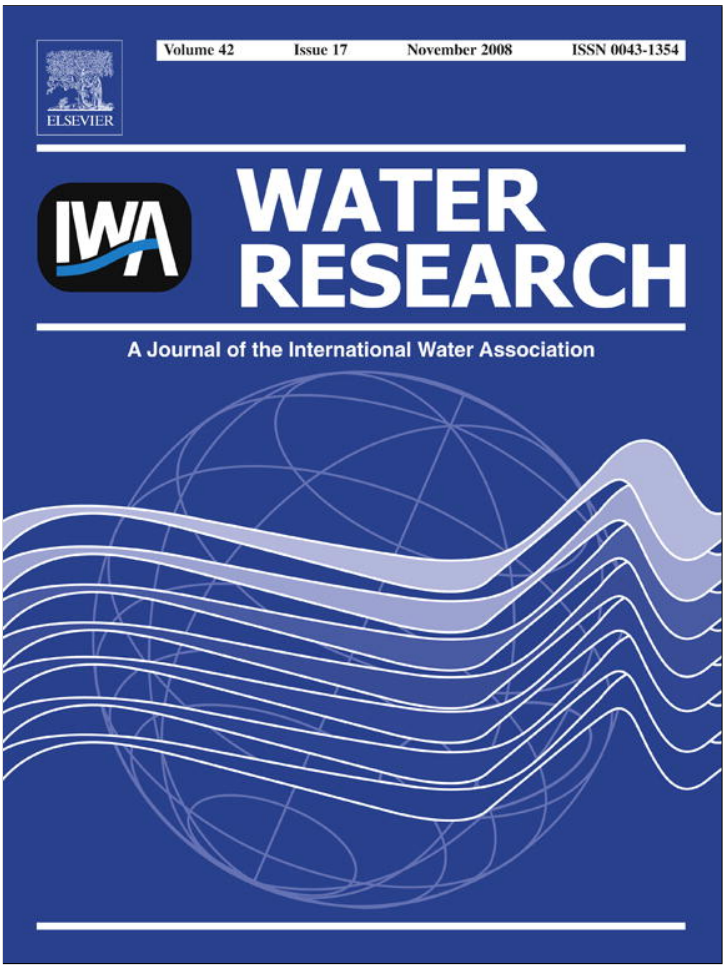

WATER RESEARCH 42 (2008) 4467–4476

Available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/watres

Optimization of DGGE community fingerprinting for characterizing *Escherichia coli* communities associated with fecal pollution

*Malak A. Esseili, Issmat I. Kassem, Von Sigler**

Department of Environmental Sciences, University of Toledo, 2801 West Bancroft Street, MS 604, Toledo, OH 43606, USA

ARTICLE INFO

Article history:
Received 14 April 2008
Received in revised form
23 July 2008
Accepted 4 August 2008
Published online 12 August 2008

Keywords:
DGGE
Community fingerprinting
Escherichia coli
Fecal pollution
Water quality

ABSTRACT

We evaluated the use of DGGE fingerprinting to differentiate communities of *Escherichia coli* from animal and geographic sources. An initial screening of 15 gene candidates revealed the ability of three target genes (*mdh*, *phoE* and *uidA*-4) to effectively differentiate *E. coli* communities originating in horses, pigs, geese and goats. Cluster and jackknife analyses performed on the communities from a more extensive number of hosts ($n = 150$) including humans (via raw sewage), horses, pigs, geese and cows revealed that the internal accuracy of classification of *E. coli* community fingerprints to their origin was similar for each of the three genes (85–86%). Each of the three genes were tested for their ability to associate *E. coli* source- and sink communities in two settings featuring contaminated water; (i) a stream receiving municipal wastewater effluent and (ii) a pond inhabited by geese. For each gene, DGGE fingerprints effectively matched effluent- and downstream *E. coli* communities (98–100% similarity) and excluded upstream communities, while communities from goose fecal material were 77–79% similar to communities in pond water, indicating fecal inputs from geese. Furthermore, each gene discriminated against *E. coli* communities from hosts non-indigenous to either setting. DGGE analysis of *E. coli* communities appears to be a promising tool to augment existing efforts aiming to address the dynamics of bacteria pollution in complex, natural environments.

1. Introduction

Despite increased monitoring and advances in detection methods, fecal pollution remains a persistent environmental challenge and a threat to public health. Primary contact with contaminated water is associated with increased incidences of gastrointestinal, respiratory, eye, ear, and skin illnesses (Pruss, 1998). While efforts are ongoing to minimize the contamination of natural waters, successful mitigation is limited when the source of the pollution is unknown (Stoner and Dorfman, 2007). Unknown pollution sources and pollution runoff are common and have accounted for 54 and 40%, respectively, of beach advisories and closings in the United States (Stoner and Dorfman, 2007). Since the most cost-effective measure to reduce frequently occurring fecal pollution is to identify and mitigate the pollution at its source (Simpson et al., 2002), a need exists for tools that can effectively detect and characterize fecal pathogens in the natural environment. Since no method currently exists that can rapidly detect the variety of pathogens associated with fecal pollution, indicator bacteria such as *Escherichia coli*, which are commonly found in animal fecal material, are typically used as a measure of fecal pollution (Santo Domingo and Sadowsky, 2007). The idea that some animal hosts are known to harbor unique *E. coli* strain

* *Corresponding author*. Tel.: +1 419 530 2897; fax: +1 419 530 4421.
E-mail address: von.sigler@utoledo.edu (V. Sigler).
0043-1354/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.watres.2008.08.002

assemblages (McLellan et al., 2003; McLellan, 2004, Anderson et al., 2006) has generated considerable interest in techniques such as community fingerprinting that can differentiate *E. coli* communities contaminating the natural environment (Farnleitner et al., 2000b; Stewart et al., 2003; Sigler and Pasutti, 2006). The ability to characterize and differentiate *E. coli* communities in a single fingerprint might not only facilitate the mitigation of fecal pollution, but also address the impact of environmental perturbations (i.e. pollution runoff, agricultural practices, land use changes) and seasonal, spatial and geographic variation on pathogen indicators.

Denaturing gradient gel electrophoresis (DGGE) analysis is a frequently used method for genetic fingerprinting of microbial communities (Muyzer, 1999). The method allows the reproducible differentiation of complex microbial communities by separating PCR amplicons that differ in DNA sequence by as little as one nucleotide (Myers et al., 1985; Muyzer, 1999). While DGGE analysis is most commonly used to analyze communities at the genus (or higher) taxonomic rank (Muyzer, 1999; Farnleitner et al., 2000b), analyses of communities at the species level, including *E. coli*, have been performed (Farnleitner et al., 2000a,b; Sigler and Pasutti, 2006). Since assemblages of *E. coli* strains tend to exhibit some degree of host specificity (Hamilton et al., 2006; Khatib et al., 2002, 2003; Ram et al., 2004), DGGE analysis has the potential to generate *E. coli* community fingerprints descriptive of a specific host- or geographic location. It follows that fingerprints of *E. coli* communities collected from differing contaminated environments could be compared as an initial step to determine the similarity, and therefore a coupling, between pollution sinks and potential sources.

The use of DGGE analysis to characterize *E. coli* communities relies on targeting a gene that meets three requirements. First, the gene must be *E. coli*-specific. Second, PCR amplification of the gene should result in DNA of appropriate size for DGGE analysis. Third, it must exhibit sequence polymorphism among differing *E. coli* strains that allows for the discrimination of *E. coli* communities of a specific host- or geographic location. Studies focusing on the detection of *E. coli* have resulted in numerous PCR primer sets that represent reasonable candidates for DGGE analysis. These include genes encoding 16S ribosomal RNA (16S rRNA) (Tsen et al., 1998; Sabat et al., 2000), 23S ribosomal RNA (23S rRNA) (Zwirglmaier et al., 2004), β-3-D-glucuronidase (*uidA*) (Bej et al., 1991; McDaniels et al., 1998), β-3-D-glucuronidase regulatory gene (*uidR*) (Bej et al., 1991), malate dehydrogenase (*mdh*) (Hsu and Tsen, 2001), lambda phage attachment site peptide (*lamB*) (Bej et al., 1990), glutamate decarboxylase (*gadAB*) (McDaniels et al., 1998), alanine racimase (*alr*) (Yokoigawa et al., 1999), maltose transport operon (*malB*) (Candrian et al., 1991), a transferase involved in Enterobacterial common antigen-ECA biosynthesis (*wecA*) (Bayardelle and Zafarullah, 2002) and an outer membrane phosphoporin (*phoE*) protein (Spierings et al., 1993). However, with the exception of one segment of *uidA* (Farnleitner et al., 2000b; Sigler and Pasutti, 2006), none have been evaluated for their potential to differentiate *E. coli* communities with regard to environmental water quality. Therefore, the purpose of this study was to comprehensively evaluate these genes for their suitability in DGGE-based differentiation of *E. coli* communities collected from primary hosts and natural waters. This was performed by assessing (i) the frequency of detection of each candidate gene within a collection of environmental *E. coli*, (ii) the ability of DGGE fingerprints generated from each gene to differentiate *E. coli* communities originating from different primary hosts, and (iii) the utility of DGGE to associate *E. coli* communities originating in potential sources with those in pollution sinks.

2. Materials and methods

2.1. Detection of target genes among environmental E. coli

2.1.1. Collection and identification of E. coli isolates

Water samples were obtained from three sites along Lake Erie Beach (LEB) in Maumee Bay State Park (OH) by inverting a sterile, 1 l bottle approximately 30 cm below the water surface. Samples were maintained on ice until analysis was performed (within 2 h of collection). A collection of *E. coli* isolates was generated by filtering an appropriate volume of water through sterile, nitrocellulose membranes (0.45 μm pore size), and then transferring membranes to Modified m-TEC agar (Difco, USA). Following incubation at 44.5 °C for 16 h, well-isolated, putative *E. coli* (purple colonies, $n = 176$) were transferred to eosin methylene blue agar (Remel, USA) and incubated at 37 °C for 24 h. Putative *E. coli* isolates were simultaneously subjected to PCR using primers specific for the 16S rRNA gene of *E. coli* (Sabat et al., 2000). All putative *E. coli* isolates displayed the expected phenotype for *E. coli* on eosin methylene blue agar and also generated a correctly sized PCR product (544 bp). A glycerol stock of each confirmed *E. coli* isolate was prepared and stored at −80 °C.

2.1.2. DNA isolation from E. coli and detection of target genes

Each of the confirmed *E. coli* isolates was grown for 16 h with shaking in Luria broth at 37 °C followed by centrifugation for 5 min at 4500 × *g*. The resulting cell pellet was suspended in 1 ml of DNA extraction buffer containing 50 mM NaCl, 50 mM Tris–HCl (pH = 7.6), 50 mM EDTA (pH = 8) and 5% SDS. The solution was transferred to a sterile tube containing 0.5 ml of glass beads (0.1 mm diameter) (Biospec Products, Inc., USA) and subjected to mechanical bead beating (15 s at 4 ms^{-1}) using a Fastprep bead beater (Thermosavant). The tubes were centrifuged for 1 min at 13,000 × *g*, and the supernatant was transferred to a new tube, followed by DNA purification and precipitation according to the phenol/chloroform method of Sambrook et al. (1989). Purified DNA was quantified by measuring A_{260}, while purity was measured by calculating A_{260}/A_{280}. The DNA was diluted with nuclease-free water to achieve a final concentration of 100 μg ml^{-1}, and stored at −20 °C. The detection frequency for each of the gene was assessed by PCR analysis of the DNA from each of the 176 *E. coli* isolates. One microliter of the isolated DNA (equivalent to 0.1 μg) was used in a 25 μl reaction mixture according to previously published PCR conditions specific for each primer set (Table 1). For each set of PCR reactions, DNA isolated from *E. coli* strain DH5α served as a positive control, while a negative control reaction contained nuclease-free water instead of DNA. PCR products were visualized following agarose gel electrophoresis and compared to a 100 bp DNA size standard

Table 1 – PCR primers used in this study

Gene	Primer sequence	PCR product (bp)	T_m (°C)	Reference
16S-1-f	5′-AAGAAGCTTGCTTCTTTGCTGAC-3′	544	72	Sabat et al., 2000
16S-1-r	5′-AGCCCGGGGATTTCACATCTGACTTA-3′			
16S-2-f	5′-GGGAGTAAAGTTATTACCTTTGCTC-3′	584	60	Tsen et al., 1998
16S-2-r	5′-TTCCCGAAGGCACATTCT-3′			
23S-1-f	5′–TGGTTCTCYCCGAAA–3′	670	50	Zwirglmaier et al., 2004
23S-1-r	5′-GCTTAAACCGGGACAACC-3′			
alr-f	5′-CTGGAAGAGGCTAGCCTGGACGAG-3′	366	72	Yokoigawa et al., 1999
alr-r	5′-AAAATCGGCACCGGTGGAGCGATC-3′			
gadAB-f	5′-ACCTGCGTTGCGTAAATA-3′	670	58	McDaniels et al., 1998
gadAB-r	5′-GGGCGGGAGAAGTTGATG-3′			
lamB-f	5′-CTGATCGAATGGCTGCCAGGCTCC-3′	309	60	Bej et al., 1990
lamB-r	5′-CAACCAGACGATAGTTATCACGCA-3′			
malB-f	5′-TCGCCACACGCTGACGCTGACCA-3′	596	65	Candrian et al., 1991
malB-r	5′-TTACATGACCTCGGTTTAGTTCACAGA-3′			
mdh-f	5′-ACTGAAAGGCAAACAGCCAAG-3′	392	60	Hsu and Tsen, 2001
mdh-r	5′-CGTTCTGTTCAAATGGCCTCAGG-3′			
phoE-f	5′-AAAGCCGTGGCACAGGCAAGCGT-3′	348	52	Spierings et al., 1993
phoE-r	5′-TCAATTTGTTATCGCTATCCAGTTGG-3′			
uidA-1-f	5′-AATAATCAGGAAGTGATGGAGCA-3′	586	60	Ram et al., 2004
uidA-1-r	5′-CGACCAAAGCCAGTAAAGTAGAA-3′			
uidA-2-f	5′-AAAACGGCAAGAAAAAGCAG-3′	147	50	Bej et al., 1991
uidA-2-r	5′-ACGCGTGGTTACAGTCTTGCG-3′			
uidA-3-f	5′-AAAAGCCAGACAGAGT-3′	623	58	McDaniels et al., 1998
uidA-3-r	5′-GCACAGCACATCCCCAAAGAG-3′			
uidA-4-f	5′-TATGGAATTTCGCCGATTTT-3′	166	50	Bej et al., 1991
uidA-4-r	5′-TGTTTGCCTCCCTGCTGCGG-3′			
uidR-f	5′-TGTTACGTCCTGTAGAAAGCCC-3′	152	59	Bej et al., 1991
uidR-r	5′-AAAACTGCCTGGCACAGCAATT-3′			
wecA-f	5′-GGTGTTCGGCAAGCTTTATCTCAG-3′	762	60	Bayardelle and Zafarullah, 2002
wecA-r	5′-GGTTAAATTGGGGCTGCCACCACG-3′			

(Promega, USA) to identify the size of the product. All the candidate gene targets were of appropriate size (≲600 bp) for DGGE analysis (Table 1). In the event that PCR generated a negative result, the reaction was repeated to confirm the negative reaction. It should be noted that with the exception of 16S-1 (Sabat et al., 2000), all primer sets evaluated in this study provide limited detection of *Shigella* spp., and in the case of *phoE*, detection of *Escherichia fergusonnii*. These bacteria are two of the most closely related species to *E. coli* (Whittam, 1989; Lawrence et al., 1991; Hariharan et al., 2007), and are also components of animal fecal material. Therefore, since all samples were treated equally, it was assumed that the systematic, limited co-detection of these bacteria would not compromise the evaluation.

2.2. DGGE analysis of host E. coli communities

2.2.1. Fecal material collection and DNA isolation

Fresh fecal material collected from randomly selected, individual geese ($n = 3$), goats ($n = 3$), pigs ($n = 3$) and horses ($n = 4$) served as a source of an *E. coli* assemblage used to screen the differentiating ability of the candidate genes. The goats, horses and pigs were housed separately at the University of Findlay Equestrian Farm (Findlay, OH) among a population of approximately 25 goats, 300 horses and 12 pigs. Freshly deposited goose fecal material was collected from among a population of approximately 120 geese inhabiting a local pond at the Toledo Botanical Garden. All fecal material was aseptically sampled into 50 ml sterile tubes with a flame-sterilized spatula, placed into a cooler and processed within 6 h of collection. To isolate DNA, 5 g of thoroughly mixed fecal material were suspended in 45 ml of 10 mM sodium phosphate buffer (pH 7.6) and dispersed by shaking for 1 h at 25 °C. One hundred μl of the suspension was inoculated onto Modified m-TEC media and incubated at 44.5 °C for 16 h. The resulting confluent purple lawn of *E. coli* was scraped from the plate (Khatib et al., 2002) and DNA was isolated from the cells as described above.

2.2.2. Generating E. coli community fingerprints with PCR-DGGE

PCR products were generated in 50 μl volumes using each primer set (Table 1) and 2 μl of DNA isolated from cells collected as described above. To facilitate subsequent PCR

product separation in DGGE, a GC-clamp (Myers et al., 1985) was added to each forward primer. DGGE was performed according to the method of Sigler and Pasutti (2006) with the exception of voltage and duration, which were set at 60 V and 16 h to optimize band clarity and separation. The optimal DGGE denaturant range for each PCR product was empirically determined by first performing DGGE of each PCR product in a gel containing a broad denaturing range of 20–60% (100% denaturant concentration is defined as 7 M urea and 40% deionized formamide), and then subsequently narrowing the denaturing range until optimized band clarity and separation were achieved. All PCR products were best separated in a denaturing range of 40–60% with the exception of *phoE* (20–35%) and *alr*, *malB* and *wecA* (40–50%). Additionally, the PCR product generated for *uidA*-1 exhibited visually superior band separation when analyzed in 6% polyacrlyamide (vs. 8% for the other genes). Following DGGE, gels were stained for 15 min with 50 ml of a 1:10,000 dilution of GelStar nucleic acid stain (Biowhittaker) and fingerprint images were documented using a Kodak Gel Logic 200 image analysis system. To facilitate fingerprint comparisons, a DGGE marker was loaded such that no more than four fingerprint lanes separated each marker lane. The DGGE marker was developed by combining equal volumes of the PCR amplicons generated by amplifying the 16S rDNA (Muyzer et al., 1993) of *Bacillus subtilis*, *E. coli* strain DH5α, *Pseudomonas aeruginosa*, *Ralstonia pickettii* and four unknown environmental bacteria. Fingerprint reproducibility was confirmed by repeating the DGGE analysis at least twice.

2.2.3. DGGE fingerprint image analysis

All fingerprint images were imported to GelCompar II software (version 4.5, Applied Maths) for pairwise analysis of fingerprint similarity. Fingerprints were normalized using the DGGE marker as an external reference, while bands that were visually determined to be common among at least five fingerprints in the gel were used as internal reference markers. Since the ''auto search bands'' function in GelCompar II often misidentified bands, we followed two approaches to ensure accurate band identification and fingerprint comparisons. First, multiple images of each gel were captured at increasingly higher exposure times in an effort to identify all bands prior to performing cluster analysis. Second, a comparison of both fingerprint intensity/curve- and band-based similarity coefficients was performed while simultaneously comparing band position tolerance (0.5 and 1%) and optimization (0 and 3%) settings in the software. In all cases, the Dice band-based coincidence index (Dice, 1945) in combination with 3% optimization and 1% position tolerance was found to be consistent with the visual inspection and resulted in highest similarity recognition (95%) among identical *E. coli* fingerprints generated on different gels.

Cluster analysis was performed on the resulting similarity matrix using the unweighted pair group method with arithmetic means (UPGMA) algorithm, resulting in dendrograms that graphically displayed the similarities among fingerprints. The candidate genes that differentiated and clustered the *E. coli* communities according to their primary host were further screened on a larger collection of *E. coli* communities originating from raw sewage ($n = 30$), horses ($n = 30$), pigs ($n = 30$), geese ($n = 30$) and cows ($n = 30$). Ten samples of 150 ml of raw sewage were obtained from the influents of each of three wastewater treatment plants in Ohio. Horse and goose fecal material were sampled from the same locations as described above, while pig and cow fecal material were collected from two differing livestock operations in Ohio. Lawns of *E. coli* from raw sewage were generated by filtering 100 ml of each sample followed by enrichment as described above. Following DGGE analysis of the *E. coli* communities, jackknife analysis (McLellan et al., 2003) was used to determine how accurately DGGE fingerprints of *E. coli* communities could be assigned to each host group. To perform jackknife analysis, *E. coli* community fingerprints were manually assigned to their respective host. The software then removed each fingerprint from the data set individually and queried the data set to determine from which host group the fingerprint was most similar. User-set parameters included maximum-similarity coefficients and ties spread equally among host groups. The internal accuracy of classification was calculated as the percentage of community fingerprints assigned to the host to which the sample was known to belong. To avoid biases in the jackknife analysis that could result from a single host being represented by a disproportionately large number of very similar fingerprints (Johnson et al., 2004), the fingerprint collection was ''decloned'', and only one representative fingerprint from clusters that were >95% similar was analyzed.

2.2.4. Matching a known pollution source to its sink

Genes that exhibited high internal accuracy were further assessed under two natural settings for their ability to match *E. coli* communities in a pollution sink to a known source. The first setting included water samples collected from (i) 10 and 20 m upstream of a wastewater treatment plant effluent discharge,(representing a background) (ii) the effluent discharge pipe (representing a source) and (iii) 10, 20 and 30 m downstream (representing a sink) of the effluent discharge. The second setting included water samples ($n = 5$) collected from a pond (sink) inhabited by a community of approximately 80 geese, and samples of freshly deposited goose fecal material (source, $n = 4$). To determine the optimum volume of water necessary to generate representative fingerprints, varying volumes (10, 100, 200, 300, 400, 500, 600, 800 and 1000 ml) were filtered and processed as described above. All volumes generated identical fingerprints. Therefore, a mid-range volume of 500 ml of each water sample from each site, in addition to each fecal material sample, was analyzed as described above. *E. coli* community fingerprints originating from animals not indigenous to the creek or pond environment (raw sewage, goats, horses and pigs) were also included in the cluster analysis to assess the ability of the method to discriminate against known, noncontributing sources.

3. Results

3.1. Detection of target genes among environmental E. coli isolates

PCR analysis using the 15 primer sets revealed that all candidate genes were detected in each *E. coli* isolate ($n = 176$) with

the exception of 16S-2 (not detected in five isolates) and *uidA*-3 (eight isolates) (data not shown). Since the targeted segments of these genes were not detected among 100% of the tested isolates, they were excluded from further analysis. In every case of positive gene identification, one band of expected size (bp) was detected.

3.2. Generating E. coli community fingerprints with DGGE

An initial DGGE analysis of *E. coli* communities from four hosts ($n = 13$ individuals) (geese, goats, horses and pigs) using the remaining 13 genes revealed that the number of bands (a proxy for sequence types) was variable and dependent on the gene targeted. Specifically, among all animals, fingerprints consisted of 33 (*phoE*), 23 (*uidA*-4), 22 (*mdh*), 18 (*uidR*), 15 (*uidA*-2 and *malB*), 13 (*lamB, alr*), 11 (23S, *gadAB*, *uidA*-1), 7 (*wecA*) and 3 (16S-1) bands.

Cluster analysis of DGGE fingerprints was used to group *E. coli* communities based on their host-origin. Fingerprints generated targeting *mdh*, *phoE* and *uidA*-4 resulted in four clusters, each corresponding to only the animal host from which the *E. coli* communities originated. The remaining 10 genes exhibited less discrimination. Fingerprints generated targeting *alr*, *lamB*, *gadAB*, *malB*, *uidA*-1, *uidA*-2 and *uidR* discriminated *E. coli* communities from one host, while the remaining three genes 16S-1, 23S, and *wecA* resulted in fingerprints that exhibited no host-discrimination (data not shown).

Since only genes that provided the highest degree of host-discrimination were desired for future applications, discriminant analyses of *E. coli* community fingerprints generated from a larger sample size (five hosts, $n = 150$ individuals) were limited to data generated using *mdh*, *phoE* and *uidA*-4. Each of the three genes exhibited a similar internal accuracy of fingerprint classification among the five hosts (86, 85 and 85% for *phoE*, *uidA*-4 and *mdh,* respectively). However, higher rates of internal accuracy were observed for *E. coli* communities from raw sewage, horses and cows, as compared with those from pigs (Table 2). Pig *E. coli* communities were most often misclassified as goose communities (12, 27 and 34% misclassification for *phoE*, *mdh* and *uidA*-4, respectively). Goose *E. coli* communities were most often correctly classified by using *uidA*-4 (68%), while *phoE*, *mdh* and *uidA*-4 often misclassified them as communities from pigs (10, 33 and 27%, respectively) and horses (23% using *phoE*) (Table 2).

3.3. Using the target genes to match a pollution source to its sink

DGGE analysis of *mdh*, *phoE* and *uidA*-4 was further assessed for its ability to match *E. coli* communities originating from a known source to a pollution sink. In setting one, cluster analysis of fingerprints generated from *mdh*, *phoE* and *uidA*-4 consistently revealed that *E. coli* communities that originated in WWTP effluent contributed to downstream pollution. Effluent *E. coli* communities were between 98 and 100% similar to those from downstream sites (Fig. 1A), while *E. coli* communities in upstream sites exhibited relatively lower similarity (65–68%) to effluent- and downstream community fingerprints (Fig. 1A). In the second setting, fingerprints generated with *mdh*, *phoE* and *uidA*-4 consistently indicated that *E. coli* communities originating from goose fecal material and water samples were between 71 and 79% similar and lacked distinct geographic (water) or host (goose)-specific clustering (Fig. 1B).

The ability of DGGE fingerprinting to discern *E. coli* communities from true, indigenous sources from those originating in non-indigenous hosts was assessed by performing cluster analysis of fingerprints generated from both scenarios with *E. coli* community fingerprints from raw sewage and hosts (goats, horses and pigs) known not to inhabit the environment of the two settings. Cluster analysis of *E. coli* community fingerprints generated by each gene showed that none of the non-indigenous host community fingerprints were observed to cluster within the fingerprints representing the effluent-stream or goose-pond water settings (Fig. 2).

4. Discussion

We investigated the use of 15 gene targets in DGGE analysis to differentiate *E. coli* communities in fecal material and polluted water. Our analyses focused on identifying genetic targets that were (i) frequently detected in a collection of environmental *E. coli*, (ii) capable of discriminating *E. coli* communities among several primary hosts and (iii) able to match pollution sources to their sinks in test settings.

Table 2 – Jackknife analysis results

	mdh					*uidA*-4					*phoE*				
	Sewage (3)	Horse (10)	Cow (15)	Pig (24)	Goose (27)	Sewage (5)	Horse (13)	Cow (23)	Pig (25)	Goose (30)	Sewage (3)	Horse (17)	Cow (22)	Pig (25)	Goose (30)
Sewage	*100*					*100*					*100*				
Horse		*100*					*100*					94	4		23
Cow			*100*		12			92		5		6	91	2	7
Pig				69	33			4	66	27				*86*	10
Goose				27	55			4	34	68			5	12	60

Italic numbers represent the percentage of *E. coli* communities correctly assigned to each source group. The numbers in parentheses represent the number of fingerprints included in the Jackknife analysis following ''decloning'' as described in the text.

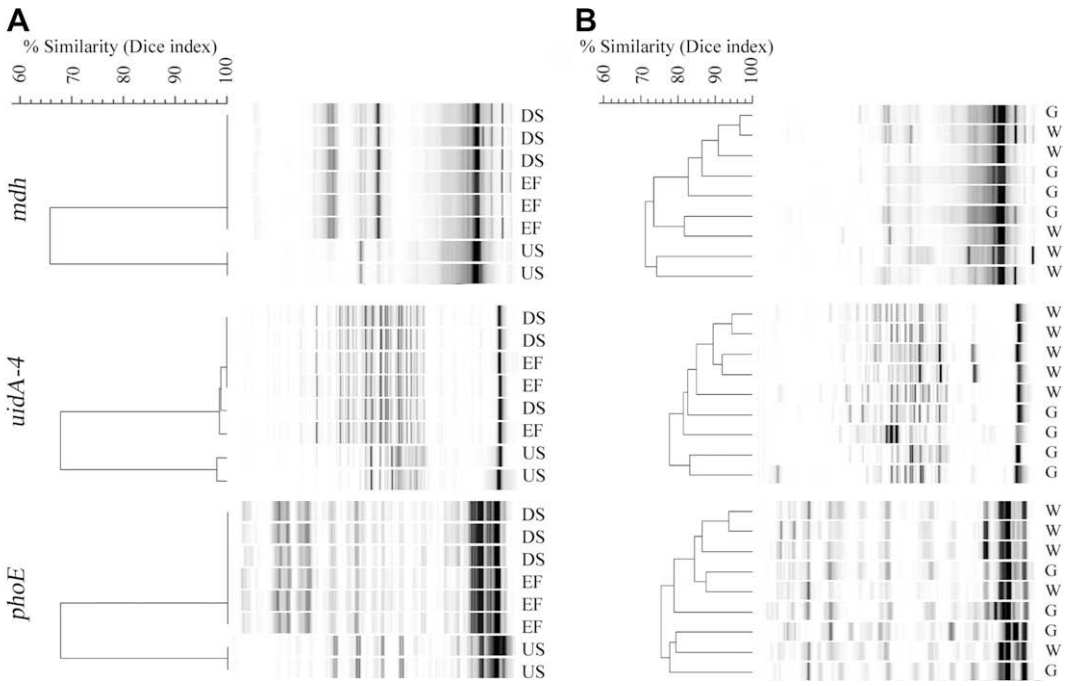

Fig. 1 – Cluster analysis of *E. coli* community DGGE fingerprints generated with *mdh*, *uidA* and *phoE* of (A) the WWTP effluent-creek scenario (setting one) and (B). the goose-water scenario (setting two). Legend: US, upstream water; DS, downstream water; EF, effluent; G, goose fecal material; W, pond water.

With the exception of *uidA*, the genes evaluated in this study were initially used as diagnostic markers to detect *E. coli* in clinical- and food safety studies. Therefore, to appropriately test the detection of the genes within an environmental context, it was necessary to generate a collection of *E. coli* that mimicked a natural, contaminating population, i.e. one originating from a broad cross-section of different host- or geographic sources. Our results indicated that each gene was detected among all 176 *E. coli* in our environmental collection with the exception of 16S-2 and *uidA*-3. While these two genes are essential components of the *E. coli* genome, limited detection of the 16S rRNA gene segment, in particular, was observed previously among a library of clinical strains (Tsen et al., 1998), perhaps the result of strain-specific differences in 16S rRNA genes (Ehresmann et al., 1975; Martınez-Murcia et al., 1999). For two primary reasons, we were confident that the collection of *E. coli* isolates used to evaluate the genes was appropriate. First, Lake Erie Beach is frequently impacted by bacterial pollution that originates from several local contributors (Lauber et al., 2003; Huang, 2007), possibly including waterfowl, suburban and agricultural runoff, faulty septic tanks, a horse farm and other wildlife. Second, of the 176 *E. coli* isolates in the collection, 66 unique fingerprints were identified by BOX-PCR fingerprinting (discriminated at 95% similarity), indicating that our *E. coli* community collection was genetically diverse (data not shown). The combination of geographic and host variability, and the genetic diversity within the library suggested that the *E. coli* isolates collected from Lake Erie Beach were of diverse host-origin and could subsequently be used to investigate the environmental detection of the genes.

An initial screening showed that cluster analysis of fingerprints generated with *mdh*, *phoE* and *uidA*-4 resulted in the discrimination of the *E. coli* communities collected from four primary hosts (horses, pigs, geese and goats). In contrast, analysis of the remaining genes resulted in clusters containing communities originating from multiple hosts. DGGE analysis operates on the principle that DNA segments of similar size can be separated based on sequence differences, resulting in bands that represent differing gene sequence types (Muyzer, 1999; Myers et al., 1985). Multiple sequence types of *uidA* have been identified previously in DGGE-based analyses of *E. coli* in freshwater (10 types among 47 isolates; Farnleitner et al., 2000a), freshwater and sediment (15 among 175 isolates; Sigler and Pasutti, 2006) and soil (nine among 205 isolates; Lasalde et al., 2005). Sequence-based surveys of *mdh*, from 29 pathogenic *E. coli* (Pupo et al., 1997) and 21 *E. coli* isolated from rat fecal material (Pupo et al., 2000) revealed 31 *mdh* sequence polymorphisms, 18 of which were located in the 392 bp *mdh* segment used in DGGE analysis in the current study. Additionally, Boyd et al. (1994) identified 40 polymorphic nucleotides within the entire *mdh* gene (939 bp) among 20 *E. coli* strains in the ECOR collection. While no previous studies have characterized polymorphisms in *phoE*, the gene is known to encode eight cell surface hypervariable regions (Nikaido, 2003) that undergo rapid mutational alterations as they interact with elements in the external environment such as antibodies, bacteriocins and phages (Nikaido, 2003). The utility of *mdh*, *phoE* and *uidA*-4 was also reflected in jackknife analyses performed on a larger sample size from five primary hosts, which indicated that DGGE analyses of the three genes could effectively classify *E. coli* communities from raw sewage, horses and cows to their

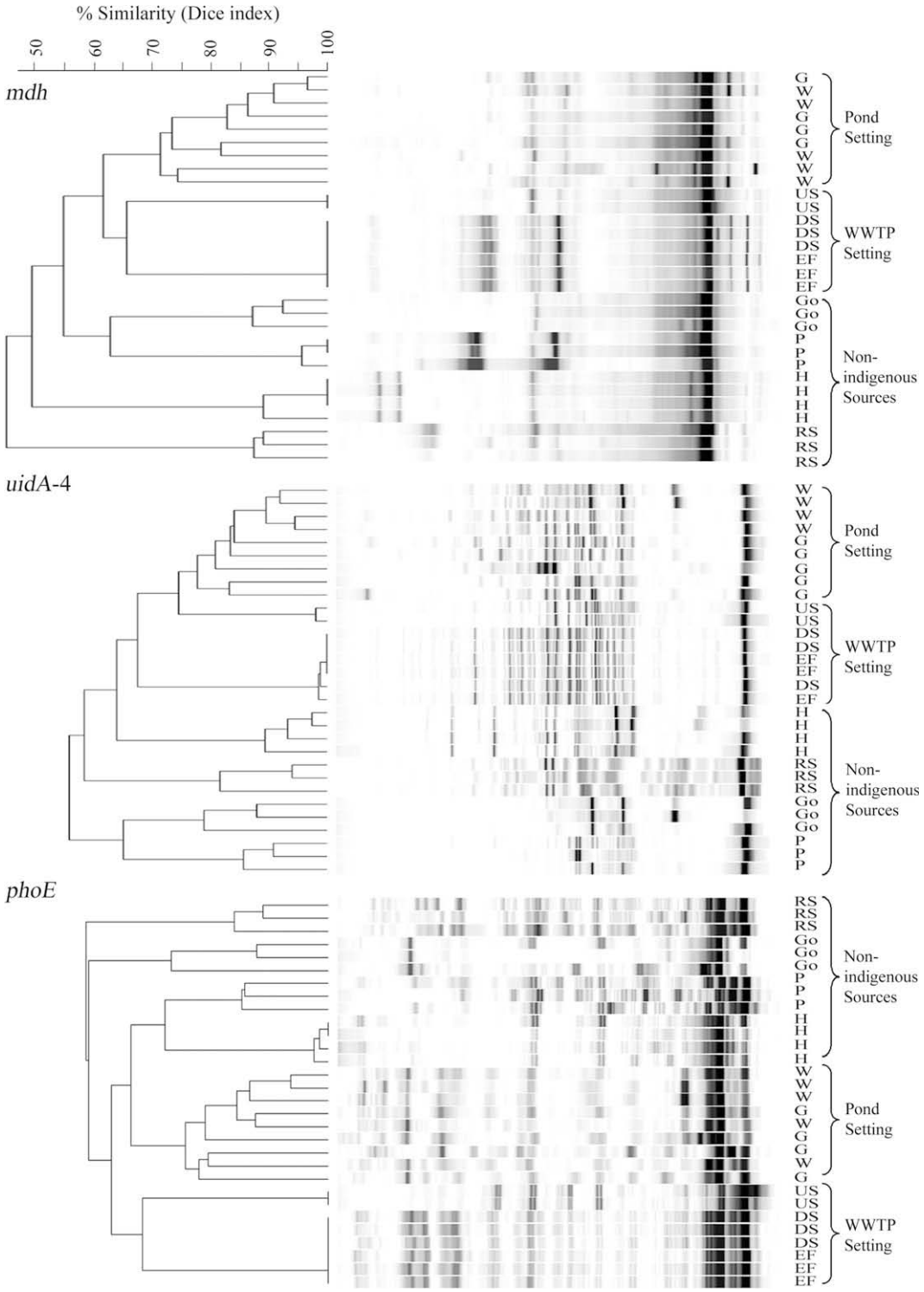

Fig. 2 – Cluster analysis of *E. coli* community fingerprints from the goose-water and effluent-creek settings, as compared with fingerprints generated from *E. coli* communities of known, non-indigenous sources including raw sewage (RS), goats (Go), horses (H) and pigs (P). Legend is consistent with that of Fig. 1.

proper host group. It is noteworthy that a moderate rate of incorrect classification was also observed, particularly among geese. Classification inaccuracy could result among hosts exhibiting relatively high gut *E. coli*, which would subsequently require a greater sampling effort to acquire fecal samples that contain diagnostic *E. coli* communities. One established driver of gut *E. coli* diversity is alterations in the host diet (Hartel et al., 2003; Bettelheim et al., 2005). In this study, the population of geese was wild and likely fed on a diverse and changing diet, facilitating a greater diversity of

gut *E. coli*, whereas the cows, horses and pigs were fed a standard livestock diet.

E. coli communities in natural waters often result from multiple inputs (Anderson et al., 2006). This factor, in addition to the recognized drawbacks of library-dependent and independent approaches (Santo Domingo and Sadowsky, 2007) limits the effectiveness of many of our current methods to accurately trace the host- or geographic origin of *E. coli*. Although efforts to adopt a community approach for studying environmental *E. coli* have been limited, DGGE analysis has been used to characterize *E. coli* communities in polluted waters (Farnleitner et al., 2000b; Sigler and Pasutti, 2006). Our current results expand on these findings and show that DGGE analyses targeting *mdh*, *phoE* and *uidA*-4 can be useful to identify a geographic origin of bacteria pollution. Specifically, DGGE fingerprints generated using each of the three genes provided strong evidence that a freshwater stream was contaminated with *E. coli* communities originating from the WWTP effluent (Fig. 1A). We are aware that high similarity (≥98%) between the *E. coli* communities in the effluent- and downstream sites is not an exhaustive indicator of the effluent's contribution to the pollution. However, the fact that the upstream *E. coli* communities were, at most, 68% similar to those in the effluent- and downstream sites provided substantial evidence of the effluent's contribution to the downstream pollution. While the upstream *E. coli* communities were relatively dissimilar to effluent- and downstream communities, DGGE did reveal some contribution from the upstream sites (68% similarity), suggesting the presence of other sources of pollution, such as wild animals or overland inputs occurring upstream of the treatment plant. In contrast with the first setting, the second setting (a small pond inhabited by approximately 120 geese) represented a more diffused discharge of *E. coli* into the secondary habitat. This was reflected in the results by a lack of host (goose) or geographic (pond)-specific clustering of *E. coli* community fingerprints (Fig. 1B). The marked difference between the two settings in the ability of DGGE analysis to link sources with sinks is likely the result of differing pollution pathways. For example, the WWTP effluent discharged directly into the creek, resulting in a relatively concentrated, point source. Alternatively, goose fecal inputs in the pond setting likely result from a combination of (i) direct deposition into the pond, and (ii) more diffuse introduction, i.e. as runoff mixed with the fecal material from animal sources other than geese. Decreased similarity between source- and sink *E. coli* communities can result from the mixing of various host contributions (Hartel et al., 2003) and because of changes in the composition of *E. coli* communities during transition from the primary- to the secondary environment (Gordon et al., 2002; Anderson et al., 2006).

The overall utility of DGGE analysis should be measured by its ability to characterize *E. coli* communities contributing to a polluted environment (i.e. true positives). In addition, the ability to discern against noncontributing sources (i.e. false positives) is an equally important quality. It is important to note that heteroduplex formation and the co-migration of bands containing DNA of different sequences but similar melting behavior can lead to misinterpretations of community structure (Von Wintzingerode et al., 1997; Cassamayor et al., 2000). Therefore, evaluations of DGGE fingerprints should be performed cautiously. Nevertheless, on at least a small scale, this study showed that DGGE analysis discriminated against non-indigenous hosts, while simultaneously linking *E. coli* communities from polluted water with their respective sources (Fig. 2), suggesting the potential utility of the method in *E. coli* community studies and pollution source tracking applications. In addition, cophenetic correlation coefficients (a parameter used to express how well the resulting dendrogram branching represents the similarity matrix) of each cluster were found to range between 80 and 100%, indicating appropriate stability of each cluster.

Although DGGE analysis is primarily viewed as a culture-independent method, an enrichment step was included prior to DNA isolation to increase the *E. coli* density in the analyzed samples. Enrichment was included following assays that revealed inconsistent PCR amplification of DNA from non-enriched samples. Furthermore, it has been shown that enrichment effectively dilutes PCR inhibitors commonly found in fecal material (Kreader, 1996). It was also observed that without enrichment, the limit of PCR-based detection of each gene varied as a function of *E. coli* density (data not shown). While this was not surprising, without enrichment it would have been difficult to compare the discriminatory performance of these genes, especially in fecal or water samples exhibiting low *E. coli* densities. We are aware of the methodological biases associated with culturing bacteria (Amann et al., 1995). Therefore, we standardized our culturing methods throughout the entire series of experiments, including consistent enrichment conditions and *E. coli*-selective culturing procedures.

5. Conclusions

- Three gene segments (*mdh*, *uidA*-4 and *phoE*) were identified that are suitable for DGGE-based differentiation of *E. coli* communities in both primary host and polluted water.
- DGGE analysis of *mdh*, *uidA*-4 and *phoE* resulted in descriptive representations of *E. coli* communities and facilitated discrimination of communities originating from several animal hosts.
- In simple but natural settings, DGGE analysis exhibited the potential to link pollution sources with sinks, while discriminating against noncontributing sources.
- DGGE analysis of *E. coli* communities targeting *mdh*, *uidA*-4 and *phoE* appears to be a promising tool to augment existing efforts aiming to address the dynamics of bacteria pollution in complex, natural environments

Acknowledgment

We would like to thank Dr. Michael Kerns (University of Findlay Equestrian Farm, Findlay, OH) and April Ames Coordinator (University of Toledo), for facilitating access to the farm and WWTPs. This work was supported by the Ohio Lake Erie Commission's Lake Erie Protection Fund Large Grants Program (grant number LEPF 04-21), the USDA CSREES

(grant number 2007-31100-06039) and the National Science Foundation Research Experience for Undergraduates (REU) program (grant number 0522522).

REFERENCES

Amann, R.I., Ludwig, W., Schleifer, K.H., 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. 59 (1), 143–169.

Anderson, M.A., Whitlock, J.E., Harwood, V.J., 2006. Diversity and distribution of *Escherichia coli* genotypes and antibiotic resistance phenotypes in feces of humans, cattle, and horses. Appl. Environ. Microbiol. 72 (11), 6914–6922.

Bayardelle, P., Zafarullah, M., 2002. Development of oligonucleotide primers for the specific PCR-based detection of the most frequent *Enterobacteriaceae* species DNA using wec gene templates. Can. J. Microbiol. 48 (2), 113–122.

Bej, A.K., Steffan, R.J., DiCesare, J., Haff, L., Atlas, R.M., 1990. Detection of coliform bacteria in water by polymerase chain reaction and gene probes. Appl. Environ. Microbiol. 56 (2), 307–314.

Bej, A.K., DiCesare, J.L., Haff, L., Atlas, R.M., 1991. Detection of *Escherichia coli* and *Shigella* spp. in water by using the polymerase chain reaction and gene probes for uid. Appl. Environ. Microbiol. 57 (4), 1013–1017.

Bettelheim, K.A., Kuzevski, A., Gilbert, R.A., Krause, D.O., McSweeney, C.S., 2005. The diversity of *Escherichia coli* serotypes and biotypes in cattle faeces. Appl Microbiol. 98 (3), 699–709.

Boyd, E.F., Nelson, K., Wang, F.S., Whittam, T.S., Selander, R.K., 1994. Molecular genetic basis of allelic polymorphism in malate dehydrogenase (mdh) in natural populations of *Escherichia coli* and *Salmonella enterica*. Proc. Natl. Acad. Sci. U.S. A. 91 (4), 1280–1284.

Candrian, U., Furrer, B., Höfelein, C., Meyer, R., Jermini, M., Lüthy, J., 1991. Detection of *Escherichia coli* and identification of enterotoxigenic strains by primerdirected enzymatic amplification of specific DNA sequences. Int. J. Food Microbiol. 12 (4), 339–352.

Casamayor, E.O., Schäfer, H., Baneras, L., Pedros-Alio, C., Muyzer, G., 2000. Identification of and spatio-temporal differences between microbial assemblages from two neighboring sulfurous lakes: comparison by microscopy and denaturing gradient gelelectrophoresis. Appl. Environ. Microbiol. 66 (2), 499–508.

Dice, L.R., 1945. Measures of the amount of ecological association between species. Ecol 26 (1), 297–302.

Ehresmann, C., Stiegler, P., Mackie, G.A., Zimmermann, R.A., Ebel, J.P., Fellner, P., 1975. Primary sequence of the 16S ribosomal RNA of *Escherichia coli*. Nucleic Acids Res. 2 (2), 265–278.

Farnleitner, A.H., Kreuzinger, N., Kavka, G.G., Grillenberger, S., Rath, J., Mach, R.L., 2000a. Comparative analysis of denaturing gradient gel electrophoresis and temporal temperature gradient gel electrophoresis in separating *Escherichia coli* uidA amplicons differing in single base substitutions. Lett. Appl. Microbiol. 30 (6), 427–431.

Farnleitner, A.H., Kreuzinger, N., Kavka, G.G., Grillenberger, S., Rath, J., Mach, R.L., 2000b. Simultaneous detection and differentiation of *Escherichia coli* populations from environmental freshwaters by means of sequence variations in a fragment of the beta-D-glucuronidase gene. Appl. Environ. Microbiol. 66 (4), 1340–1346.

Gordon, D.M., Bauer, S., Johnson, J.R., 2002. The genetic structure of *Escherichia coli* populations in primary and secondary habitats. Microbiology 148 (5), 1513–1522.

Hamilton, M.J., Yan, T., Sadowsky, M.J., 2006. Development of goose- and duckspecific DNA markers to determine sources of *Escherichia coli* in waterways. Appl. Environ. Microbiol. 72 (6), 4012–4019.

Hariharan, H., López, A., Conboy, G., Coles, M., Muirhead, T., 2007. Isolation of *Escherichia fergusonii* from the feces and internal organs of a goat with diarrhea. Can. Vet. J. 48 (6), 630–631.

Hartel, P.G., Summer, J.D., Segars, W.I., 2003. Deer diet affects ribotype diversity of *Escherichia coli* for bacterial source tracking. Water Res. 37 (13), 3263–3268.

Hsu, S.C., Tsen, H.Y., 2001. PCR primers designed from malic acid dehydrogenase gene and their use for detection of *Escherichia coli* in water and milk samples. Int. J. Food Microbiol. 64 (1), 1–11.

Huang, X., 2007. Identification of Putative Geographic Sources of Bacterial Pollution in Lake Erie by Molecular Fingerprinting. M. S. thesis. University of Toledo, Ohio.

Johnson, L.K., Brown, M.B., Carruthers, E.A., Ferguson, J.A., Dombek, P.E., Sadowsky, M.J., 2004. Sample size, library composition, and genotypic diversity among natural populations of *Escherichia coli* from different animals influence accuracy of determining sources of fecal pollution. Appl. Environ. Microbiol. 70 (8), 4478–4485.

Khatib, L.A., Tsai, Y.L., Olson, B.H., 2003. A biomarker for the identification of swine fecal pollution in water, using the STII toxin gene from enterotoxigenic *Escherichia coli*. Appl. Microbiol. Biotechnol. 63 (2), 231–238.

Khatib, L.A., Tsai, Y.L., Olson, B.H., 2002. A biomarker for the identification of cattle fecal pollution in water using the LTIIa toxin gene from enterotoxigenic *Escherichia coli*. Appl. Microbiol. Biotechnol. 59 (1), 97–104.

Kreader, C.A., 1996. Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. Appl. Environ. Microbiol. 62 (3), 1102–1106.

Lasalde, C., Rodriguez, R., Toranzos, G.A., Smith, H.H., 2005. Heterogeneity of uidA gene in environmental *Escherichia coli* populations. J. Water Health 3 (3), 297–304.

Lauber, C.L., Glatzer, L., Sinsabaugh, R.L., 2003. Prevalence of pathogenic *Escherichia coli* in recreational waters. J. Great Lakes Res. 29 (1), 301–306.

Lawrence, J.G., Ochman, H., Hartl, D.L., 1991. Molecular and evolutionary relationships among enteric bacteria. J Gen. Microbiol. 137 (8), 1911–1921.

Martınez-Murcia, A.J., Anton, A.I., Rodrıguez-Valera, F., 1999. Patterns of sequence variation in two regions of the 16S rRNA multigene family of *Escherichia coli*. Int. J. Sys. Bacteriol 49 (2), 601–610.

McDaniels, A.E., Rice, E.W., Reyes, A.L., Johnson, C.H., Haugland, R.A., Stelma, G.N., 1998. Confirmational identification of *Escherichia coli*, a comparison of genotypic and phenotypic assays for glutamate decarboxylase and beta-D-glucuronidase. Appl. Environ. Microbiol. 64 (10), 4113–4119.

McLellan, S.L., 2004. Genetic diversity of *Escherichia coli* isolated from urban rivers and beach water. Appl Environ Microbiol. 70 (8), 4658–4665.

McLellan, S.L., Daniels, A.D., Salmore, A.K., 2003. Genetic characterization of *Escherichia coli* populations from host sources of fecal pollution by using DNA fingerprinting. Appl. Environ. Microbiol. 69 (5), 2587–2594.

Muyzer, G., de Waal, E.C., Uitterlinden, A.G., 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl. Environ. Microbiol. 59 (3), 695–700.

Muyzer, G., 1999. DGGE/TGGE a method for identifying genes from natural ecosystems. Curr. Opin. Microbiol. 2 (3), 317–322.

Myers, R.M., Fischer, S.G., Lerman, L.S., Maniatis, T., 1985. Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. Nucleic Acids Res. 13 (9), 3131–3145.

Nikaido, H., 2003. Molecular basis of bacterial outer membrane permeability revisited. Microbiol. Mol. Boil. Rev. 6 (4), 593–656.
Pruss, A., 1998. Review of epidemiological studies on health effects from exposure to recreational water. Int. J. Epidemiol. 27 (1), 1–9.
Pupo, G.M., Karaolis, D.K., Lan, R., Reeves, P.R., 1997. Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and mdh sequence studies. Infect. Immun. 65 (7), 2685–2692.
Pupo, G.M., Lan, R., Reeves, P.R., Baverstock, P.R., 2000. Population genetics of *Escherichia coli* in a natural population of native Australian rats. Environ. Microbiol. 2 (6), 594–610.
Ram, J.L., Ritchie, R.P., Fang, J., Gonzales, F.S., Selegean, J.P., 2004. Sequence based source tracking of *Escherichia coli* based on genetic diversity of beta-glucuronidase. J. Environ. Qual. 33 (3), 1024–1032.
Sabat, G., Rose, P., Hickey, W.J., Harkin, J.M., 2000. Selective and sensitive method for PCR amplification of *Escherichia coli* 16S rRNA genes in soil. Appl. Environ. Microbiol. 66 (2), 844–849.
Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular cloning: a Laboratory Manual, second ed. cold Spring Harbor Laboratory Press, New York.
Santo Domingo, J.W., Sadowsky, M.J., 2007. Microbial Source Tracking: Emerging Issues in Food Safety, first ed. ASM Press, Washington, DC.
Sigler, V., Pasutti, L., 2006. Evaluation of denaturing gradient gel electrophoresis to differentiate *Escherichia coli* populations in secondary environments. Environ. Microbiol. 8 (10), 1703–1711.
Simpson, J.M., Santo Domingo, J.W., Reasoner, D.J., 2002. Microbial source tracking: state of the science. Environ. Sci. Technol. 36 (24), 5279–5288.
Spierings, G.C., Ockhuijsen, C., Hofstra, H., Tommassen, J., 1993. Polymerase chain reaction for the specific detection of *Escherichia coli/Shigella*. Res. Microbiol. 144 (1), 557–564.
Stewart, J.R., Ellender, R.D., Gooch, J.A., Jiang, S., Myoda, S.P., Weisberg, S.B., 2003. Recommendations for microbial source tracking: lessons from a methods comparison study. J. Water Health 1 (4), 225–231.
Stoner, N., Dorfman, M., 2007. Testing the Waters: a Guide to Water Quality at Vacation Beaches. Natural Resources Defense Council. Available from: http://www.nrdc.org/water/oceans/ttw/ttw2007.pdf.
Tsen, H.Y., Lin, C.K., Chi, W.R., 1998. Development and use of 16S rRNA gene targeted PCR primers for the identification of *Escherichia coli* cells in water. J. Appl. Microbiol. 85 (3), 554–560.
Whittam, T.S., 1989. Clonal dynamics of *Escherichia coli* in its natural habitat. Antonie Van Leeuwenhoek 55 (1), 23–32.
Von Wintzingerode, F., Göbel, U.B., Stackebrandt, E., 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol. Rev. 21 (3), 213–229.
Yokoigawa, K., Inoue, K., Okubo, Y., Kawai, H., 1999. Primers for amplifying an alanine racemase gene fragment to detect *E. coli* strains in foods. J. Food Sci. 64 (1), 571–575.
Zwirglmaier, K., Ludwig, W., Schleifer, K.H., 2004. Improved method for polynucleotide probe-based cell sorting, using DNA-coated microplates. Appl. Environ. Microbiol. 70 (1), 494–497.