Environmental Microbiology (2006) doi:10.1111/j.1462-2920.2006.01105.x

Evaluation of denaturing gradient gel electrophoresis to differentiate *Escherichia coli* populations in secondary environments

Von Sigler[1]* and Lauren Pasutti[2]
[1]*Department of Earth, Ecological and Environmental Sciences, University of Toledo, Toledo, OH 43606, USA.*
[2]*Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824, USA.*

Summary

The development of methodology to differentiate mixed populations of *Escherichia coli* in the secondary habitat might improve monitoring of fecal pollution indicators and facilitate the development of strategies to mitigate bacterial pollution. The objective of this study was to determine the ability of denaturing gradient gel electrophoresis (DGGE) to differentiate mixed assemblages of *E. coli* in the natural environment. After confirming the identity of 184 environmental bacterial isolates as *E. coli*, each was subjected to polymerase chain reaction (PCR) of the β-glucuronidase gene (*uidA*) followed by DGGE fingerprinting. The ability of DGGE to discriminate individual isolates at the strain level was determined by comparing fingerprints to those resulting from a standard, library-dependent fingerprinting method, BOX-PCR. Computerized analysis of fingerprints indicated that DGGE and BOX-PCR identified 15 and 21 unique phylotypes respectively. Rank–abundance plots comparing the numerical distribution of unique *E. coli* phylotypes detected by both methods revealed no difference in resolution at the population level. In water and sediment samples from two beaches, DGGE effectively distinguished indigenous *E. coli* populations with an average rate of correct classification (site-based) of 83%. Denaturing gradient gel electrophoresis of *uidA* genes isolated and PCR-amplified from environmental samples appears to be an effective tool to differentiate unique *E. coli* populations and should be useful to characterize *E. coli* dynamics in the secondary environment.

Received 22 July, 2005; accepted 31 March, 2006. *For correspondence. E-mail von.sigler@utoledo.edu; Tel. (+1) 419 530. 2897; Fax (+1) 419 530.4421.

Introduction

A recent response by the American Society for Microbiology (ASM) to the United States Environmental Protection Agency (USEPA)'s recently proposed water quality criteria for coastal and Great Lakes waters (USEPA, 2004) stressed the need for studies that focus on *Escherichia coli* population dynamics and ecology (ASM, 2004). Naturally occurring populations of *E. coli* are of particular public health interest for two reasons. First, because the primary habitat of *E. coli* is the intestinal microflora of warm-blooded animals, elevated levels of *E. coli* in natural waters can indicate the presence of fecal pollution introduced through animal or human waste (Dufour, 1977; USEPA, 1986a). Second, several strains of *E. coli* are toxic to humans and are infectious via routes including the food chain (Mead *et al*., 1999) and contact with polluted recreational water (USEPA, 1986b; 2000a). Monitoring *E. coli* abundance provides limited information concerning the dynamics and structure of *E. coli* populations, which could provide clues about the geographic and host origin of fecal pollution. Furthermore, it has been established that the secondary habitat can influence *E. coli* population composition, survival and transport (LaLiberte and Grimes, 1982; Burton *et al*., 1987; Davies *et al*., 1995; Fish and Pettibone, 1995; Rothmaier *et al*., 1997; Gordon *et al*., 2002). The ability to monitor *E. coli* populations in the secondary habitat can improve both our overall understanding of *E. coli* in the environment as well as strategies to identify bacterial pollution sources.

Two broad categories of methodology to characterize populations of environmental *E. coli* have rapidly evolved in response to interest in microbial source tracking. First, library-dependent methods match phenotypic or genotypic patterns in bacteria occurring in libraries generated from polluted sites with those from known sources. An example of a popular library-dependent fingerprinting method is BOX-PCR, which involves the amplification of a repetitive DNA sequence found throughout the bacterial genome (de Bruijn, 1992). BOX-PCR has been evaluated to assess the composition of primary and secondary habitat *E. coli* populations and to facilitate host-origin classification of *E. coli* isolates (Dombek *et al*., 2000; Albert *et al*., 2003; Carson *et al*., 2003; McClellan *et al*., 2003).

For example, studies that focused on the primary habitat revealed average rates of correct host classification using BOX-PCR of 87% (Seurinck *et al*., 2003), 87.5% (Dombek *et al*., 2000) and 88.1% (Carson *et al*., 2003) among four, seven and eight hosts respectively. Library-dependent methods can require a comprehensive isolate database of potentially several thousand isolates from the sink site as well as each potential source (Wiggins *et al*., 2003), resulting in a significant expenditure of time and money. Even upon the construction of extensive known-source libraries, the genetic diversity of environmental *E. coli* populations is variable (McClellan, 2004), high (McClellan *et al*., 2003; McClellan, 2004) and often accounts for limited classification accuracy (Johnson *et al*., 2004). Given the high diversity of naturally occurring *E. coli* populations, isolate-based fingerprinting methods such as BOX-PCR might prove to be excessively resource-intensive for source tracking applications over broad geographic scales.

While recent reviews have recognized the reliance of most current source tracking techniques on library-dependent methods (Scott *et al*., 2002; Simpson *et al*., 2002), it was shown that library-independent methods outperformed library-based methods in the ability to identify/exclude any of five different fecal sources artificially added to a water sample (Griffith *et al*., 2003). Therefore, methodology that can differentiate the structure of *E. coli* populations in secondary habitats without the prerequisite development of isolate libraries might enhance efforts to understand the spatial and temporal dynamics of *E. coli* populations. Library-independent methods are currently utilized only to provide information concerning the presence or absence of a given source-specific genetic marker. For example, library-independent methodology has been recently developed to detect pathogenic *E. coli* (Watterworth *et al*., 2004) and to discriminate human from animal fecal bacteria (Bernhard *et al*., 2003; Field *et al*., 2003). Denaturing gradient gel electrophoresis (DGGE) can provide highly reproducible fingerprints of complex microbial communities (Díez *et al*., 2001) by separating polymerase chain reaction (PCR)-generated DNA fragments that vary in nucleotide sequence by as little as one in several hundred base-pairs (Fisher and Lehrman, 1983). Denaturing gradient gel electrophoresis has been shown to differentiate laboratory isolates of *E. coli* based on polymorphisms in the β-glucuronidase gene (*uidA*) (Farnleitner *et al*., 2000a,b), which is active in approximately 95% of environmental *E. coli* (Martins *et al*., 1993). *uidA* is considered an appropriate molecular target for *E. coli* detection (Bej *et al*., 1991), and its activity provides the foundation of several commercially available *E. coli* enumeration media (e.g. Colilert, modified m-TEC) (USEPA, 2000b). Recently, Ram and colleagues (2004) reported that nucleotide sequence analysis of *uidA* from 182 human and animal fecal *E. coli* isolates resulted in approximately 75% correct assignment to the primary host, suggesting the utility of *uidA* as a microbial source-tracking target. Despite the demonstrated potential of DGGE and analyses based on *uidA*, a thorough evaluation of the ability of DGGE of *uidA* to differentiate *E. coli* populations in the secondary habitat has yet to be performed.

As environmental *E. coli* populations are maintained by a constant fecal input from a primary habitat (Savageau, 1983), tools that characterize the dynamics of *E. coli* throughout secondary habitats are potentially valuable for ecological and source tracking studies. The overall goal of this study was to evaluate the utility of DGGE of the *uidA* gene as a tool to differentiate environmental *E. coli* populations with the long-term goal of developing a population-based fingerprinting method to match fecal indicators from potential sources with pollution sinks.

Results

Escherichia coli *isolation and identification*

A library of *E. coli* isolates was generated from four environments represented by water and sediments from two different beaches: Lake Erie Beach (LEB) and Inland Lake Beach (ILB). Culturing of the water and sediment samples on modified m-TEC medium resulted in the isolation of a total of 190 presumptive *E. coli* isolates. Physiological and molecular tests confirmed that over 96% ($n = 184$) of the isolates were *E. coli*, which were distributed among LEB water ($n = 53$), LEB sediment ($n = 23$), ILB water ($n = 57$) and ILB sediment ($n = 51$).

Polymerase chain reaction-DGGE of uidA *genes from* E. coli *isolates*

A portion of the *uidA* gene (166 bp) was successfully PCR-amplified from all confirmed *E. coli* isolates. Denaturing gradient gel electrophoresis of the *uidA* fragments resulted in 175 'fingerprints', each containing one band, that were suitable for further analysis. Nine of the fingerprints were either diffuse or contained more than one band and were not further processed. Bands were considered to derive from unique phylotypes based on differential migration as determined by computerized image analysis. Fifteen unique phylotypes were detected and these were distributed among the water and sediment samples (Fig. 1).

Many of the phylotypes detected by DGGE were not unique to a single environment. Nine phylotypes were distributed among LEB water samples while 10 were distributed among each of the remaining three environments (LEB sediment, ILB water and ILB sediment). For example,

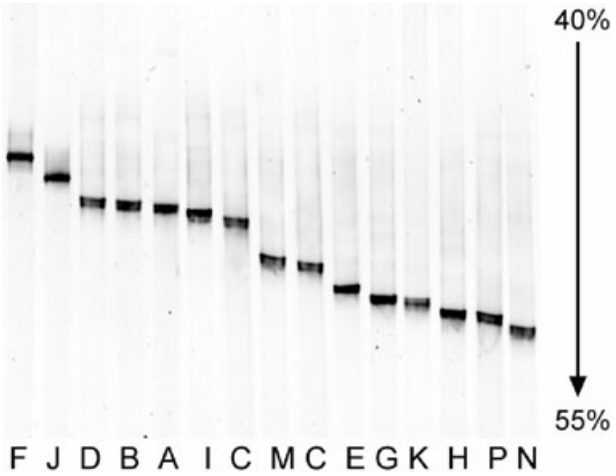

Fig. 1. Parallel DGGE separation of the 15 *E. coli* phylotypes detected among LEB and ILB water and sediment. Labels indicate phylotype designation (A through P) as described in the text and Table 1. The denaturant range is noted by the arrow on the right.

DGGE band type 'A' was detected in all four environments and generated by 73 of the 184 *E. coli* isolates. Eight band types (B, D, E, F, G, I, K, M) were detected in three environments and two types (N, P) were detected in two environments (Table 1). Four other phylotypes (C, H, J and O) were detected in a single environment.

BOX-PCR of E. coli *isolates*

Each of the confirmed *E. coli* isolates produced a BOX-PCR fingerprint useful for image analysis. Isolates exhibiting greater than 90% fingerprint similarity as determined by Pearson's product moment correlation coefficient were identified using cluster analysis of the BOX-PCR fingerprints and classified as identical strains. This analysis resulted in the identification of 21 unique phylotypes. As with the DGGE-detected *E. coli* phylotypes, those detected by BOX-PCR were distributed among all four environments. Specifically, five phylotypes were detected in all four environments, while three, four and nine phylotypes were detected in three, two and one environment(s), respectively (Table 1).

Jackknife analysis of phylotypes established by Pearson's product moment correlation coefficients was used to determine the confidence of BOX-PCR to correctly assign a given phylotype to its environment of origin (LEB water, LEB sediment, ILB water, or ILB sediment). Rates of correct classification were 58.5% (LEB water), 47.8% (LEB sediment), 67.1% (ILB water) and 54.9% (ILB sediment), with an average rate of 57.1%.

Rank–abundance analysis

The numerical distribution (rank–abundance) of *E. coli* phylotypes was used as a measure of the similarity of *E. coli* population structure detected by each method. Rank–abundance plots were compared using ANCOVA analysis and showed that the evenness of *E. coli* phylotypes at each site detected by DGGE was statistically similar to that detected by BOX-PCR (LEB water, $P = 0.60$; LEB sediment, $P = 0.62$; ILB water, $P = 0.41$; and ILB sediment $P = 0.78$) (Fig. 2).

Polymerase chain reaction-DGGE of mixed E. coli *assemblages*

Denaturing gradient gel electrophoresis fingerprints of *E. coli* populations inhabiting LEB and ILB water and sediment were complex and consisted of between 16 (LEB sediment-01) and 31 bands (LEB sediment-02)

Table 1. Distribution of DGGE and BOX-PCR phylotypes in the sampled environments.

	Occurrences in environment				
Phylotype	LEB water	LEB sediment	ILB water	ILB sediment	Total
DGGE					
A	33	8	28	4	73
E	8	0	4	3	15
D	2	3	0	8	13
F	0	2	5	5	12
O	0	0	0	11	11
G	0	2	4	4	10
I	1	0	5	2	8
K	0	1	3	5	8
B	4	2	1	0	7
M	0	1	1	2	4
P	3	0	1	0	4
N	1	0	0	2	3
C	0	2	0	0	2
H	0	1	0	0	1
J	0	1	0	0	1
Total	52	22	52	46	172
BOX-PCR					
18	35	5	26	9	75
15	4	2	8	11	25
17	0	7	2	7	16
10	2	1	5	2	10
1	2	3	2	2	9
8	0	1	0	7	8
12	0	1	0	3	7
14	2	1	4	0	7
7	1	2	1	1	5
3	3	0	0	1	4
13	1	0	0	2	3
2	0	0	2	0	2
4	0	0	2	0	2
11	0	0	1	1	2
16	0	0	2	0	2
19	0	0	0	2	2
5	0	0	0	1	1
6	0	0	0	1	1
9	0	0	1	0	1
20	0	0	0	1	1
21	0	0	1	0	1
Total	53	23	57	51	184

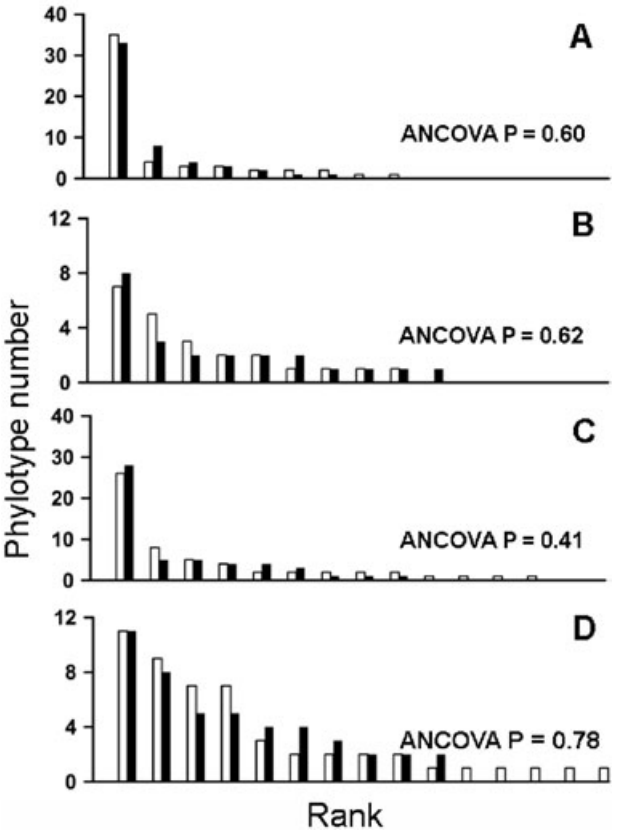

Fig. 2. Rank–abundance distribution of *E. coli* phylotypes as detected by BOX-PCR (open bars) and DGGE (shaded bars). Legend: (A), LEB water; (B), LEB sediment; (C) ILB water; and (D), ILB sediment. *P*-values refer to level of significant difference as determined by ANCOVA.

(Fig. 3). Despite the presence of several phylotypes that appeared to be common to each environment, DGGE fingerprinting successfully detected variation in the *E. coli* populations and differentiated the *E. coli* population structure in all samples. Cluster analysis showed that those *E. coli* populations that originated from a common matrix (water or sediment) or beach (LEB or ILB) were most similar. For example, LEB and ILB water *E. coli* populations were at least 97% and 84% similar among the triplicate samples respectively (Fig. 3). The population structure of sediment *E. coli* exhibited less similarity among triplicate samples than did populations from the water samples (LEB, 28%; ILB, 77%). Two subsequent samplings that occurred approximately 6 and 8 weeks following the described sampling revealed a similar relationship in *E. coli* population structure among the four environments (data not shown).

Jackknife analysis of population fingerprint Pearson's product moment correlation coefficients indicated the average rate of correct classification of each mixed *E. coli* sample into its correct environment was 83.3%. Specifically, all of the population fingerprints generated from LEB water, ILB water and ILB sediment were classified to the correct environment. Of the triplicate fingerprints that characterized the LEB sediment *E. coli* populations, one was correctly classified, while the two incorrectly classified fingerprints were split between LEB and ILB water.

Discussion

Escherichia coli populations in secondary habitats are maintained by the constant input from a primary habitat (Savageau, 1983). Tools that can characterize the dynamics of *E. coli* populations in secondary habitats are

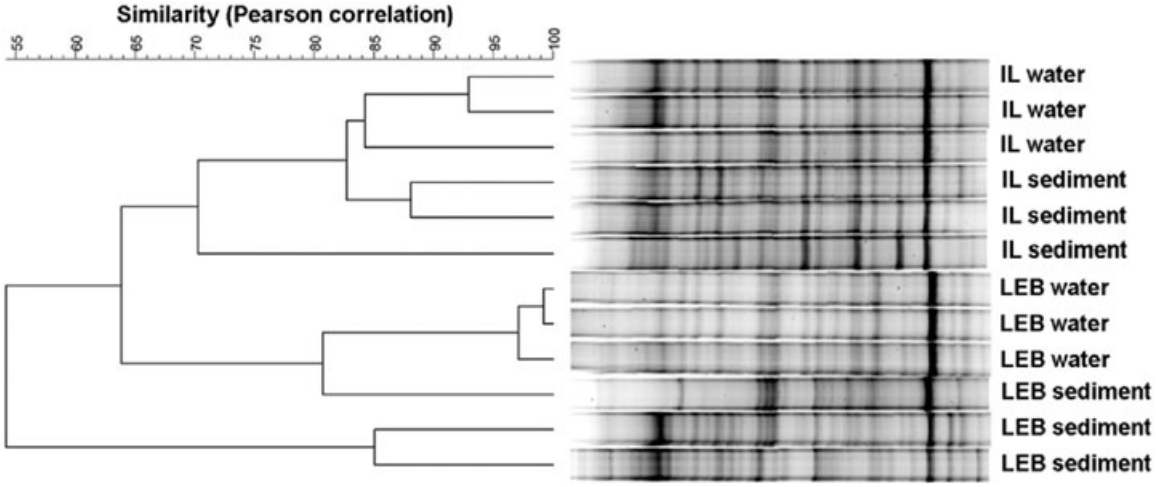

Fig. 3. Relatedness of DGGE fingerprints of *E. coli* populations (*uidA* gene) associated with triplicate LEB and ILB water and sediment membrane filtration/cultures. Comparisons were based on Pearson product moment correlation of fingerprint intensity curves as described in the text.

potentially valuable in assessing changes in population composition, identifying pollution vectors and characterizing *E. coli* transport. Therefore, the primary goal of this study was to determine the ability of DGGE to effectively discriminate populations of mixed *E. coli* phylotypes inhabiting water and sediment.

The resolving capacity of DGGE was first determined on a single isolate basis prior to testing the ability of DGGE to discriminate complex *E. coli* populations. We standardized the resolution of DGGE by comparing it with the resolution afforded by an isolate-based fingerprinting method (BOX-PCR). While it is an uncommon practice to assess the resolution power of a 'population' fingerprinting method (e.g. DGGE) at the isolate level, we felt that the feasibility of DGGE to resolve mixed *E. coli* populations could be best predicted and verified by the ability of DGGE to detect polymorphisms in isolates. Although *uidA* exhibits a moderate degree of sequence polymorphism (Farnleitner *et al*., 2000a; Ram *et al*., 2004) the number of unique *E. coli* phylotypes detected by DGGE was lower than that of BOX-PCR (Table 1), indicating the superior resolution of BOX-PCR at the isolate level. This result was not unexpected, as the *uidA* sequence is likely conserved in order to preserve its functional qualities. This would effectively limit the sequence polymorphism-based separation by DGGE. Furthermore, bands that contain differing DNA sequences but similar melting behaviour can comigrate, resulting in multiple bands assuming the same position in the gel and decreased phylotype resolution (Nübel *et al*., 1996; Casamayor *et al*., 2000).

Rank–abundance plots can provide a valuable metric to detect differences in the structure of microbial communities (Jackson *et al*., 1998; Sigler *et al*., 2002) and were used to ascertain the similarity of numerical distribution of the *E. coli* phylotypes detected by DGGE and BOX-PCR. The statistical similarity of the rank–abundance plots suggested that although the number of phylotypes detected by each method differed (DGGE, 15; BOX-PCR, 21), both methods appeared to present a similar view of the overall *E. coli* population structure (Fig. 2). Therefore, (i) because BOX-PCR is an established method to characterize mixed *E. coli* populations (Dombek *et al*., 2000; Carson *et al*., 2003), and (ii) the rank–abundances expressed following DGGE of *uidA* and BOX-PCR analyses were statistically similar, DGGE of *uidA* was considered effective to differentiate populations of *E. coli* in the secondary habitat.

The high number of DGGE bands shared among the sample fingerprints (Fig. 3) suggested that many phylotypes were common to both beaches. Furthermore, *E. coli* populations inhabiting the water and sediment within each beach also appeared to possess similar phylotype composition (Fig. 3). These results suggest the possibility that LEB and ILB share common fecal waste vectors, especially considering the close proximity of LEB and ILB

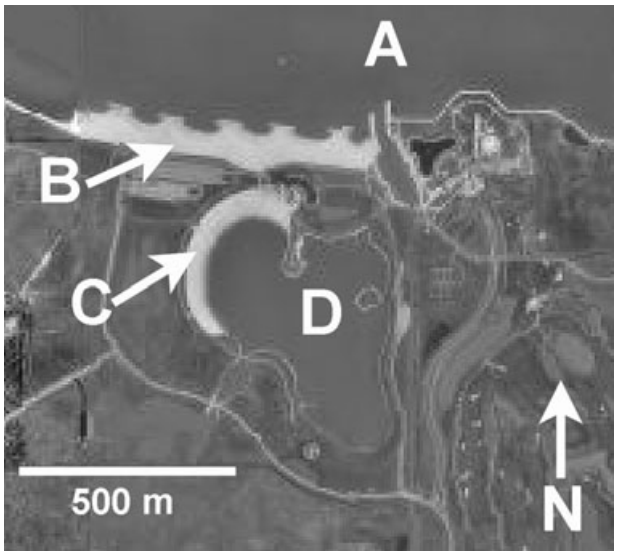

Fig. 4. Arial photo of (A), Lake Erie; (B), Lake Erie Beach; (C), Inland Lake Beach; and (D) the Inland Lake at Maumee Bay State Park, Oregon, OH, USA.

(~150 m, Fig. 4). We further hypothesize that the origin of the bacteria is likely a mobile (animal or human) vector common to both sites because the water supply differs for each beach (LEB is in direct contact with water from Lake Erie while ILB is fed through groundwater sources). Therefore, our future work aims to confirm this hypothesis by using DGGE fingerprinting to further describe the relationship between *E. coli* populations associated with potential vectors such as birds, humans and drainage ditches and populations inhabiting LEB and ILB.

Similarity coefficients showed that triplicate LEB and ILB water samples were 89% and 91% similar, respectively, while sediment samples exhibited greater heterogeneity (LEB, 60% similarity; ILB, 86% similarity). It is logical that wave action and other physical disruptions can result in a greater mixing of bacterial populations in water versus those in underlying sediment (Whittman *et al*., 1999). This result suggests that the use of DGGE in source tracking exercises might be better suited to the water environment, where subtle changes in the population structure might not be masked by site heterogeneity.

Few studies have considered the use of whole-population fingerprinting for *E. coli* ecology studies or source tracking activities. The limited use of DGGE for source tracking efforts likely results from two established shortcomings of DGGE analysis including biases associated with PCR amplifying complex mixtures of DNA (reviewed by Wintzingerode *et al*., 1997; Suzuki and Giovannoni, 1996), and the comigration of bands of different DNA sequence but similar melting behaviour (Nübel *et al*., 1996). Although we recognize these limitations, the intended utility of the method in the current study was to

identify changes in *E. coli* populations more rapidly than can be accomplished with library-dependent techniques. Denaturing gradient gel electrophoresis appeared to accomplish this goal, as it detected a comparable number of unique phylotypes and generated statistically similar rank–abundance plots of *E. coli* populations in comparison with BOX-PCR.

The ASM has recently stated the need for tools and studies that investigate *E. coli* population dynamics in the secondary environment (ASM, 2004). Our results suggest that DGGE of *uidA* can be a useful tool for comparing mixed *E. coli* populations in natural habitats. The current study represents one effort in a extended family of studies in which we are developing the use of DGGE fingerprinting of *E. coli* and other pathogen populations as a screening tool to determine the similarity (or dissimilarity) of bacterial communities in known sinks to potential point and non-point sources. Future research will attempt to utilize this system to qualitatively screen potential sources of fecal pollution for those most likely contributing to the pollution sink. Such analysis might allow for a more efficient identification of sources of bacterial pollution by limiting the significant expenditure of resources often associated with library-dependent methods.

Experimental procedures

Study site and sampling

Water and sediment samples were collected from two beaches located at Maumee Bay State Park in Oregon, Ohio, USA (Fig. 4). Lake Erie Beach is located on the southern shore of Lake Erie and is approximately 750 m in length. The second sampling site was the Inland Lake Beach also located at Maumee Bay State Park. The Inland Lake Beach is located on a 40-ha, ground-water-fed lake located approximately 150 m inland (south) from Lake Erie Beach. From each of three locations at each beach, triplicate subsamples were collected and mixed. Water samples were collected in approximately 50 cm of water depth by inverting sterile 1-l bottles approximately 30 cm below the water surface according to standard protocols (Myers and Wilde, 2003). Sediments were collected simultaneously by scraping sterile screw-top jars across the upper 3–5 cm of lake bottom sediment. Water samples were always harvested before sediments to avoid the possibility of cosampling disturbed, suspended sediments during water sampling. All samples were immediately transported to the laboratory and analysed within 2 h of collection.

Escherichia coli *isolation and confirmation*

Escherichia coli were cultured from water samples according to the modified *E. coli* method (USEPA, 2000b). Sediment samples were also processed with the USEPA method, but with slight modifications. Briefly, 5 g of sediment were suspended by hand-shaking (50 shakes) in 50 ml of 10 mM sodium phosphate buffer (pH 7.0) and allowed to settle for 10 min. The supernatant was serially diluted 10-fold to 1:100, and then 20 ml of the 1:10 and 1:100 dilutions were filtered (0.45 μm) and incubated as previously described (USEPA, 2000b). Following incubation, presumptive *E. coli* colonies from water and sediment samples (as noted by magenta coloration of the colony on modified m-TEC media) were transferred twice onto Luria–Bertani agar to generate pure isolates. The isolates were then streaked onto eosin methylene blue (EMB) agar and incubated overnight at 37°C. A representative colony of each isolate that presented a metallic green sheen on EMB agar was resuspended in 50 μl of water. One μl of this suspension was subjected to PCR using primers specific for the 16S rRNA gene of *E. coli* (Sabat *et al*., 2000). Isolates that were magenta in colour on modified m-TEC, displayed a metallic green sheen on EMB agar, and resulted in a PCR product of proper size (~0.544 kb) following amplification using *E. coli*-specific PCR primers were designated as *E. coli* isolates and used for subsequent studies.

Polymerase chain reaction-DGGE of E. coli *isolate* uidA *genes*

Polymerase chain reaction of *uidA* genes (0.166 kb) was performed with primers UAL-1939 and UAR-2105 and 1 μl of the cell suspension described above according to the method of Bej *et al*., 1991). To facilitate PCR product separation in subsequent DGGE, a 40-bp GC-clamp (Muyzer *et al*., 1993) was attached to the 5′ end of primer UAL-1939. Denaturing gradient gel electrophoresis of the PCR product from each isolate was performed with a DCODE Universal Mutation Detection System (Bio-Rad) in 8% polyacrylamide gels supplemented with 2% (v/v) glycerol. The denaturant concentration ranged from 40 to 55% and electrophoresis was performed for 1000 Vh (5 h at 200 V). Gels were stained for 20 min with a 1:10 000 dilution of GelStar nucleic acid stain and visualized with a Kodak Gel Logic 200 image analysis system.

BOX-PCR

BOX-PCR was performed according to the method of Rademaker and colleagues (1998) using the BOX A1R primer and 1 μl of the cell suspension described above.

Following the initial DGGE and BOX-PCR fingerprinting attempts, isolates that either failed to produce a fingerprint, or produced a fingerprint of limited quality were reanalysed a maximum of three times to generated a suitable fingerprint.

DNA isolation and DGGE fingerprinting of mixed E. coli *populations*

DNA was isolated directly from the same membrane filters described in the section '*E. coli* isolation and confirmation' above. In an effort to maximize the likelihood of sampling a phylotype distribution representative of the water and sediment populations, only membranes that exhibited dense growth of magenta colonies (too numerous to count) were selected. Membranes were cut into pieces with sterile scis-

sors and placed into a 50-ml centrifuge tube with 10 ml of 10 mM sodium phosphate buffer (pH 7.0). Tubes were shaken for 30 s on a vortex mixer, and then centrifuged for five min at 13 000 *g* to pellet the dislodged cells. The supernatant and membrane pieces were removed, the cells were resuspended in 1 ml of DNA extraction buffer by pipetting, transferred to a 2-ml microcentrifuge tube containing 0.3 ml of 0.1 mm diameter glass beads, and extracted for DNA as previously described (Sigler and Zeyer, 2002). Polymerase chain reaction-DGGE and fingerprint analysis of *uidA* genes from water and sediment *E. coli* populations was performed as described above.

Denaturing gradient gel electrophoresis and BOX-PCR fingerprint analysis

All DGGE and BOX-PCR fingerprint patterns were analysed with GelCompar II software (version 3.5, Applied Maths). For both fingerprinting methods, markers were loaded such that a maximum of five samples separated each marker lane. Denaturing gradient gel electrophoresis fingerprints were normalized using a custom marker synthesized from seven *uidA* gene fragments that upon electrophoresis were distributed along the length of the gel. BOX-PCR fingerprints were normalized using an external DNA marker (1 kb Ladder, Promega). The similarity of DGGE fingerprints of isolate *uidA* genes as well as those generated from the mixed, environmental populations was calculated using the band-based Dice coefficient (Dice, 1945) with 3% optimization and 1% band position tolerance settings in the GelCompar II software.

Because *uidA* is found in a single copy in the *E. coli* genome, DGGE of *uidA* PCR products generated from *E. coli* isolates produced a single band. Furthermore, it was previously established that the above DGGE parameters were capable of detecting band mobility differences in *uidA* fragments differing by as little as one base pair (Farnleitner *et al*., 2000a). Therefore, when analysing the DGGE fingerprints of *E. coli* isolates, we concluded that any two bands that did not comigrate (i.e. $<$ 100% similarity) represented differing nucleotide sequences and originated from unique *E. coli* phylotypes.

BOX-PCR fingerprint similarities were determined by calculating the Pearson's product moment correlation coefficient (Jobson, 1991). A 3% optimization was used in the GelCompar II software to accounting for electrophoretic shifts between any two identical bands/patterns to provide the most appropriate fingerprint matching. Because of the difficulty involved in resolving low molecular weight bands in the BOX-PCR gels, only bands that were larger than approximately 300 bp were considered in the analysis. To our knowledge, no study has confirmed an appropriate, discrete BOX-PCR fingerprint similarity threshold for defining bacterial isolates as unique. Although many authors fail to report a threshold similarity value for BOX-PCR fingerprints of *E. coli* isolates, the majority of reported threshold similarity values range from 80% to 90% (Myoda *et al*., 2003; Johnson *et al*., 2004; McClellan, 2004). In the current study, *E. coli* isolates were considered to be identical if they exhibited BOX-PCR fingerprints that were at least 90% similar.

After archiving the DGGE and BOX-PCR fingerprints, each fingerprint was manually assigned to it's environment of origin using GelCompar II software. Discriminant analysis determined how accurately the BOX-PCR fingerprint similarity coefficients could predict the environment from which isolate was obtained. This was accomplished using the Jackknife method, in which each fingerprint was removed from the database and individually compared with all remaining fingerprints. The percentage of instances in which the query fingerprint matched the environment from which it originated was a measure of (i) the stability of that environmental grouping, and (ii) the rate of correct classification.

Comparison of the apparent population structure detected by DGGE and BOX-PCR

Rank–abundance plots compared the numerical distribution of unique *E. coli* phylotypes detected by both DGGE and BOX-PCR. The resulting plots were compared for statistically significant differences within each environment using analysis of covariance (ANCOVA) following log-transformation of the abundance data.

References

Albert, J.M., Munakata-Marr, J., and Siegrist, R.L. (2003) Statistical evaluation of bacterial source tracking data obtained by rep-PCR DNA fingerprinting of *Escherichia coli*. *Environ Sci Technol* **37:** 4554–4560.

American Society for Microbiology (ASM) (2004) *Response to: Water quality criteria for bacteria for coastal and Great Lakes waters.* [WWW document]. URL http://www.asm.org/Policy/index.asp?bid=29781.

Bej, A.K., Dicesare, J.L., Haff, L., and 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:** 1013–1017.

Bernhard, A.E., Goyard, T., Simonich, M.T., and Field, K.G. (2003) Application of a rapid method for identifying fecal pollution sources in a multi-use estuary. *Wat Res* **37:** 909–913.

de Bruijn, F. (1992) Use of repetitive (repetitive extragenic palindromic and erterobacterial intergenic consensus) sequences and the polymerase chain reaction to fingerprint the genomes of *Rhizobium meliloti* isolates and other soil bacteria. *Appl Environ Microbiol* **58:** 2180–2187.

Burton, G.A., Gunnison, D., and Lanza, G.R. (1987) Survival of pathogenic bacteria in various freshwater sediments. *Appl Environ Microbiol* **53:** 633–638.

Carson, C.A., Shear, B.L., Ellersiek, M.R., and Schnell, J.D. (2003) Comparison of ribotyping and extragenic palindromic-PCR for identification of fecal *Escherichia coli* from humans and animals. *Appl Environ Microbiol* **69:** 1836–1839.

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

Davies, C.M., Long, J.A.H., Donald, M., and Ashbolt, N.J. (1995) Survival of fecal microorganisms in marine and

freshwater sediments. *Appl Environ Microbiol* **61:** 1888–1896.

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

Díez, B., Pedrós-Alió, C., Marsh, T.L., and Massana, R. (2001) Application of denaturing gradient gel electrophoresis (DGGE) to study the diversity of marine picoeukaryotic assemblages and comparison of DGGE with other molecular techniques. *Appl Environ Microbiol* **67:** 2942–2951.

Dombek, P.E., Johnson, L.K., Zimmerley, S.T., and Sadowsky, M.J. (2000) Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. *Appl Environ Microbiol* **66:** 2572–2577.

Dufour, P. (1977) *Escherichia coli*: the faecal coliform. In *Bacterial Indicators / Health Hazards Associated With Waler*. ASTM STP 635. Hoadley, A. and Dutka, B.J. (eds). Conshohocken, PA, USA: American Society for Testing and Materials, pp. 48–58.

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

Farnleitner, A.H., Kreuzinger, N., Kavka, G.G., Grillenberger, S., Rath, J., and Mach, R.L. (2000b) 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:** 427–431.

Field, K.G., Chern, E.C., Dick, L.K., Fuhrman, J., Griffith, J., Holden, P.A., *et al.* (2003) A comparative study of culture-independent, library-independent genotypic methods of fecal source tracking. *J Water Health* **1:** 181–194.

Fish, J.T., and Pettibone, G.W. (1995) Influence of freshwater sediment on the survival of *Escherichia coli* and *Salmonella* sp. as measured by three methods of enumeration. *Lett Appl Microbiol* **20:** 277–281.

Fisher, S.G., and Lehrman, L.S. (1983) DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: Correspondence with melting theory. *Proc Nat Acad Sci* **80:** 1579–1583.

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

Griffith, J.F., Weisberg, S.B., and McGee, C.D. (2003) Evaluation of microbial source tracking methods using mixed fecal sources in aqueous test samples. *J Water Health* **1:** 141–151.

Jackson, C.R., Roden, E.E., and Churchill, P.F. (1998) Changes in bacterial species composition in enrichment cultures with various dilutions of inoculum as monitored by denaturing gradient gel electrophoresis. *Appl Environ Microbiol* **64:** 5046–5048.

Jobson, J.D. (1991) *Applied Multivariate Data Analysis*, Vol. 1. New York, USA: Springer Verlag.

Johnson, L.K., Brown, M.B., Carruthers, E.A., Ferguson, J.A., Dombek, P.E., and 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:** 4478–4485.

LaLiberte, P., and Grimes, D.J. (1982) Survival of *Escherichia coli* in lake bottom sediment. *Appl Environ Microbiol* **43:** 1089–1098.

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

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

Martins, M.T., Rivera, I.G., Clark, D.L., Stewart, M.H., Wolfe, R.L., and Olson, B.H. (1993) Distribution of *uidA* gene sequences in *Escherichia coli* isolates in water sources and comparison with the expression of beta-glucuronidase activity in 4-methylumbelliferyl-beta-D-glucuronide media. *Appl Environ Microbiol* **59:** 2271–2276.

Mead, P.S., Slutsker, L., Dietz, V., Mccaig, L.F., Breese, J.S., Shapiro, C., *et al.* (1999) Food related illness and death in the United States. *Emerg Infect Dis* **5:** 607–625.

Muyzer, G., De Waal, E.C., and 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:** 695–700.

Myers, D.N., and Wilde (2003) *Biological Indicators (3rd edn): U.S. Geological Survey Techniques of Water-Resources Investigations*, Book 9, Chapter A7. [WWW document]. URL http://pubs.water.usgs.gov/twri9A7/.

Myoda, S.P.A.C., Carson, Fuhrmann, J.J., Hahm, B., Hartel, P.G., Yampara-Iquise, H., *et al.* (2003) Comparison of genotypic-based microbial source tracking methods requiring a host origin database. *J Wat Health* **1:** 167–180.

Nübel, U., Engelen, B., Felske, A., Snaidr, J., Wieshuber, A., Amann, R., *et al.* (1996) Sequence heterogeneities of genes encoding 16S rRNAs in *Paenibacillus polymyxa* detected by temperature gradient gel electrophoresis. *J Bacteriol* **178:** 5636–5643.

Rademaker, J.L.W.F.J., Louws and de Brujin (1998) Characterization of the diversity of ecologically important microbes by rep-PCR genomic fingerprinting. In *Molecular Microbial Ecology Manual*, Vol. 3.4.3. Akkermans, A.D.L., Van Elsas, J.D. and De Bruijn, F. (eds). Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 1–27.

Ram, J.L., Ritchie, R.P., Fang, J., Gonzalez, F.S., and Selegean, J.P. (2004) Sequence-based source tracking of *Escherichia coli* based on genetic diversity of β-glucuronidase. *J Environ Qual* **33:** 1024–1032.

Rothmaier, R., Weidenmann, A., and Botzenhart, K. (1997) Transport of *Escherichia coli* through soil to groundwater traced by randomly amplified polymorphic DNA (RAPD). *Wat Sci Techn* **35:** 351–357.

Sabat, G., Rose, P., Hickey, W.J., and Harkin, J.M. (2000) Selective and sensitive method for PCR amplification of *Escherichia coli* 16S rRNA genes in soil. *Appl Environ Microbiol* **66:** 844–849.

Savageau, M.A. (1983) *Escherichia coli* habitats, cell types, and molecular mechanisms of gene control. *Am Nat* **122:** 732–744.

Scott, T.M., Rose, J.B., Jenkins, T.M., Farrah, S.R., and Lukasik, J. (2002) Microbial source tracking: current methodology and future directions. *Appl Environ Microbiol* **68:** 5796–5803.

Seurinck, S., Verstraete, W., and Siciliano, S.D. (2003) Use of 16S–23S rRNA intergenic spacer region PCR and repetitive extragenic palindromic PCR analyses of *Escherichia coli* isolates to identify nonpoint fecal sources. *Appl Environ Microbiol* **69:** 4942–4950.

Sigler, W.V., and Zeyer, J. (2002) Microbial diversity and activity along the forefields of two receding glaciers. *Microb Ecol* **43:** 397–407.

Sigler, W.V., Crivii, S., and Zeyer, J. (2002) Bacterial succession in glacial forefield soils characterized by community structure, activity and opportunistic growth dynamics. *Microb Ecol* **44:** 306–316.

Simpson, J.M., Santo Domingo, J.W., and Reasoner, D.J. (2002) Microbial source tracking; state of the science. *Environ Sci Tecnol* **36:** 5279–5288.

Suzuki, M.T., and Giovannoni, S.J. (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl Environ Microbiol* **62:** 625–630.

United States Environmental Protection Agency (USEPA) (1986a) Ambient water quality criteria for bacteria-1986, Washington, D.C., Office of Water, EPA-823-F-02–008.

United States Environmental Protection Agency (USEPA) (1986b) Bacteriological ambient water quality criteria; availability. *Federal Register* **51:** 8012–8016.

United States Environmental Protection Agency (USEPA) (2000a) *National Water Quality Inventory: 1998 Report to Congress.* EPA-841-R-00–001. Washington DC, USA: Office of Water, U.S. Environmental Protection Agency.

United States Environmental Protection Agency (USEPA) (2000b) *Improved enumeration methods for the recreational water quality indicators* – Enterococci *and* Escherichia coli. Washington DC, EPA/821/R-97/004 March 2000.

United States Environmental Protection Agency (USEPA) (2004) Water quality standards for coastal and Great Lakes recreational waters; proposed rule. *Federal Register* **69:** 41720–41743.

Watterworth, L., Topp, E., Schraft, H., and Leung, K.T. (2004) Multiplex PCR-DNA probe assay for the detection of pathogenic Escherichia coli. *J Microbiol Meth* **60:** 93–105.

Whittman, R.L., Nevers, M.B., and Gerovac, P.J. (1999) Interaction of ambient conditions and fecal coliform bacteria in southern Lake Michigan beach waters: monitoring program implications. *Nat Areas J* **19** (166)**:** 171.

Wiggins, B.A., Cash, P.W., Creamer, W.S., Dart, S.E., Garcia, P.P., Gerecke, T.M., *et al.* (2003) Use of antibiotic resistance analysis for representativeness testing of multiwatershed libraries. *Appl Environ Microbiol* **69:** 3399–3405.

Wintzingerode, F.V., Goebel, U.B., and Stackebrandt, E. (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* **21:** 213–229.