Host-specific Genetic Markers of Fecal Bacteria for Fecal Source Tracking in Food and Water

Guolu Zheng1 and Zhenyu Shen2

1Cooperative Research Program, Lincoln University, Jefferson, MO 65101, USA
2Veterinary Diagnostic Laboratory, University of Missouri, Columbia, MO 65211, USA

Abstract

Fecal pollution in food and water is the major cause of disease outbreaks, so accurate identification of this fecal pollution in these locations is critical to prevent such outbreaks. Fecal source tracking (FST) is an effective tool to identify the sources (human vs. animal) of any particular instance of fecal pollution. Many FST technologies were developed over the past two decades and used mainly to monitor and manage water quality, but their applications have recently attracted more attention in food systems. With the great advances of technologies in DNA sequencing and related bioinformatic tools, a significant number of novel genetic markers of fecal indicator bacteria (FIB) have been identified to be associated with particular host species, and thus, suitable for FST. The host-specific genetic markers were originally limited to 16S rDNA sequences of FIB, which are relatively conserved. Novel genetic markers include genes of the host-bacterium interaction and intervening sequences within the 16S- and 23S-rDNA of fecal bacteria. However, most of the genetic markers were only evaluated in laboratory settings, and their competive values in the monitoring and management of food and water safety need further assessments in the field.

Keywords: Fecal Source Tracking (FST); Food safety; Water safety; Fecal contamination; Monitoring; Foodborne illness; Waterborne illness; Fecal Indicator Bacteria (FIB)

Introduction

The feces of human and food animals are the main sources of pathogens associated with foodborne and waterborne illness outbreaks. Not only ill people and animals can shed pathogens in their feces; healthy animals can be reservoirs of pathogens. Cattle have been determined to be the major reservoir for enterohemorrhagic Escherichia coli O157:H7 [1], chicken and turkey for Salmonella spp. and Campylobacter jejuni [2], and pig for Campylobacter coli [3]. Thus, one of the main concerns regarding the microbiological quality of food and water is the presence of pathogens associated with human and animal feces. It is impossible to monitor all pathogens on a routine basis because each of these microorganisms requires a specific test method. Therefore, fecal indicator bacteria (FIB) were introduced to evaluate the food and water systems to assess the microbiological quality and to predict the presence of pathogens therein [4,5].

While the presence of FIB is indicative of fecal pollution, it provides no information about the sources of pollution. That information is needed so that effective steps can be taken for food and water safety. To identify fecal pollution sources, fecal source tracking (FST) or microbial source tracking (MST) methods have been developed. Although current FST methods are primarily used in the field of water monitoring and management to reduce outbreaks of waterborne diseases, it is suggested that MST methods can be applied to track fecal pollution sources in food systems to better prevent the contamination of food by feces-carrying pathogens [6]. The principle of FST is to compare the characteristics of microorganisms isolated from the polluted water with those of fecal microorganisms in host-known feces [6,7]. The assumption of FST is that the distribution of fecal microorganisms in human or animal intestinal tracts is not random, i.e., some fecal microorganisms are host-specific. Obviously, the goal of FST is to determine which fecal microorganism(s) or associated trait(s) (i.e., phenotypic or genotypic) is/are specific to an animal species.

Over the last two decades, FST technologies have expanded greatly, from phenotype-bases to genotype-bases, from detection of standard FIB to alternative FIB, from differentiations of bacterial 16S-RNA genes to other less conserved genes, and from fecal bacteria to viruses. There have been several comprehensive review papers on FST technologies and their application in the monitoring and management of water quality [6-11]. This review evaluates the recent developments in FST technologies, with an emphasis on those using host-specific genetic markers of fecal bacteria and their application in the management of food and water safety.

Fecal Indicator Bacteria

Historically, total and fecal coliforms, E. coli, and enterococci bacteria were used as standard FIB in many countries to monitor the microbiological quality of food and water. Recently, E. coli has been proven to be a better fecal indicator bacterium than total and fecal coliforms for fresh water, while enterococci are optimal when testing marine water [12]. The Beaches Environmental Assessment and Coastal Health (BEACH) Act of 2000 requires states with marine or Great Lakes to adopt enterococci as the fecal indicator to monitor recreational water quality [13].

However, the standard FIB (E. coli and enterococci) have at least two drawbacks. First, they are found in relatively low numbers in human and animal intestinal tracts (feces), resulting in low detection sensitivity. In human and animal intestines, prokaryotes are the dominant microbes, with approximately 10^{11} cells/g feces [14], but the
Genetic Marker Based FST Methods

With advances in science and the advent of new technologies, many FST methods have been developed and used to track fecal pollution in water. Categorically, these methods can be divided into library-dependent and library-independent methods [6]. The library-dependent methods require the construction of a library or database containing a set of either phenotypic or genotypic characteristics of FIB, commonly E. coli or enterococci bacteria, isolated from the feces of known host sources. To determine the fecal sources of a feces-polluted site, strains of E. coli or enterococci are isolated from the site and their phenotypic or genotypic characteristics are then compared with those in the library to find the matches [6]. On the other hand, library-independent FST methods do not require a reference library, relying instead on the detection of host-specific markers associated with FIB or animal feces.

A host-specific marker can be either a genetic or chemical marker that is unique to the feces of an animal species or human being. A host-specific genetic marker refers to a unique nucleic acid (i.e., DNA or RNA) sequence that is exclusively or strongly associated with particular host sources. This sequence can be either a nucleic acid sequence of fecal cells (e.g., bacteria and viruses) or of host cells. A number of chemicals and metabolites have been found to be associated with humans or animals because of different life styles and digestion systems. Chemical marker-based methods have been comprehensively reviewed recently [41], so this review will focus on the host-specific genetic makers of fecal bacteria.

### Host-specific Genetic markers of Fecal Bacteria

Theoretically, genes in microorganisms involved in host-microbe interactions are ideal potential host genetic markers. However, most fecal microorganisms are uncultured, and their genomes remain largely unknown. For many years, bacterial 16S rDNA were the only genetic markers used in FST.

#### 16S rDNA

The pioneer work of developing host-specific markers derived from 16S rDNA sequences of fecal bacteria is credited to research by Field and others [42]. Since 2000, a considerable number of host-specific genetic markers for the detection of major sources of fecal pollution have been identified in 16S rDNA. Among them, host-specific genetic markers of Bacteroidales 16S rDNA have been the most intensively studied, including through field studies, as these markers are the most promising for FST [43]. Therefore, 16S rDNA has become the most widely used genetic marker in FST.

#### Table 1: 16S rDNA Host-specific Marker

<table>
<thead>
<tr>
<th>Target Host</th>
<th>Target bacteria</th>
<th>Marker</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>Faecalibacterium</td>
<td>Bi-AOD, Bi-DEN</td>
<td>Bonjoch et al. [30]</td>
</tr>
<tr>
<td></td>
<td>Bacteroidales</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bifidobacterium</td>
<td>HM</td>
<td>Gómez-Dofar et al. [89]</td>
</tr>
<tr>
<td></td>
<td>Bifidobacterium</td>
<td>HFB-F3/R5</td>
<td>Zheng et al. [28]</td>
</tr>
<tr>
<td></td>
<td>Faecalibacterium</td>
<td>CF128F, CF193F</td>
<td>Bernhard and Field [21]</td>
</tr>
<tr>
<td></td>
<td>Bacteroidales</td>
<td>BacCow</td>
<td>Kildare et al. [86]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cow-Bac</td>
<td>Okabe and Shimazu [90]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CH125F</td>
<td>Stricker et al. [91]</td>
</tr>
<tr>
<td></td>
<td>Bifidobacterium</td>
<td>CW</td>
<td>Gómez-Dofar et al. [89]</td>
</tr>
<tr>
<td>Deer/Eland</td>
<td>Bacteroidales</td>
<td>EF447F/990R</td>
<td>Dick et al. [93]</td>
</tr>
<tr>
<td>Dog</td>
<td>Bacteroidales</td>
<td>DF475F</td>
<td>Dick et al. [93]</td>
</tr>
</tbody>
</table>

Table 1: Host-specific genetic markers derived from bacterial 16S rDNA.

Although 16S rDNA is highly conserved across species of the same genus of bacteria, variable regions of 16S rDNA provide a tool to discriminate between bacteria at the subspecies level, according to their host [44]. In addition, 16S rDNA has multiple copies in most bacterial cells [45]. For example, each E. coli cell has seven copies of almost identical 16S rDNA in its genome [46]. Multiple numbers of markers in cells can increase the detection sensitivity.

Recently, studies of the physiology, ecology, and biodiversity of intestinal flora have resulted in an enormous number of 16S rDNA sequences of microorganisms from human and animal guts or feces. Those sequences are available via public databases, such as the Ribosomal Database Project (RDP) [47] and Genbank [48]. By comparing these sequences using bioinformatics approaches, host-specific markers can be found in 16S rDNA [26].

However, 16S rDNA-based genetic markers are not without drawbacks. Cross-reaction is a common issue [49-52].

Table 2: Host-specific genetic markers derived from bacterial virulence genes
Bacterial virulent genes

Several virulent genes of pathogenic strains of the standard FIB have been reported to be host associated Table 2. The heat labile toxin Ila (LTIIa) gene and the heat stable toxin II (STII) gene in enterotoxigenic E. coli (ETEC) have been found to be responsible for diarrheal diseases of cattle and pig [53]. Sequence analysis of LTIIa and STII in addition to PCR reactions showed the two markers were 100% specific to cattle and pig feces, respectively [54-56]. Other studied but less successful FST markers in the toxin genes of E. coli include a heat-stable enterotoxin gene (STIIb) for humans [57], a major fimbrial subunit gene (raG) for rabbits, a P fimbrial adhesin gene (papG) for dogs, and a temperature-sensitive hemagglutinin gene (tsh) for birds [58]. Seven goose/duck specific genetic markers were identified in E. coli isolates by the suppression subtractive hybridization (SSH) method, and their specificities were tested with the colony hybridization method. Among the markers, sequences of GA9 and GC11 are homologues to the genes encoding type III secretion proteins in E. coli O157:H7, and both GB2 and GE11 have been found to encode adhesion-like proteins in E. coli O157:H7 [59].

In Enterococcus faecium, a putative virulence gene encoding the enterococcal surface protein (esp) was also proposed for identification of human fecal pollution, with controversial results [60-64]. The detection sensitivity of virulent genes has also been an issue because healthy humans and animals usually do not shed such pathogens in their feces, and the pathogens are present in low numbers in ill humans and animals.

Bacterium-host interaction genes as markers

As host-specific genetic markers, genes involved in bacterium-host interactions are superior to other genetic markers. Although knowledge of such genes is very limited for most intestinal microbes, studies have identified some host-specific genetic markers that might be associated with bacterium-host interactions, primarily through gene annotations Table 3.

Bacteroides thetaiotaomicron is a symbiont living in the human gut, and it contributes to degrading indigestible polysaccharides to provide calories and maintain the health of the intestinal ecosystem. The gene encoding α-1,6-mannanase is believed to be involved in this degradation [65], and this gene has been found to be highly associated with human feces [66,67].

There are several genetic markers reportedly associated with domestic animals. By the method of genome fragment enrichment of fecal metagenomes, Shanks et al. [68] identified 26 genetic markers in B. thetaiotaomicron highly associated with humans. Three of them showed high similarities to the genes encoding outer membrane protein (hum39), genes encoding the outer membrane efflux protein precursor (hum336), and genes of a protein associated with remodeling bacterial surface polysaccharides and lipopolysaccharides (hum163) [68]. With the same metagenomic approach, a series of cattle-specific markers were identified, and three Bacteroidales-like markers (i.e., Bac1, Bac2, and Bac3) were randomly selected for PCR assay development. The markers are predicted to be associated with membrane protein and secretion [69].

Other bacterial genes as markers

There are various bacterial genes, whose functions are unknown or not directly involved in bacterium-host interactions, found to be associated with particular host sources Table 3. With DNA microarray hybridization, Soule et al. [70] identified eight genetic markers in the Enterococcus genome specific to cattle (M15 and M19), two to elk/deer (M40 and M48), and four to humans (M67, M68, M77, and M81). M15 is homologous to the helicase gene, M40 belongs to the MutS2 family, M68 is related to carbohydrate kinase PfkB, M77 encodes a transcriptional regulator, and the other markers do not have significant similarity to function-known genes [70].

Using the metagenomic method, Lu et al. [71] identified 21 chicken-specific genetic markers having various predictive functions, including cellular process, metabolism, and information storage. The three markers (i.e., CP2-9, CP3-49, and CB8-24) that performed best were associated with bacterial metabolism.

Recently, research by Zheng et al. [26] used the approach of bioinformatics in the comparative analysis of 7,458 sequences of Faecalibacterium 16S rDNA, reportedly associated with human and animal species. They identified an intervening sequence (IVS), IVS-p, within Faecalibacterium 16S rDNA. IVS-p appeared to be specific to poultry (chicken and turkey) feces. IVS was considered to be a bacterial adaptation to a close working relationship with the host species [72].

Applications of FST in Food Systems

Seafood

Using FST technologies in tracking fecal pollution sources in foods has attracted increasing interest recently [73-75]. Pathogens carried by feces can enter the food production chain through various paths, from farm to table. Food crops can be polluted in the field, by untreated human or animal manure used as fertilizer or by feces-polluted irrigation water. Pathogens can enter a water body through point (i.e. specific) pollution sources, such as effluent from wastewater treatment plants, lagoons, and septic tanks, or through non-point (i.e. non-specific) pollution sources, such as wildlife or storm runoff from urban and agricultural areas [76]. FST methods were first introduced into aquaculture and then into vegetable production for food safety management.

Molluscan shellfish, including oysters, clams, mussels, and scallops, are an important food commodity in the United States. Shellfish can accumulate waterborne pathogens in their bodies, posing health risks to consumers [7]. Mauffret et al. [77] investigated the presence of host-specific genetic markers in oysters, cockles, and clams grown in water artificially or naturally polluted by feces. HF183 (human-specific, Bacteroides), Rum2Bac (cattle-specific, Bacteroides), Swine1Bac (pig-specific, Bacteroides), and Gull2 (gull-specific, C. marimallium) were quantified in 100% of intravelar liquid samples of oysters living in artificially polluted water, while HF183 and Rum2Bac were quantified in 31% and 23% of those shellfish living in naturally polluted water. However, none of the host-specific markers was detected in the cockle or clam. This study suggests that FST methods can be used to manage at least some seafood products.

Fresh produce

Fresh produce has been associated with rising numbers of foodborne disease outbreaks in the United States, as the products are increasingly consumed as a part of healthy diets. Most of the pathogens causing outbreaks are of fecal origin. Feces-polluted irrigation water and/or improperly composted human/animal wastes can be the
sources of fecal pollution in produce [78,79]. The work of Ravaliya et al. [74] in Northern Mexico has provided evidence that Bacteroidales 16S rDNA markers may serve to distinguish both general and human-feces contamination in the production environment. Furthermore, FST can be used for the rapid identification of human or livestock fecal contamination in fresh produce [73]. It is generally believed that FST technologies can be a powerful tool to manage food safety in produce production.

<table>
<thead>
<tr>
<th>Target host</th>
<th>Target bacteria</th>
<th>Target gene</th>
<th>Gene function</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>Bacteroides thetaiotaomicron</td>
<td>α-1,6, mannanase</td>
<td>Degradation of polysaccharides</td>
<td>Yampara-Iquise et al. [87]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hum39</td>
<td>Putative outer membrane protein</td>
<td>Shanks et al. [68]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hum163</td>
<td>Putative remodeling of bacterial surface polysaccharides and lipopolysaccharides</td>
<td>Shanks et al. [68]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hum336</td>
<td>Putative efflux protein precursor</td>
<td>Shanks et al. [68]</td>
</tr>
<tr>
<td>E. coli</td>
<td></td>
<td>ycjM</td>
<td>Putative degradation of polysaccharides</td>
<td>Deng et al. [102]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H8</td>
<td>Sodium/hydrogen exchanger precursor</td>
<td>Gomi et al. [101]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H12</td>
<td>Putative phage protein</td>
<td>Gomi et al. [101]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H14</td>
<td>ATP/GTP-binding protein</td>
<td>Gomi et al. [101]</td>
</tr>
<tr>
<td>Cattle</td>
<td>Bacteroidales</td>
<td>Bac1, Bac2, Bac3</td>
<td>Putative membrane secretion protein</td>
<td>Shanks et al. [69]</td>
</tr>
<tr>
<td></td>
<td>E. coli</td>
<td>Co3</td>
<td>Putative integrase</td>
<td>Gomi et al. [101]</td>
</tr>
<tr>
<td>Pig</td>
<td>E. coli</td>
<td>P1</td>
<td>F1C fimbral usher</td>
<td>Gomi et al. [101]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P3</td>
<td>DNA fragment</td>
<td>Gomi et al. [101]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P4</td>
<td>Hypothetical protein</td>
<td>Gomi et al. [101]</td>
</tr>
<tr>
<td>Poultry</td>
<td>E. coli</td>
<td>Ch9</td>
<td>Hypothetical protein</td>
<td>Gomi et al. [101]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ch12</td>
<td>Type I restriction-modification system</td>
<td>Gomi et al. [101]</td>
</tr>
</tbody>
</table>

Table 3: Host-specific genetic markers derived from bacterium-host interaction or other genes

Conclusion

Although 16S rDNA are currently the dominant FST genetic markers, their high degree of conserved sequences makes the cross-reaction an inherent issue. Genetic markers with higher host-specificity may be found in microbial genes involved with microbiome-host interactions. In fact, the known human-feces marker, the α-1,6-mannanase gene of B. thetaiotaomicron, is highly associated with humans [66,67]. However, the determination of microbiome-host interaction genes relies much on the current knowledge of the microbial genome, which excludes most fecal microbes. It has been proposed to use the ribosomal intervening sequence (IVS) of fecal bacteria as genetic markers for FST, which would combine the advantages of the enormous data about 16S rDNA available in public databases with the desirable host specificity of the genes (DNA fragments) involved in microbiome-host interactions. Ribosomal IVSs are insertion sequences in the 16S or 23S rDNA of prokaryotes. They are post-transcriptionally excised by RNase III without religation, which causes RNA fragmentation [80]. The fragmentation may enhance the rRNA degrading rate by creating more targets for certain ribonucleases [80]. It is known that bacterial cells adjust their RNA levels based on environmental changes. The fact that IVSs are mainly found in symbionts and pathogens of eukaryotic hosts also supports the conclusion that IVSs may contribute to the host-microbe interaction [72]. The presence of ribosomal IVS in prokaryotes is relatively uncommon but does occur in many bacterial species [81,82]. IVS is more commonly found in 23S rRNA than in 16S rRNA [80,81,83]. With the development of next-generation sequencing (NGS) methods, a vast amount of 23S rDNA data has been increasingly accumulated, which will facilitate the identification of novel, host-specific IVSs in microbial 23S rDNA.

Phylogenetic analysis of fecal microbes from different host species is a common method for the identification of host-specific genetic markers. This method is based on alignment analysis against the DNA sequences of potential FST molecules of the target fecal microorganisms. Before the availability of NGS technologies, DNA cloning and Sanger DNA sequencing were the necessary steps for phylogenetic analysis. However, DNA cloning is time-consuming and subject to cloning bias. NGS provides a high-throughput, time-efficient, and cost-effective tool for the identification of host-specific FST markers.

The major challenge of using FST technologies for the safety of water and food is that the correlation between the FST markers/indicators and pathogens/diseases has not been well established [6,84]. Nevertheless, the future of FST should not be underestimated, especially in the area of food safety.

Acknowledgement

This work was supported by a USDA NIFA’s Evans-Allen Grant.
References


