tel: 573-681-5964; Fax: 573-681-5955; E-mail: ZhengG@LincolnU.edu

Received date: January 5, 2016; Accepted date: January 27, 2016; Published date: January 31, 2016

Copyright: © 2016 Zhenyu S, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Ribosomal intervening sequences (IVSs) were recently proposed as genetic markers for microbial source tracking (MST). This study comprehensively investigated host specificities of IVSs within the 16S rDNA of 73 genera of dominant fecal bacteria using the approaches of bioinformatics and next generation sequencing (NGS). Thirteen types of IVSs were identified *in silico* to be associated with particular host species; they were found within bacteria of the genera *Anaerovibrio*, *Bacteroides*, *Faecalibacterium*, *Mitsuokella*, *Peptostreptococcus*, *Phascolarctobacterium*, and *Subdoligranulum*. Based on the DNA sequences of the thirteen types of IVSs, polymerase chain reaction (PCR) assays were developed. PCR amplifications using fecal DNA samples of target and non-target host species demonstrated that eight out of the 13 IVSs were highly associated with human, chicken/turkey, beef cattle/pig, or horse/pig/human feces. Based on the IVS polymorphisms, NGS was applied to search for single-host-associated IVSs from those linked to multiple host species. Consequently, a new type of IVS specific to beef cattle was found and confirmed by PCR amplification using cattle and non-cattle fecal samples. The results suggest that some IVSs may be used as the genetic markers for MST and that NGS may be useful in identifying novel host-specific genetic markers.

Keywords: Microbial source tracking; Human feces; Beef cattle feces; Host-specific PCR

Introduction

Identification of the source (human vs. animal) of fecal pollution is critical in the assessment and mitigation of fecal pollution [1]. In the past two decades, many microbial source tracking (MST) methods have been developed to determine the sources of fecal pollution in water, and these procedures have been comprehensively reviewed [2-5]. Recently, host-specific polymerase chain reaction (PCR) methods are the most popular MST approaches for determining the sources of fecal pollution by detecting the genetic markers associated with the fecal bacteria unique to a host animal species [2].

Among the host-specific genetic markers, 16S rDNA sequences of fecal bacteria are the most common. As an outcome of studies on the physiology, ecology, and biodiversity of intestinal flora, vast amounts of 16S rDNA have been sequenced for fecal microorganisms associated with human and animal intestinal contents or feces. These sequences are available in several public databases, such as the Ribosomal Database Project (RDP) [6] and Genbank [7]. Although the 16S rDNA sequences are highly conserved, their variable regions can provide discrimination capability to the subspecies level [8] at which host-specific microbes (subsequently genetic markers) can be found. In fact, many host-specific genetic markers have been identified in the 16S rDNA of fecal bacteria, such as *Bacteroides* [9-11], *Faecalibacterium* [12], *Bifidobacterium* [13], *Brevibacterium* [14], and *Catellibacterium* [15]. However, cross-reaction is not unusual for 16S rDNA genetic

markers; this is because of the high degree of sequence similarity among 16S rDNA molecules [16-20].

In contrast, the genes involved in fecal microorganism-host interactions, such as the genes required for bacterial colonization in host intestinal tracts or microbe-host symbiosis, are believed to be more host-specific and are ideal candidates as MST genetic markers [21]. Unfortunately, this type of gene is rarely identified in most fecal microorganisms [22]. Initially, we used the ribosomal intervening sequences (IVSs) of fecal bacteria as genetic markers for MST, taking advantage of the vast 16S rDNA data and the host specificity of IVSs [23]. Ribosomal IVSs are believed to be a result of the coevolution of the host and the bacteria, and IVSs appear to cause a high level of variation among both 16S rDNA and 23S rDNA in adapting to different environments [24,25]. An IVS within the 16S rDNA sequence of the genus *Faecalibacterium* were identified to be specific to poultry (chicken and turkey) feces and found to be distributed widely across a huge geographic area, indicating that IVSs might be alternative host-specific genetic markers, if not superior to the conventional 16S rDNA sequences for MST [23].

The study herein seeks to comprehensively examine, *in silico*, by PCR, and with NGS approaches, the host specificity of IVSs within the 16S rDNA sequences of 73 genera of fecal bacteria dominant in the intestinal tracts of humans and important agricultural animals. The study also aimed to provide information essential to better understand the potential of IVSs as host-specific genetic markers. In this study, term "composite sample" referred to a mixture of an equal amount of DNA extracted from the feces of at least 20 individual animals (or

persons), “target host” an animal species (including human) whose feces was the target of a genetic marker.

Materials and Methods

Fecal sample collection and total DNA extraction

All animal fecal samples were collected as certainly as possible from separate animal individuals in animal farms in Missouri, USA, including samples from chickens, turkeys, beef cattle, dairy cattle, goats, sheep, horses, and pigs. Human fecal samples were also collected in Missouri from the sewage at waste treatment plant inflow. All fecal and sewage samples were kept on ice during transportation and stored at -70°C before DNA extraction. The total fecal DNA was extracted from the samples using the PowerSoil® DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA) and stored at -20°C before use. The DNA was adjusted to a concentration of 5 ng/ μl for PCR amplification.

In silico analyses of host-specific IVSs within the 16S rDNA of the fecal bacteria

Seventy-three genera of bacteria (Table 1) were selected for this study as they are generally considered to be dominant in human and

animal intestinal tracts [26-28]. The aligned 16S rDNA sequences of each genus were downloaded from the RDP database and viewed by the visualization software, BioEdit [29]. IVSs of each genus were manually identified and compiled in an Excel spreadsheet. The IVSs were then sorted by length with the Excel add-in tool, DigDB (<http://www.digdb.com/>). All IVSs shorter than 70 bp were filtered off, as they might be too short for Tagman qPCR, which would be developed in future study. The IVSs thus obtained were used to do a BLAST search against the National Center for Biotechnology Information’s (NCBI) GenBank database [30] to retrieve any DNA sequences containing these IVSs. The information about the host and the host location of the IVS-containing DNA sequences was also subsequently retrieved from GenBank. Any IVS that appeared to be associated with a particular bacterial genus and with no more than three host species was verified by PCR assay for its host specificity (section Verification of the host specificity of the IVSs by PCR assays). In total, 13 IVSs (110-140 bp) were thus identified.

Genus	No. of 16S rDNA	Genus	No. of 16S rDNA
<i>Firmicutes phylum</i>			
<i>Acetanaerobacterium</i>	22	<i>Lachnobacterium</i>	29
<i>Acetivibrio</i>	1606	<i>Lactobacillus</i>	38569
<i>Acidaminococcus</i>	247	<i>Megasphaera</i>	1378
<i>Allisonella</i>	42	<i>Mitsuokella</i>	313
<i>Allobaculum</i>	2717	<i>Mogibacterium</i>	788
<i>Anaerobacter</i>	53	<i>Oribacterium</i>	479
<i>Anaerofilum</i>	122	<i>Oscillibacter</i>	5016
<i>Anaerosporobacter</i>	332	<i>Papillibacter</i>	24
<i>Anaerovibrio</i>	307	<i>Parasporobacterium</i>	1
<i>Blautia</i>	10823	<i>Peptococcus</i>	246
<i>Butyrivibrio</i>	1208	<i>Phascolarctobacterium</i>	980
<i>Catenibacterium</i>	580	<i>Pseudobutyrvibrio</i>	394
<i>Catonella</i>	227	<i>Roseburia</i>	3657
<i>Clostridium</i>	65450	<i>Ruminococcus</i>	5541
<i>Coproccoccus</i>	1687	<i>Sarcina</i>	684
<i>Dialister</i>	3218	<i>Selenomonas</i>	2292
<i>Dorea</i>	2050	<i>Sharpea</i>	49
<i>Erysipelothrix</i>	94	<i>Sporacetigenium</i>	36
<i>Ethanoligenens</i>	101	<i>Sporobacterium</i>	1
<i>Eubacterium</i>	4425	<i>Streptococcus</i>	89257

<i>Faecalibacterium</i>	15823	<i>Subdoligranulum</i>	147
<i>Fastidiosipila</i>	51	<i>Turicibacter</i>	1287
<i>Hespellia</i>	17	<i>Weissella</i>	2680
<i>Howardella</i>	35		
<i>Bacteroidetes phylum</i>			
<i>Bacteroides</i>	41413	<i>Paraprevotella</i>	1205
<i>Barnesiella</i>	7933	<i>Prevotella</i>	33515
<i>Butyricococcus</i>	870	<i>Tannerella</i>	355
<i>Hallella</i>	257	<i>Xylanibacter</i>	13
<i>Parabacteroides</i>	10044		
<i>Actinobacteria phylum</i>			
<i>Adlercreutzia</i>	52	<i>Eggerthella</i>	335
<i>Atopobium</i>	981	<i>Enterorhabdus</i>	123
<i>Bifidobacterium</i>	4842	<i>Olsenella</i>	493
<i>Butyricimonas</i>	262	<i>Slackia</i>	121
<i>Collinsella</i>	1974		
<i>Proteobacteria phylum</i>			
<i>Campylobacter</i>	2107	<i>Lebetimonas</i>	19
<i>Desulfovibrio</i>	4343	<i>Succinivibrio</i>	784
<i>Escherichia</i>	13457		
<i>Spirochaetes phylum</i>			
<i>Spirochaeta</i>	6683	<i>Treponema</i>	6055
<i>TM7 phylum</i>			
<i>TM7 genera incertae sedis</i>	3145		

Table 1: Bacterial phyla and genera along with the number of 16S rDNA sequences used in this study.

Verification of the host specificity of the IVSs by PCR assays

For each of the 13 IVSs selected through the *in silico* analysis described in section *In silico* analyses of host-specific IVSs within the 16S rDNA of the fecal bacteria, a PCR primer set was designed using the NCBI Primer-BLAST program [31]. The specificity of each primer pair was examined with the ProbeMatch program [31] against the Ribosomal Database Project (RDP) database, allowing a maximum of three nucleotide mismatches in each primer. The optimal annealing temperature (TA) of each primer pair was determined through gradient PCR reactions using a gradient TA from 45 to 65°C with the composite fecal DNA samples of the target host species (human or animal). If an expected-size PCR amplicon was generated, the PCR primer set was further tested by PCR amplification against all composite fecal DNA samples from both target and non-target host sources. Each composite DNA sample was a mixture of an equal

amount of DNA extracted from the feces of at least 20 individual animals. The PCR reactions were conducted with 40 cycles of the following thermocycle after initial denaturation at 95°C for 2 min: denaturation at 95°C for 30 sec, annealing at the optimal TA (Table 2) for 30 sec, and elongation at 72°C for 20 sec. The final elongation was at 72°C for 6 min. The 25 µl of PCR reaction cocktail consisted of 10 ng of composite fecal DNA, 25 pmol of each PCR primer (Integrated DNA Technologies, Coralville, IA), and 10 µl of 2.5 × Taq 5 Prime MasterMix (5 Prime Inc., Gaithersburg, MD). PCR products were separated by electrophoresis in 2.0% agarose gels. The PCR amplicons with expected molecular sizes were assumed to be IVS amplicons and were then gel-purified using the GelElute Extraction Kit (5 Prime Inc., Gaithersburg, MD). The purified IVS amplicons were used as the DNA templates in the Illumina® PCR amplifications as described in section Search of novel host-specific IVSs by NGS.

Target IVS	Primer pair (5'→ 3')	Amplicon size (bp)	AT1 (°C)	Host specificity
Anaerovibrio_1	Ana_F1: gtgaggccccgaaaacag Ana_R1: tccatttctaacacccagctct	76	45 to 65 gradient	NSA2
Anaerovibrio_2	Ana_F2: gcttgatagacagcgtcgtg Ana_R2: atccattcctaacacccagctcac	91	45 to 65 gradient	NSA
Anaerovibrio_3	Ana_F3: tgcaaagcaaaaccggctac Ana_R3: agcattccacgctaactca	82	59	Beef cattle and pig
Bacteroides_1	Bac_F1: tcattcgagagcagcggtaaa Bac_R1: ttctcgtcgtcatgtgtgt	83	59	Beef cattle and pig
Faecalibacterium_1	Fae_F1: tccgagccttttcagcgtca Fae_R1: gccactaagcgatttttactctcc	104	55	Chicken and turkey
Faecalibacterium_2	Fae_F2: actccgagttcttcgctgggt Fae_R2: tcccagctctctgctccacct	120	55	Chicken and turkey
Faecalibacterium_3	Fae_F3: actccgggttcttctgctggct Fae_R3: actcccagctctccgctcca	107	45 to 65 gradient	NSA
Faecalibacterium_4	Fae_F4: ttgcgagtcagccggaagcg Fae_R4: tgcccttctgctcagcctactgg	75	60	Beef cattle and pig
Mitsuokella_1	Mit_F1: gagcggactgatgagagggc Mit_R1: catcactcagcctaaagttaagcact	81	62	Horse, human, and pig
Peptostreptococcus_1	Pep_F1: agggcttactcggaattgag Pep_R1: cggcactaatagtgctgtg	95	55	Human
Phascolarctobacterium_1	Pha_F1: cggagcaaatatttactgctctt Pha_R1: cgccactaagagatccaacac	93	45 to 65 gradient	NSA
Subdoligranulum_1	Sub_F1: gctgaaaggtgagcgtcaatc Sub_R1: attttcttactcccagccc	91	55	Chicken and turkey
Subdoligranulum_2	Sub_F2: tgacgcttccggcaagc Sub_R2: actcccaatcctctgccc	84	55	Pig

1Annealing temperature
2No specific amplification (NSA) was observed.

Table 2: Host specificity of the 13 types of IVSs, determined by PCR assays.

PCR without DNA template served as a negative control, and PCR with the universal bacterial 16S rDNA primers, Bac1070F and Bac1392R [32], was used to determine the presence of possible PCR inhibitors. All PCR reactions were repeated at least in duplicate.

Search of novel host-specific IVSs by NGS

To search for single-host-associated IVSs based on IVS polymorphisms from those linked to multiple host species, the following approach of NGS was applied: The Illumina® MiSeq platform using the 2 × 250 nt run was performed at the DNA Core Facility of the University of Missouri in Columbia, Missouri. Briefly, the DNA containing possible IVSs were generated through the PCR amplifications (designated as Illumina® PCRs) using the Illumina® primers with the purified IVS amplicons as the DNA templates. The Illumina® primers consisted of a pair of Illumina® universal forward/reverse adapters, a pair of the universal forward/reverse binding

primers, and a pair of the host-specific IVS forward/reverse primers. A unique barcode sequence was incorporated into each Illumina® reverse primer so that all different Illumina® PCR amplifications were coded. The Illumina® PCR was performed using the conditions detailed in section Verification of the host specificity of the IVSs by PCR assays, except that the TA was 55°C and the elongation time was 60 sec. The resulting Illumina® PCR amplicons were purified with the GelElute Extraction Kit and then with Agencourt AMPure XP beads (Beckman Coulter, Inc., Pasadena, CA), following the manufacturers' protocols. Purified DNA samples were quantified with a Qubit® Fluorometer (Thermo Fisher Scientific, Waltham, MA). The samples were combined in equal molar amounts to form a final 10 nM multiplex pool and were submitted for sequencing through the Illumina® MiSeq platform. The resulting raw reads were used to search for novel host-specific IVSs, as illustrated in Figure 1.

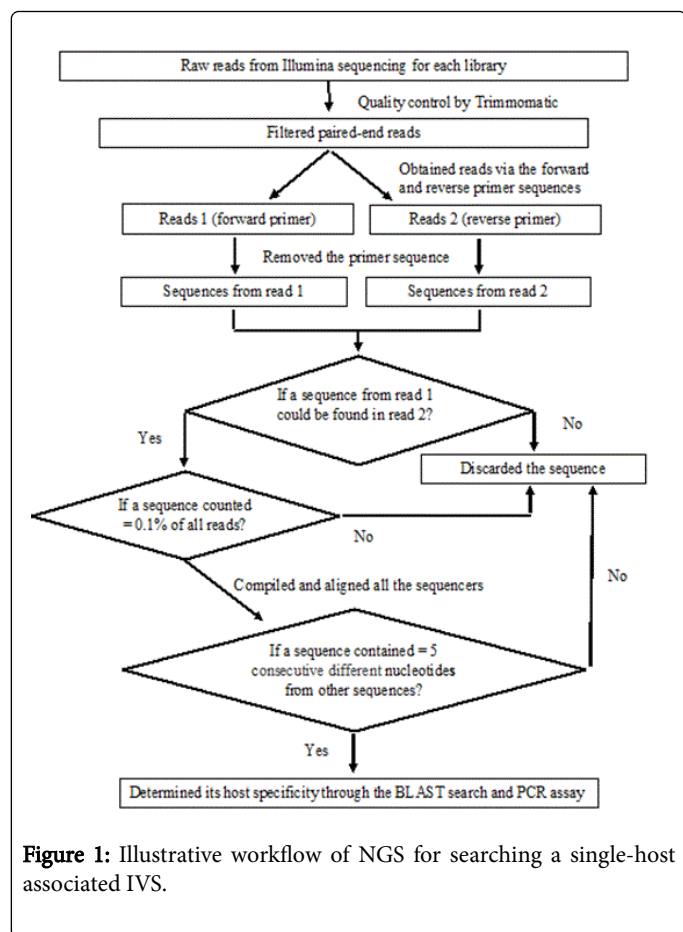


Figure 1: Illustrative workflow of NGS for searching a single-host associated IVS.

Briefly, the raw sequences were pre-processed using Trimmomatic [33] to remove low quality reads and trim any sequencing adaptors. PCR duplicated and sequences containing no Illumina® primers and/or with an ambiguous read were removed. The resulting sequences were aligned, and the identical sequences were clustered and identified as one read type. Only the read types accounting for 0.1% or more of the read population were compared to search for single-host-associated IVSs. Only host-specific IVSs longer than 70 nt were further examined for their host specificities *in silico*, as described in section *In silico* analyses of host-specific IVSs within the 16S rDNA of the fecal bacteria, and by PCR, as explained in section Verification of the host specificity of the IVSs by PCR assays.

The beef cattle-specific PCR assay

Based on the sequence of IVS_BacBC (Figure 2) identified by NGS followed by the multiple sequence alignment comparisons, a PCR assay specific to beef cattle feces was developed. The forward primer is Bac_BCF: 5'-GTAAAGCGTGCCGAAGACTG-3'; the reverse primer, Bac_BCR: 5'-TATCGGGGACTTGTAAGCCG-3'. The thermocycle conditions were those used in section Verification of the host specificity of the IVSs by PCR assays, except that the TA was 55°C. The composite fecal DNA samples of beef cattle and non-cattle host species were used to verify this PCR assay.

```

1                                     50
IVS_BacBC: TCATTTCGAGA GCAGCGGTAA AGCGTGCCGA AGACTGTAAG CAACGTCTCG
Bacteroides_1: TCATTTCGAGA GCAGCGGTAA AAATTTTCG. .... CTTG
Consensus   TCATTTCGAGA GCAGCGGTAA AaagTgcCG. .... CTcG

51                                     100
IVS_BacBC: CGACTTTCTC AAACGCAAGG CGGCTTACAA GTCCCCGATA CAACACATGC
Bacteroides_1: CGAAAATTTT TA...CTACG CG..TTCCAA AGC...ATA CAACACATGC
Consensus   CGAaaaTcTc aA...CaAcG CG..TTaCAA agC...ATA CAACACATGC

101    110
IVS_BacBC: AGCGAGGAA
Bacteroides_1: AGCGAGGAA
Consensus   AGCGAGGAA
    
```

Figure 2: Sequence alignment of IVSs Bacteroides_1 and IVS-BacBC.

- Dots indicate the missing nucleotides in IVS Bacteroides_1, which could be found in both beef cattle and pig feces.
- The capital letters in consensus sequences indicate the identical nucleotides, and the lowercase letters, the different ones.
- The underlined letters indicate the targeting sequence of Bacteroides_1 PCR primers.
- The letters in bold indicate the targeting sequences of IVS_BacBC PCR primers (IVS_Bac BC could only be found in beef cattle feces).

Results

In silico analyses of ribosomal IVSs in fecal bacteria

A total of 406,466 aligned 16S rDNA sequences from 73 bacterial genera were compiled from the RDP database [31]. In order from the highest to the lowest frequency (Table 1), the 16S rDNA sequences were gathered from bacteria of the phyla Firmicutes, Bacteroidetes,

Actinobacteria, Proteobacteria, Spirochaetes, and TM7. Visualized by the BioEdit program [29], IVSs were sporadically and manually found within the variable region 1 (V1) of the 16S rDNA sequences from the genera *Anaerovibrio*, *Bacteroidis*, *Faecalibacterium*, *Mitsuokella*, *Peptostreptococcus*, *Phascolarctobacterium*, and *Subdoligranulum*. Among them, 13 types of IVSs were identified to be associated with fecal or intestinal samples from not more than three host sources, as detailed in Table 3.

Name of IVS ¹ : sequence (length in bp)	Animal host	Sample location
Anaerovibrio-1: ggaagacatttctaaccaccgagtagctgagtagatgactgtcactgagacagcgtcgtagggccccgaaa acagagcctcaccacactatatttagttagtaaacatttagagactgggttagaaatggattct (136)	Dairy cattle	Japan and USA
Anaerovibrio-2: ggaagacatttctaaccaccgagctactaagttggtgtgcttagatagacagcgtcgtagagctcgaagagcct caccacactatataaaagttagtaaacatttagttagctgggttaggaatggattct (132)	Dairy cattle	USA
Anaerovibrio-3: ggctgacagcctaaccactgaactttcagcgttagacgacagcgtgcccgaagtttcaaagcaaacggct acgctttcaaaaagtctaacacaactagaatgtgaaaggttagctgtagctggatgct (138)	Beef cattle and pig	Canada and USA
Bacteroides-1: gagaaagtcaacataaaagcgttcaatcattcgagagcagcggtaaaaatttcgctgcgaaaaattttactacgc gttccaaagcacaacacatgcagcggaggaatcttgagtgctttatgtgaactttct (137)	Beef cattle and pig	China, Korea, and USA
Faecalibacterium-1 (or IVS-p)²: gaaagattttctactccgagttctcgggtcttaaggagagcgtgatcaatcgaagcatcgaagatgcgag cattgatccagcgtttatagaagactaacacaaggtggagcagagactgggtagtggaaaatctttt (148)	Chicken and turkey	USA
Faecalibacterium-2: gagcgtattttctactccgagcctttgcagcgtcaatcaatcggagcattgattagccttataagtaagctgaca catgcagatggtgggtagtaaaaaatcgctt (111)	Chicken and turkey	Australia, China, and USA
Faecalibacterium-3: ggaagattttctactccggttcttctgctgcttaaaagagcgtcaatcaatcggagcattgattcaggcttttaa agaagactaacacagagatggagcggagagctgggtagtggaaaatctttt (132)	Chicken and turkey	China and USA
Faecalibacterium-4: gggtgaaatttctactgagttttgagcagcgggaagcgcattgatgcgaagcatcaataggcttattccagt aaggctgacacaaaggcgaagactgagtgaggaaatcattc (124)	Beef cattle and pig	Korea and USA
Mitsuokella-1: tttcatcactgaacttaaaagagcggactgatgagagcgtcgtaagcgaagcattcaccacactttataaagtgc ttaacttaaggctgagtgatgaaagatgc (110)	Horse, human, and pig	Australia, China, India, Spain, and USA
Peptostreptococcus-1: agggcttactcggaaatgagtagtattcactcagcgtttttctgacccctagtagcagattgggaaacgaaacac gcactataagtgcaagtacgcactattgagtcagtaaaagtagagattgaattctgagtaagacc (149)	Human	USA
Phascolarctobacterium-1: agagcttcggtggcggtagcgggataaaaatttcggagcaaatttattacgctcttaaaagtatgcaacacatgca taaggaaccaacacctgaggtttcagtggtgatactcttagtgccg (127)	Human	China
Subdoligranulum-1: gagtaatttttctactccgaaccttagctgaagctgaaaggtgagcgtcaatcaacgcgaagcgttgattaggctt atttagtaagactaacaccaagcggaggctgggagtgaaagaaatctt (131)	Chicken and turkey	USA

Subdoligranulum-2: gagttattttcttactccgagttctttgcacgctccggcaagcggcaattcatttgcgaagcaaaaggaatttaggctt atttagtaagacgaacacggaagcgagagattggagtaagaaaataact (135)	Pig	Korea and USA
¹ Name of each IVS begins with the name of its associated bacterial genus. ² Faecalibacterium-1 was previously named and reported as IVS-p [23].		

Table 3: Characteristics of the host-specific IVSs identified *in silico*.

Using the aforementioned 13 types of IVSs to do a BLAST search against the GenBank database [30], IVS-containing DNA sequences were retrieved. All the retrieved IVS-containing sequences were 16S rDNA molecules, and each IVS was associated with only one genus of bacteria (Table 3). Single nucleotide polymorphism (SNP) was common among the retrieved IVSs, indicated by “nucleotide mismatches” between the retrieved IVSs and the inquiry IVSs (data not shown). In addition, IVSs found in the genera *Anaerovibrio*, *Bacteroides*, *Faecalibacterium*, and *Mitsuokella* appeared to have wide geographic distribution as the IVSs had been identified in more than one country, while the other IVSs were reported only in one country by the time this study was conducted (Table 3).

Verification of host specificity of the IVSs

Based on the sequences of the 13 types of host-specific IVSs, 13 pairs of PCR primer sets were designed, and the corresponding PCR assays were performed using composite fecal samples from target host species. Among the 13 pairs of primers, nine produced the expected size of amplicons, while the remaining four failed to generate the expected amplicons at all annealing temperatures (45 to 65°C) tested (Table 2). The corresponding IVSs of the nine primer sets were then examined for their host specificities by PCR assays using composite fecal DNA samples from both target and non-target host sources, including beef cattle, chickens, dairy cattle, goats, sheep, horses, humans, pigs, sheep, and turkeys. The results demonstrated that the nine IVSs were only present in the fecal DNA of the target host source(s), which was in agreement with the results of the *in silico* analyses (Table 3).

Target IVS	Source of fecal DNA	Number of valid reads	Number of IVS-BacBC reads	Number of <i>Bacteroides_1</i> reads
<i>Bacteroides_1</i>	Beef cattle	620,942	2, 608	618, 334
	Pig	663,883	0	663, 883

Table 4: Distributions of IVS *Bacteroides_1* and IVS-BacBC determined by the NGS.

Discussion

This study investigated the potential of 16S rDNA IVSs from fecal bacteria as MST genetic markers. The 73 genera represented most, if not all, abundant bacteria in human and agriculturally important animal feces [27,34-37]. Thirteen types of IVSs were identified to be specific to one, two, or three host sources (Tables 2 and 3). PCR assays were developed, which were able to detect nine out of the 13 IVSs (Table 2). Derived from the nine IVSs and with the approach of NGS, a novel IVS, IVS_BacBC, was found to be unique to the feces of beef cattle. Most of the host-specific IVSs found in this study could be

Search for novel IVSs by NGS

Most NGS analyses generated from 400,000 to 700,000 quality sequenced reads, which could be clustered into from 1,000 to 10,000 read types, but no quality reads were obtained for IVS *Mitsuokella_1* against all the associated host sources (horse, human, and pig) and for either *Faecalibacterium_1* or *Subdoligranulum_1* against the turkey source. The length and nucleotide distribution of reads derived from each type of IVS exhibited a high degree of nonuniformity, which was observed among each associated host source. Single nucleotide polymorphism (SNP) was ubiquitous throughout the sequences of the nine types of IVSs, but the mosaic nucleotide variations provided no differential power to discriminate among the same types of IVSs between or among different host sources. One exception was a unique read type of IVS *Bacteroides_1*, accounting for 0.42% of the read population (Table 4). The sequence was found in the feces of beef cattle but not in pig feces, which is another source associated with *Bacteroides_1* (Table 2). This unique IVS, significantly different from the IVS *Bacteroides_1* (Figure 2), was designated IVS_BacBC (stands for an IVS from genus *Bacteroides* and specific to beef cattle). Using the sequence IVS_BacBC to do a BLAST search against the Genbank, not one similar sequence was found, suggesting that the IVS is a novel one. A pair of PCR primers (IVS_BCF and IVS_BCR) were designed for IVS_BacBC, and the corresponding PCR assay was developed. The result of the PCR assay, using target and nontarget fecal samples, confirmed IVS_BacBC to be specific to beef cattle feces.

detected in fecal samples from at least two countries, demonstrating their wide geographic distribution (Table 3). This suggests that these potential MST markers might be useful for a huge geographic range.

IVSs of 16S- and 23S-rDNA are collectively called ribosomal IVSs. They were first found in the 23S rDNA of *Salmonella enterica* Typhimurium [38] and then in the 16S and 23S rDNA of other bacteria [39-42]. Ribosomal IVSs are excised from 16S- or 23S-rRNA by ribonucleases after transcription, resulting in rRNA fragmentation [43]. The fragmentation, which is a bacterial response to living environments by adjusting rRNA levels, is believed to increase the rRNA degrading rate by creating more targets for ribonucleases [43].

The presence of ribosomal IVSs might be an adaptation of microbes to their living environments, e.g., the intestinal tracts of their hosts, in the case of fecal microorganisms. In other words, IVSs might be host-adapted. In addition, IVSs contribute to the diversity of bacterial rDNA [24,44], thus providing a genetic basis for differentiating among the fecal microbes of different host sources. Therefore, ribosomal IVSs can be used as host-specific genetic markers for MST.

We previously identified an IVS specific to the feces of poultry (chicken and turkey) within the 16S rDNA of the genus *Faecalibacterium* [23]. This study identified 13 types of host-specific IVSs in seven genera. Surprisingly, although the bacteria of the genus *Bacteroides* are among the most abundant fecal microbes in many host species [25,37,45,46], only one host-specific IVS suitable for PCR detection was identified from this genus (Tables 2 and 3). IVSs of *Faecalibacterium* appeared to be more abundant and diverse. Three types of *Faecalibacterium* IVSs were identified to be associated with poultry (chicken and turkey) feces, including the previously reported IVS-p [23], while one type of *Faecalibacterium* IVS was linked to the feces of beef cattle and pigs. The genus *Faecalibacterium* and the closely related genus *Subdoligranulum* account for 49% of the bacterial population in the cecum of chickens [45]. For detecting poultry feces, IVSs of these two genera might be superior, in terms of abundance, to IVSs of fecal bacteria from other genera.

16S rDNA IVSs were not common but were widely distributed in many bacterial genera [43]. Our result supports this conclusion. At least seven out of 73 genera of bacteria were found to contain the 16S rDNA IVSs. However, frequency of occurrence of IVSs in 16S rDNA might in fact be higher than what we observed because smaller IVSs (< 70 bp) were not accounted for in this study, and some large IVS-containing 16S rDNA sequences might have been excluded from the GenBank and RDP databases. This is due to the techniques used to obtain the 16S rDNA sequence data. Before the era of NGS, the data were commonly obtained through 16S rDNA cloning followed by DNA sequencing, where the larger-than-usual 16S rDNA would have had less chance to be cloned or even was intentionally excluded as non-16S rDNA. Although an IVS can be longer than 350 bp [47], our analysis did not find any IVSs larger than 200 bp, indicating that large-IVS-containing 16S rDNA sequences are missing in the databases.

To better understand the diversity and occurrence of ribosomal IVSs in fecal bacteria, research beyond the publically available data is needed.

Fecal pollution from humans, compared with other sources, poses the highest risk to human health because it can spread human diseases. In our study, three IVSs (*Mitsuokella_1*, *Peptostreptococcus_1*, and *Peptostreptococcus_1*) were identified *in silico* to be associated with human feces. We were only able to develop a PCR assay to detect the IVSs *Mitsuokella_1* and *Peptostreptococcus_1*. However, IVS *Mitsuokella_1* was not only found in human feces but also in horse and pig feces (Tables 2 and 3). Bacteria of *Peptostreptococcus* might be a good fecal indicator of a human source. It is a genus of anaerobic, Gram-positive, non-spore-forming bacteria [48] that is unable to survive in an environment where oxygen is present and therefore can be used as a fecal indicator for fresh fecal pollution. Furthermore, *Peptostreptococcus* is among the most abundant genera of human enteric microbiota [49]. However, further experiments are needed to understand the complete value of the IVS *Peptostreptococcus_1* as a genetic marker for tracking human fecal pollution in the environment, although the current study presents an initial investigation.

IVS_BacBC appeared to be a novel genetic marker for the detection of beef cattle feces. Further tests with fecal samples from wider geographic locations are needed to verify the value of this marker. Surprisingly, beef cattle shared the same types of IVSs more often with pigs than dairy cattle. For example, the IVSs *Anaerovibrio-3*, *Bacteroides-1*, and *Faecalibacterium-4* were found in both beef cattle and pigs but not in other animal species (Table 3). This result was in agreement with some previous research [25,37] reporting that beef cattle and pigs, but not dairy cattle, had similar intestinal microbiota. However, the cause of the observations remains unknown.

Conclusions

Our results demonstrated that some ribosomal IVSs may be used as genetic markers in MST and that the large-size (>200 bp) IVSs were absent from the public databases such as RDP and GenBank. This study suggests that ribosomal IVSs are a group of closely related sequences with various host specificity. Our data suggest that the IVS *Peptostreptococcus_1* and the IVS *IVS_BacBC* may be useful genetic markers for identification of human and beef cattle feces, respectively.

Acknowledgments

This research was financially supported by a USDA-NIFA Evans-Allen Grant (Project NO: MOX-Zheng, grant# 0223248, to Zheng).

References

- USEPA (2005) Microbial source tracking guide document. US Environmental Protection Agency, Office of Research and Development, National Risk Management Research Laboratory, Cincinnati, OH.
- Harwood VJ, Staley C, Badgley BD, Borges K, Korajkic A (2014) Microbial source tracking markers for detection of fecal contamination in environmental waters: Relationships between pathogens and human health outcomes. *FEMS Microbiol Rev* 38: 1-40.
- Roslev P, Bukh AS (2011) State of the art molecular markers for fecal pollution source tracking in water. *Appl Microbiol Biotechnol* 89: 1341-1355.
- 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, Domingo JWS, Reasoner DJ (2002) Microbial source tracking: State of the science. *Environ Sci Technol* 36: 5279-5288.
- Cole JR, Chai B, Farris RJ, Wang Q, Kulam SA, et al. (2005) The ribosomal database project (RDP-II): Sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res* 33: D294-D296.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2008) Genbank. *Nucleic Acids Res* 36: D25-D30.
- Zhang C, Zheng G, Xu SF, Xu D (2012) Computational challenges in characterization of bacteria and bacteria-host interactions based on genomic data. *J Comput Sci Technol* 27: 225-239.
- Bernhard AE, Field KG (2000) Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. *Appl and Environ Microbiol* 66: 1587-1594.
- Shanks OC, Domingo JWS, Lamendella R, Kelty CA, Graham JE (2006) Competitive metagenomic DNA hybridization identifies host-specific microbial genetic markers in cow fecal samples. *Appl Environ Microbiol* 72: 4054-4060.
- Reischer GH, Kasper DC, Steinborn R, Mach RL, Farnleitner AH (2006) Quantitative PCR method for sensitive detection of ruminant fecal pollution in freshwater and evaluation of this method in Alpine Karstic regions. *Appl Environ Microbiol* 72: 5610-5614.

12. Zheng G, Yampara-Iquise H, Jones JE, Andrew Carson C (2009) Development of *Faecalibacterium* 16S rRNA gene marker for identification of human faeces. *J Appl Microbiol* 106: 634-641.
13. Bonjoch X, Lucena F, Blanch AR (2009) The persistence of bifidobacteria populations in a river measured by molecular and culture techniques. *J Appl Microbiol* 107: 1178-1185
14. Weidhaas JL, Macbeth TW, Olsen RL, Sadowsky MJ, Norat D, et al. (2010) Identification of a *Brevibacterium* marker gene specific to poultry litter and development of a quantitative PCR assay. *J Appl Microbiol* 109: 334-347.
15. Lu J, Domingo JWS, Lamendella R, Edge T, Hill S (2008) Phylogenetic diversity and molecular detection of bacteria in gull feces. *Appl Environ Microbiol* 74: 3969-3976.
16. Ahmed W, Goonetilleke A, Powell D, Gardner T (2009) Evaluation of multiple sewage-associated *Bacteroides* PCR markers for sewage pollution tracking. *Water Res* 43: 4872-4877.
17. Haugland RA, Varma M, Sivaganesan M, Kelty C, Peed L, et al. (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.
18. Layton A, McKay L, Williams D, Garrett V, Gentry R, et al. (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.
19. Ryu H, Griffith JF, Khan IU, Hill S, Edge TA, et al. (2012) Comparison of gull feces-specific assays targeting the 16S rRNA genes of *Catelicoccus marimammalium* and *Streptococcus* spp. *Appl Environ Microbiol* 78: 1909-1916.
20. Lamendella RJ, Yannarell AC, Ghosh S, Di G, Mackie RI, et al. (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.
21. Domingo JWS, Sadowsky MJ (2007) *Microbial source tracking*. ASM Press, Washington, D.C.
22. Domingo JWS, 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.
23. Shen Z, Duan C, Zhang C, Carson A, Xu D, et al. (2013) Using an intervening sequence of *Faecalibacterium* 16S rDNA to identify poultry feces. *Water Res* 47: 6415-6422.
24. Pei A, Nossa CW, Chokshi P, Blaser MJ, Yang L, et al. (2009) Diversity of 23S rRNA genes within individual prokaryotic genomes. *PloS one* 4: e5437.
25. Durso LM, Harhay GP, Smith TP, Bono JL, Desantis TZ, et al. (2010) Animal-to-animal variation in fecal microbial diversity among beef cattle. *Appl Environ Microbiol* 76: 4858-4862.
26. Wadud S, Michaelsen A, Gallagher E, Parsci G, Zemb O, et al. (2012) Bacterial and fungal community composition over time in chicken litter with high or low moisture content. *Brit Poultry Sci* 53: 561-569.
27. Dowd SE, Callaway TR, Wolcott RD, Sun Y, Mckeehan T, et al. (2008) Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). *BMC Microbiol* 8: 125-132.
28. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, et al. (2005) Diversity of the human intestinal microbial flora. *Science* 308: 1635-1638.
29. Hall TA (1999) Bioedit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Sym Ser*: 95-98.
30. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410.
31. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, et al. (2009) The ribosomal database project: Improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* 37: D141-D145.
32. Ferris M, Muyzer G, Ward D (1996) Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot spring microbial mat community. *Appl Environ Microbiol* 62: 340-346.
33. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* 170.
34. Ley RE, Peterson DA, Gordon JI (2006) Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* 124: 837-848.
35. Lu J, Santo Domingo J (2008) Turkey fecal microbial community structure and functional gene diversity revealed by 16S rRNA gene and metagenomic sequences. *J Microbiol* 46: 469-477.
36. Steelman SM, Chowdhary BP, Dowd J, Suchodolski J, Janecka JE (2012) Pyrosequencing of 16S rRNA genes in fecal samples reveals high diversity of hindgut microflora in horses and potential links to chronic laminitis. *BMC Vet Res* 8: 231-242.
37. Kim HB, Borewicz K, White BA, Singer RS, Sreevatsan S, et al. (2011) Longitudinal investigation of the age-related bacterial diversity in the feces of commercial pigs. *Vet Microbiol* 153: 124-133.
38. Burgin AB, Parodos K, Lane DJ, Pace NR (1990) The excision of intervening sequences from *Salmonella* 23S ribosomal RNA. *Cell* 60: 405-414.
39. Meinersmann R, Phillips R, Ladely S (2009) Inter- and intra-genomic heterogeneity of the intervening sequence in the 23S ribosomal RNA gene of *Campylobacter jejuni* and *Campylobacter coli*. *Syst Appl Microbiol* 32: 91-100.
40. Tazumi A, Sekizuka T, Moore J, Millar B, Taneike I, et al. (2008) Molecular characterization of intervening sequences in 23S rRNA genes and 23S rRNA fragmentation in *Taylorella equigenitalis*. *Folia Microbiol* 53: 486-492.
41. Rossi M, Zannoni RG, Hanninen ML (2010) Delineation of two *Helicobacter bilis* genomospecies: Implications for systematics and evolution. *Int J Syst Evol Microbiol* 60: 2392-2397.
42. Pronk LM, Sanderson KE (2001) Intervening sequences in rrl genes and fragmentation of 23S rRNA in genera of the family *Enterobacteriaceae*. *J Bacteriol* 183: 5782-5787.
43. Hackenberg E (2005) Bacterial ribosomal RNA in pieces. *Mol Microbiol* 57: 318-325.
44. Pei AY, Oberdorf WE, Nossa CW, Agarwal A, Chokshi P, et al. (2010) Diversity of 16S rRNA genes within individual prokaryotic genomes. *Appl Environ Microbiol* 76: 3886-3897.
45. Bjerrum L, Engberg R, Leser TD, Jensen BB, Finster K, et al. (2006) Microbial community composition of the ileum and cecum of broiler chickens as revealed by molecular and culture-based techniques. *Poult Sci* 85: 1151-1164.
46. Scupham AJ (2007) Succession in the intestinal microbiota of preadolescent turkeys. *FEMS Microbiol Ecol* 60: 136-147.
47. Smuts HE, Lastovica AJ (2010) Molecular characterization of the 16S rRNA gene of *Helicobacter fennelliae* isolated from stools and blood cultures from paediatric patients in South Africa. *J Pathog* 2011: 1-9.
48. Moore H, Lv JJ, Moore W (1986) Genus *Peptococcus*, in *Bergey's manual of systematic bacteriology*. Williams & Wilkins, Baltimore, MD.
49. Guarner F, Malagelada JR (2003) Gut flora in health and disease. *Lancet* 361: 512-519.