
Large scale analysis of virulence genes in *Escherichia coli* strains isolated from Avalon Bay, CA

Matthew J. Hamilton^{1,2}, Asbah Z. Hadi^{1,3}, John F. Griffith and Michael J. Sadowsky^{2,3}

ABSTRACT

Contamination of recreational waters with *Escherichia coli* and *Enterococcus* species is a widespread problem resulting in beach closures and loss of recreational activity. While *E. coli* is frequently used as an indicator of fecal contamination, and has been extensively measured in waterways, few studies have examined the presence of potentially pathogenic *E. coli* strains in beach waters. In this study, a combination of high-throughput, robot-assisted colony hybridization and analyses based on polymerase chain reaction (PCR) was used to determine the genomic composition and frequency of virulence genes present in *E. coli* isolated from beach water at Avalon Bay, Santa Catalina Island, CA. A total of 24,493 *E. coli* isolates were collected from two sites at a popular swimming beach August through September 2007 and July through August 2008. All isolates were examined for the presence of shiga-like toxins (*stx1/stx2*), intimin (*eaeA*), and enterotoxins (*ST/LT*). Of the 24,493 isolates examined, 3.6% contained the *eaeA* gene, indicating that these isolates were potential enteropathogenic *E. coli* (EPEC) strains. However, after 5 days, more than 10% of the strains were potentially EPEC, suggesting a strong temporal component associated with incidence of virulence genes. No shiga toxin producing *E. coli* (STEC) or enterotoxigenic *E. coli* (ETEC) isolates were detected, and only eight (<1.0%) of the potential EPEC isolates were found to carry the EPEC adherence factor (*EAF*) plasmid. The potential EPEC isolates primarily belonged to *E. coli* phylogenetic groups B1 or B2, and carried the beta intimin subtype. Deoxyribonucleic acid (DNA) fin-

gerprint analyses of the potential EPEC strains indicate that the isolates belong to several genetically diverse groups, although clonal isolates were frequently detected. While the presence of virulence genes alone cannot be used to determine the pathogenicity of strains, results from this study show that potential EPEC strains can be found in marine beach water, and their presence needs to be considered as one of the factors used in decisions concerning beach closures.

INTRODUCTION

Contamination of recreational waters with *E. coli* and *Enterococcus* species is a common problem resulting in beach closures and loss of recreational activity. Because these bacteria are most frequently found in the intestinal tract of warm-blooded animals and are shed with feces, *E. coli* and *Enterococcus* have traditionally been used as indicators of fecal contamination. While it is generally accepted that the presence of indicator organisms in recreational waters suggests the potential presence of fecal pathogens (Cabelli *et al.* 1982, Dufour 1984, Prüss 1998), several studies have shown that these bacteria can persist in the environment without a host and can become naturalized to beach sand, soil, and algae (Hardina and Fujioka 1991; Davies *et al.* 1995; Byappanahalli and Fujioka 1998; Desmarais *et al.* 2002; Byappanahalli *et al.* 2006; Ishii *et al.* 2006a,b; Ishii *et al.* 2007a; Yan and Sadowsky 2007). Traditionally, *E. coli* has been often characterized as a harmless, commensal bacterium. However, some strains have been shown to be capable of causing human disease (Nataro and Kaper 1998, Kaper *et al.* 2004), and fecal material from some animals con-

¹ University of Minnesota, Department of Microbiology, Minneapolis, MN

² University of Minnesota, BioTechnology Institute, St. Paul, MN

³ University of Minnesota, Department of Soil, Water, and Climate, St. Paul, MN

tains high frequencies of *E. coli* strains harboring virulence genes (Ishii *et al.* 2007b). Despite increasing evidence that *E. coli* strains from several animal hosts contain virulence genes, and some have been shown to cause serious or fatal diseases in humans (Nataro and Kaper 1998), few studies have determined whether *E. coli* strains isolated from marine recreational waters contain virulence genes and are potentially pathogenic (Lang *et al.* 1994).

Pathogenic *E. coli* generally cause diarrhea and other gastrointestinal disease (Nataro and Kaper 1998, Kaper *et al.* 2004), although some strains have been found to cause extraintestinal infections (Dietzman *et al.* 1974, Manges *et al.* 2001). Diarrheagenic strains are classified into several groups based on the mechanisms of pathogenesis and the presence of various virulence factors or determinants. These groups include diffusely adhering (DAEC), enteroaggregative (EAEC), enteroinvasive (EIEC), EPEC, ETEC, and STEC. Detailed information about pathogenic *E. coli* strains can be found in review articles by Nataro and Kaper (1998) and Kaper *et al.* (2004). The STEC are defined as *E. coli* strains expressing either *stx1/stx2*, or other toxins sharing significant homology to the shiga toxin gene originally identified in *Shigella dysenteriae* (Paton and Paton 1998, Kaper *et al.* 2004). The common reservoirs for STEC strains are ruminants and swine (Djordjevic *et al.* 2004, Fratamico *et al.* 2004, Gyles 2007, Ishii *et al.* 2007b). In contrast, ETEC strains are defined by the presence of at least one of the heat stable (*ST α*) and the heat labile (*LT-1*) enterotoxin genes (Levine 1987). These strains are often associated with infantile diarrhea, particularly in the developing world where the disease is considered endemic (Black 1990, Nataro and Kaper 1998, Kaper *et al.* 2004). Enterotoxigenic *E. coli* strains are also a common cause of traveler's diarrhea, where visitors from the developed world may lack immunity to these strains (Black 1990).

Enteropathogenic *E. coli* strains are a common cause of human diarrheal diseases in developing countries, particularly among children less than two years of age (Levine and Edelman 1984, Trabulsi *et al.* 2002). These strains have been isolated from many animal host species, including: humans, cats, cows, dogs, deer, ducks, geese, and horses (Ishii *et al.* 2007b). In addition, these strains are defined by the presence of a locus of enterocyte effacement (LEE) pathogenicity island, which encodes several virulence factors including *eaeA*, and the absence of

stx1/stx2 (Nataro and Kaper 1998, Kaper *et al.* 2004). Strains carrying the *EAF* plasmid are referred to as typical EPEC, while those strains without the plasmid are referred to as atypical EPEC.

Most *E. coli* strains have been assigned to one of six major phylogenetic groups (A, B1, B2, C, D, and E) based on their evolutionary origins (Clermont *et al.* 2000, Escobar-Páramo *et al.* 2004). The association of STEC strains with certain phylogenetic groups has been tentatively established. Escobar-Páramo *et al.* (2004) determined the majority of STEC strains they examined from clinical samples belonged to phylogenetic group B1. However, while a tentative link has been established between intimin subtype and phylogenetic group (Reid *et al.* 2000, Escobar-Páramo *et al.* 2004), not all EPEC strains have a specific genetic background, and they may be distributed among several phylogenetic groups (Escobar-Páramo *et al.* 2004, Ishii *et al.* 2007b).

Detection of EPEC and STEC strains has been previously accomplished with multiplex PCR, using primers specific to *eaeA* and *stx1/stx2*, respectively (Paton and Paton 1998, Ahmed *et al.* 2007, Ishii *et al.* 2007b, Ram *et al.* 2007). While these analyses are useful for screening relatively small numbers of isolates, they suffer from the high costs and low throughput associated with PCR.

We previously reported the successful use of high throughput, semi-automated, robotic technology to detect the presence of host-source specific genes in environmental *E. coli* isolates (Yan *et al.* 2007). In the study presented here, we adapted this technology to quickly screen large numbers of *E. coli* isolates from marine recreational water for the presence of virulence determinants, and we confirmed positive hybridization results by using PCR. The objectives of this present study were to: 1) examine the distribution and frequency of potentially pathogenic *E. coli* strains isolated from beach water at a popular swimming beach in Avalon, CA; 2) characterize the virulence gene determinants; and 3) determine the genetic relatedness of the potentially pathogenic strains by using horizontal fluorophore-enhanced rep-PCR (HFERP) DNA fingerprint analyses.

METHODS

Water samples were obtained in 2007 and 2008, at either 8:00 a.m. or 12:00 p.m., from two beach sites at Avalon Bay, Santa Catalina Island, CA, as previously described (Bordner *et al.* 1978). The two

sampling locations are identified as sites B and C in Figure 1. Water samples (4 L) were collected at each time point. The temperature on most sampling days was approximately 29°C, and only a few swimmers were in the water during the 8 a.m. sampling time. Bacteria were concentrated by membrane filtration using 142 mm (dia.), 0.45 µm Supor hydrophilic polyethersulfone membranes (Pall Corporation, East Hills, NY). After filtration, membranes were cut into 4 equal sections, and each section was placed in a 50 ml conical tube containing 10 ml of sterile phosphate buffered saline (20 mM sodium phosphate, 15 mM sodium chloride, 7.2 pH) amended with 0.1% hydrolyzed gelatin and 10 g of sterile 3 mm (dia.) glass beads. Bacteria were removed from the filter surface by gentle agitation for 30 minutes using a wrist action shaker (Burrell Scientific, Pittsburg, PA); filters were removed from solutions, and 50% glycerol was added to obtain a final concentration of 10%. All samples were stored frozen at -80°C until used.

The *E. coli* reference strains Pig206 (*stx1/2+*, *eaeA-*) and Deer090 (*stx1/2-*, *eaeA+*) were used as positive and negative controls for colony hybridization with the *stx1*, *stx2*, and *eaeA* probes (Dombek *et al.* 2000, Johnson *et al.* 2004). The ETEC strain 1362 and Pig206 were used as positive and negative controls, respectively, for hybridizations with the enterotoxin gene probes. The *E. coli* strain O157:H7 (ATCC 43895) was used as the positive control for PCR-based assays for virulence genes and for ampli-

fying DNA used as hybridization probes for the *eaeA*, *stx1*, and *stx2* genes. The ETEC strain 1362 was used as template for amplifying DNA for use as hybridization probes for the *STα* and *LT-I* genes. The *E. coli* strain H120 was used a positive control for PCR reactions to detect the *EAF* plasmid (Dombek *et al.* 2000, Johnson *et al.* 2004, Ishii *et al.* 2007b), and *E. coli* strain Pig294 was used as control for HFERP DNA fingerprint analysis (Dombek *et al.* 2000, Johnson *et al.* 2004).

E. coli strains were isolated from filter washings as previously described (Yan *et al.* 2007). One to three ml aliquots of filter washings were spread-plated onto the surface of modified membrane thermo-tolerant *E. coli* (mTEC) agar medium in 22 x 22 cm Q-tray bioassay plates (Genetix Boston, MA). Modified mTEC was prepared as described, except that 500 µg of 5-bromo-4-chloro-3-indolyl-β-d-glucuronic acid (X-Gluc) per ml was used as the chromogenic indicator (USEPA 2002, Yan and Sadowsky 2007). Plates were incubated at 35°C for 2 hours, and then at 44.5°C for 22 hours. After incubation, plates were stored at 4°C overnight to facilitate development of blue pigment in colonies, and to differentiate *E. coli* from other coliform and gram-negative bacteria. Well-isolated blue colonies were picked by hand or by using a Q-Bot robotic system (Genetix, Boston, MA) into 384 well microplates containing Hogness modified freezing medium as previously described (Yan *et al.* 2007).



Figure 1. Sample sites at Avalon Bay, Santa Catalina Island, CA.

Microplates were incubated at 37°C overnight, and stored at -80°C until used. A total of 24,493 individual *E. coli* isolates were obtained in this study. A random sample of 1,024 strains were confirmed as *E. coli* by using a series of microbiological and biochemical tests, including: the indole and methyl red tests, the inability of isolates to grow on citrate, β -d-glucuronidase activity using EC-MUG broth (Difco), and color reaction on ChromAgar and MacConkey agar (Dombek *et al.* 2000, Johnson *et al.* 2004).

Automated arraying of *E. coli* isolates onto positively charged 22 x 22 cm Performa II nylon membranes (Genetix) was done using a QBot robot system (Genetix) as previously described (Yan *et al.* 2007). Each membrane was divided into 6 sections; each section contained 384 subunits, and each subunit consisted of 4 individual spots. One spot per subunit contained either a positive or negative control strain as described above. Using this format, each membrane was arrayed with a maximum of 6,912 *E. coli* isolates obtained from water samples. Arrayed membranes were placed on the surface of LB agar plates, incubated at 37°C for 8 to 10 hours, and stored at 4°C overnight.

Colony hybridizations were performed as previously described (Hamilton *et al.* 2006, Yan *et al.* 2007) to determine the reactivity of the *E. coli* isolates to DNA probes for the *stx1*, *stx2*, and *eaeA* genes. The DNA probes for the *stx1*, *stx2*, and *eaeA* genes were obtained from control strains using the PCR and primer pairs *stx1F* and *stx1R*, *stx2F* and *stx2R*, and *eaeAF* and *eaeAR*, respectively (Paton and Paton 1998). Probe DNA for *ST α* and *LT-I* was amplified using primer pairs JW14/JW7 and LTPr1/JW11, respectively (Stacy-Phipps *et al.* 1995). Probes were labeled with [32P]dCTP using the Random Primer DNA Labeling system (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol. Probes for *stx1/stx2* and ETEC were pooled before labeling. Membranes were hybridized overnight at 68°C and washed under high stringency at 68°C in 0.1 x SSC (1 x SSC is 0.15 M NaCl and 0.015 M sodium citrate) containing 0.1% sodium dodecyl sulfate (Yan *et al.* 2007). Membranes were air dried after washing, wrapped in plastic film, and exposed overnight to storage phosphor imaging screens (GE Healthcare, Chalfont St. Giles, UK). All colony hybridizations were done in triplicate. Hybridization images were captured using a STORM 840 densitometer (GE Healthcare), and quantitative

image analysis used ScanAlyze version 2.51 software (http://rana.lbl.gov/downloads/scanalyzer/scanalyzer2_versions_2_51.exe). Positive and negative hybridization signals were determined as previously described (Yan *et al.* 2007).

The presence of virulence genes in strains showing positive hybridization reactions, on any of the triplicate membranes, was confirmed using a multiplex PCR approach using primer pairs *stx1F/stx1R*, *stx2F/stx2R*, *eaeAF/eaeAR* (Paton and Paton 1998), and primer pairs JW14/JW7 and LTPr1/JW11 (Stacy-Phipps *et al.* 1995). Template DNA was extracted from cells grown overnight in LB medium as previously described (Ishii *et al.* 2006a), diluted tenfold in distilled H₂O, and 1 to 2 μ l of lysate was added to each PCR reaction. Isolates shown to carry the *eaeA* gene were also tested for the presence of the *EAF* virulence plasmid using primer pairs *EAF1* and *EAF25* (Franke *et al.* 1994).

Subtype analysis of the *eaeA* gene was done using a PCR-RFLP technique and primers EaeVF, EaeR, EaeZetaVR, and EaeIotaVR, as previously described (Ramachandran *et al.* 2003). Amplified products were digested with restriction enzymes *AluI*, *CfoI*, or *RsaI* to distinguish among 14 *eae* subtypes (Ramachandran *et al.* 2003). DNA fragments were separated by electrophoresis at 90 V for 2 to 3 hours using 2% SeaKem LE agarose gels (FMC, Rockland, ME). Gels were stained with 0.5 μ g ethidium bromide per ml and visualized with UV light.

E. coli strains were classified into one of four phylogenetic groups (A, B1, B2, and D/E) using a multiplex PCR protocol as previously described (Clermont *et al.* 2000). Strains with the Clermont (+, -, +) genotype (phylogenetic group D) were tested for the presence of *ibeA* to determine if these strains actually belonged to phylogenetic group B2 (Gordon *et al.* 2008). HFERP DNA fingerprint analyses were done using the BOXA1R primer as previously described (Johnson *et al.* 2004).

G-tests (likelihood-ratio or maximum likelihood statistical significance tests) were used to determine if *E. coli* carrying virulence factor genes were evenly distributed among samples and across sites. Statistical analysis of DNA fingerprint data was done using Bionumerics software (version 2.1; Applied Math, Kortrijk, Belgium). Dendrograms were constructed using the curve-based, Pearson's product-moment correlation coefficient and clustering was done using the unweighted pair group method with

arithmetic means (UPGMA; Johnson *et al.* 2004). Multivariate analysis of variance (MANOVA) of HFERP DNA fingerprint data was used to cluster *E. coli* strains isolated from different sites and in different years (Dombek *et al.* 2000, Byappanahalli *et al.* 2006, Ishii *et al.* 2006a).

RESULTS

A total of 12,000 and 12,493 well-isolated individual *E. coli* colonies were collected from Avalon Bay sites B and C (Figure 1) at Santa Catalina Island, CA in 2007 and 2008, respectively. The frequency of obtaining *E. coli* from water samples varied considerably by date, and the number of *E. coli* strains recovered from Avalon Bay was greater at 8 a.m. than at 12 noon. Biochemical analyses indicated that 99.8% of the isolates were confirmed to be *E. coli* as determined by methyl red and indole tests, color on ChromAgar and MacConkey agar, citrate utilization reaction, and β -d-glucuronidase activity (data not shown). These results are consistent with other reports concerning the ability of mTEC medium to selectively isolate *E. coli* from water samples (USEPA 2002, Yan *et al.* 2007).

Hybridization analyses were used to screen environmental *E. coli* isolates for the presence of virulence genes. Of the 24,493 isolates examined, 875 (3.6%) were positive for *eaeA*, and thus were considered to be potential EPEC strains. In contrast, none of the isolates carried the STEC *stx1* or *stx2* toxin genes, or the ETEC toxin genes *LT-I* or *ST α* . The percentage of *eaeA* positive isolates present in a given sample varied from 0 to 11.8% (Figure 2), both within and across sampling years. G-test analysis, done to compare individual samples collected at common dates and times at each of the two sampling sites, indicated that the presence of potential EPEC isolates was not evenly distributed among samples ($P < 0.001$). While potential EPEC strains were found at both sampling locations, isolates carrying the *eaeA* virulence factor gene were present at a greater frequency ($P < 0.001$) for site B compared to site C (4.8 and 1.5%, respectively). Since relatively few *E. coli* strains were obtained from the 12 p.m. samples, the frequency of *eaeA* positive strains was not compared across sampling times.

The 875 potential EPEC isolates, those carrying the *eaeA* gene, were also examined by PCR (Franke *et al.* 1994) for the presence of the *EAF* plasmid. Of the 875 potential EPEC isolates examined, 8 (0.9%)

were found to carry the *EAF* plasmid. Consequently, these isolates were considered to be typical EPEC strains (Ishii *et al.* 2007b). The PCR-RFLP subtype analysis indicated that 87.5% of the 875 potential EPEC strains carried the β intimin subtype (Figure 3, Table 1). While 6.2 and 3.1% of the potential EPEC strains were comprised of *eae* subtypes κ and ξ , respectively, the intimin subtypes α -2, ζ , η , θ , ι , and ν were found at lower frequencies among the strains examined.

Phylogenetic grouping analysis was used to investigate the evolutionary origins of the potentially pathogenic EPEC strains. Preliminary assignment into phylogenetic groups A, B1, B2, and D/E was

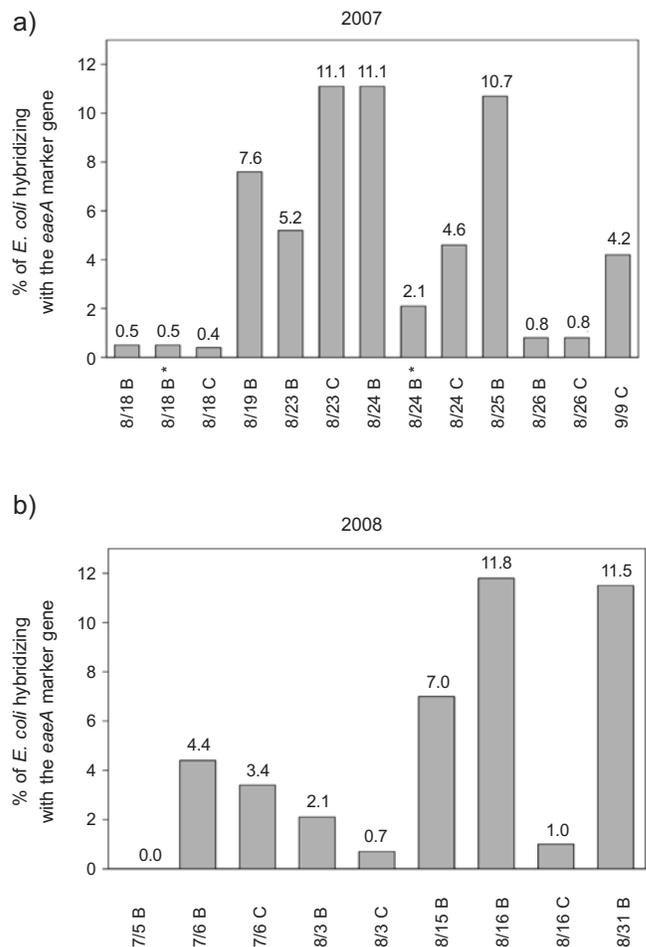


Figure 2. Frequency of *E. coli* isolates carrying the intimin (*eaeA*) gene from water samples collected at Avalon Bay. Samples were collected in 2007 and 2008. All samples were collected at 8:00 a.m., except those indicated by * were collected at 12:00 p.m. Values above the bars are percentages of isolates containing the *eaeA* gene.

based on multiplex PCR detection of the *chuA* and *yjaA* genes, chromosomal TSPE4.C2 DNA (Clermont *et al.* 2000), and a PCR reaction to detect the presence of *ibeA* (Gordon *et al.* 2008). Of the 875 isolates examined, the greatest number (70.5%) belonged to phylogenetic group B1, followed by those in phylogenetic group B2 (25.0%). Strains in phylogenetic groups A and D were represented at lower frequencies, 3.2 and 1.3%, respectively. The relationship between phylogenetic group classification and intimin subtype is shown in Figure 3. While the majority of strains in phylogenetic groups A, B1, and D/E carried the β intimin subtype, those in phylogenetic group B2 carried a greater variety of intimin subtypes. The phylogenetic groups B1 and B2 strains, and β intimin subtypes represented 95.5%, and 63 to 96% of the isolates examined, respectively.

The genetic relatedness of the potential EPEC isolates was determined by using HFERP DNA fingerprint analysis. Dendrograms created using data from all 875 *eaeA*⁺ isolates showed that the strains were distributed into several distinct and genetically diverse groups containing clonal and closely related strains. The relative similarity among these strains

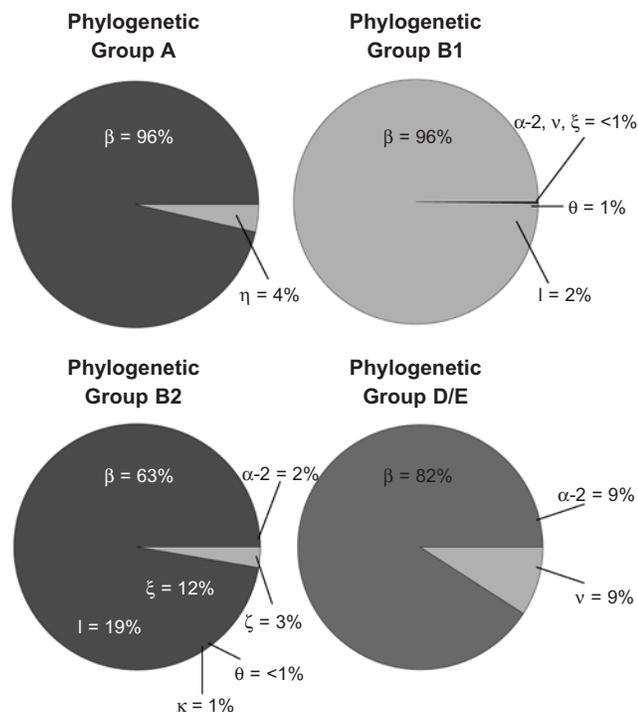


Figure 3. Intimin subtypes among potential EPEC strains separated by *E. coli* phylogenetic groups A (n = 28), B1 (n = 617), B2 (n = 224), and D/E (n = 6). The percentages of intimin subtypes α -2, β , ζ , η , θ , ι , κ , ν , and ξ among strains in each phylogenetic group are shown.

Table 1. Number and frequency of 9 intimin (*eaeA*) subtypes among 875 potential EPEC strains isolated from Avalon Bay, Santa Catalina Island, CA.

Intimin Subtype	Number and Frequency
α -2	7 (0.8%)
β	766 (87.5%)
ζ	6 (0.7%)
η	1 (0.1%)
θ	8 (0.9%)
ι	54 (6.2%)
κ	3 (0.3%)
ν	3 (0.3%)
ξ	27 (3.1%)

ranged from 3.4 to 100% (Figure 4), although the majority of strains were $\geq 40\%$ similar. The genetic relatedness of isolates from 2007 and 2008 ranged from 6.4 to 100% and 2.4 to 100%, respectively. The 875 potential EPEC strains could be divided into four large groups based on their overall genetic relatedness. The largest group consisted of 684 strains having $\geq 60\%$ similarity, the majority of which came from site B in 2007 and 2008. Strains having relative similarity values of $\geq 92\%$ were considered clones (Johnson *et al.* 2004). Based on this criterion, 128 different *eaeA*⁺ potential EPEC strains were isolated during this study. The two largest groups of clonal isolates, designated as groups I and II, contained 211 and 137 isolates, respectively (Figure 5). Interestingly, group I strains were obtained from sites B and C from several different dates in 2007, and from site B on two dates in 2008. In contrast, group II strains only contained EPEC strains from sites B and C isolated from several dates in 2007, and were not found in 2008 samples. The remaining 126 potential EPEC strains were represented by ≤ 64 isolates. Approximately 64% (82 of 128) strains found in this study were represented by a single isolate.

MANOVA analysis of HFERP DNA fingerprints showed that isolates from 2007 and 2008 generally clustered separately into two large groups (Figure 6). While this suggested that the isolates were generally different in each year of the study, the same clonal potential EPEC isolate was sometimes detected at the same or different sampling site over consecutive years (Figure 5). The first and second dis-

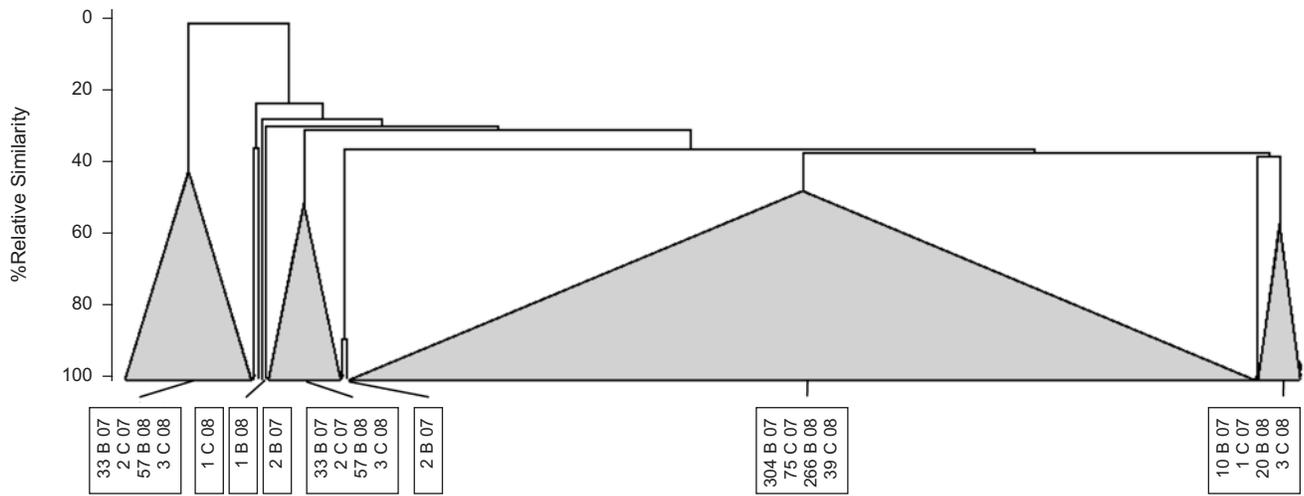


Figure 4. Dendrogram of HFERP DNA fingerprint data obtained from potential EPEC isolates from Avalon Bay, CA. The dendrogram is collapsed at 40% similarity due to size constraints.

crimnants of the MANOVA analysis accounted for ~91% of the variation (Figure 6), indicating the strains clustered well to their respective groups.

DISCUSSION

The water samples used in this study were collected adjacent to two popular swimming beach sites in Avalon Bay, CA. Both of these sites suffer from fecal contamination due to degraded sewer infrastructure and fecal indicator counts at these sites frequently exceed state and federal standards. A recent Natural Resources Defense Council report named these sites as 2 of the 10 most highly contaminated beaches in CA, based on the frequency of samples exceeding the single sample standard of

104 enterococci per 100 ml water (Dorfman and Rosselot 2008).

In this study, we examined Avalon Bay water for the presence of potentially pathogenic *E. coli*. While *E. coli* has historically been used as an indicator of fecal contamination in inland, freshwater environments, this bacterium is generally not quantified in marine environments (USEPA 1986), in part, due to its perceived low survival rates (McCambridge and McMeekin 1981, Anderson *et al.* 1983, Martinez *et al.* 1989). However, this makes *E. Coli* an ideal candidate to examine recent fecal contamination. In contrast, enterococci species are used as the indicator for fecal contamination in marine systems because *Enterococcus* counts have been shown to have strong

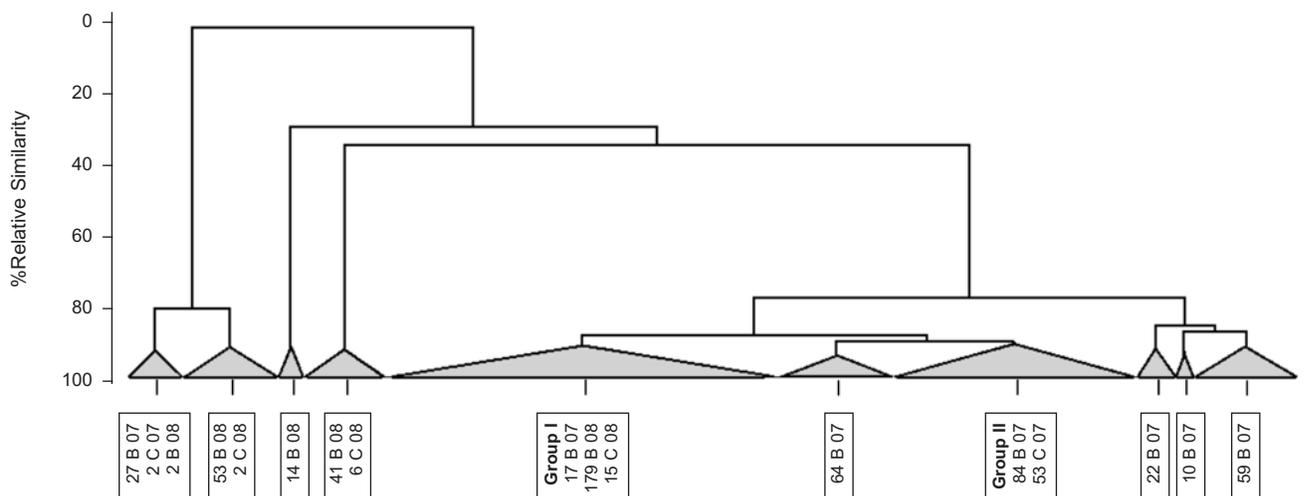


Figure 5. Dendrogram of HFERP DNA fingerprint data obtained from strains represented by 10 or more clonal isolates. Strains having $\geq 92\%$ similarity are considered to be clones. Strain groups I and II are designated.

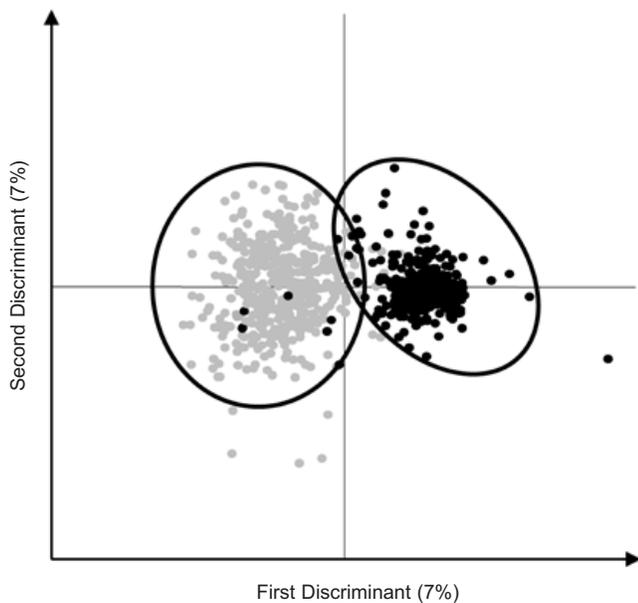


Figure 6. MANOVA analysis of HFERP DNA fingerprint data obtained from all potential EPEC strains isolated in Avalon Bay. Only isolates from site B are shown. Strains isolated in 2007 and 2008 are indicated by (•) and (◦), respectively. Only the first and second discriminants are shown; together they represent 98% of the variation.

correlation to human disease rates (Cabelli *et al.* 1982, Dufour 1984), and this bacterium survives well in marine water (Hanes and Fragala 1967). Despite this fact, *E. coli* was readily isolated from marine waters at Avalon Bay, and potentially pathogenic strains were detected. This may impact the health of swimmers that ingest contaminated water.

Initial screening of the 24,493 *E. coli* isolates collected from marine water samples was done using an automated arraying system and methodology which was originally developed to determine sources of fecal bacteria in waterways using host source-specific marker genes (Yan *et al.* 2007). Results obtained in the study presented here indicate that the high throughput technology can be adapted to determine the presence of any gene among large populations of culturable bacteria. Of the isolates examined, 875 (3.6%) were found to be potential EPEC strains based on detection of the *eaeA* gene. Since the majority (>99%) of the potential EPEC strains described in this study did not harbor the *EAF* plasmid, these strains were classified as atypical EPEC (aEPEC; Nataro and Kaper 1998, Kaper *et al.* 2004).

Currently, there is controversy regarding the pathogenicity of aEPEC, and it has been suggested

that these strains likely arose from STEC that have lost bacteriophages carrying *stx* genes, or EPEC strains that have lost the *bfpA* encoding *EAF* plasmid (Levine and Edelman 1984, Bielaszewska *et al.* 2008). Recently, Tennant *et al.* (2009) showed that 80% of the EPEC strains they examined from clinical and water samples were genetically distinct, carried adhesins that may serve as replacements for the lack of *bfpA*, and suggested that different populations of aEPEC may have varying degrees of pathogenicity. Historically, most cases of EPEC in industrialized countries have been associated with typical strains, but more recently aEPEC strains have been linked to outbreaks of human disease affecting both adults and children around the world (Bokete *et al.* 1997, Hedberg *et al.* 1997, Yatsuyanagi *et al.*, 2002, Afset *et al.* 2003, Nataro 2006, Nguyen *et al.* 2006, Afset *et al.* 2008). More recently, Nguyen *et al.* (2006) reported that patients infected with aEPEC strains were more likely to experience diarrhea lasting longer than two weeks, increasing the risk for serious illness and death. Taken together, these results indicate that aEPEC strains in contaminated recreational water may represent a public health risk to swimmers.

Interestingly, no STEC or ETEC strains were found among the 24,493 isolates screened. This result is similar to that reported by Higgins *et al.* (2005), where the EPEC virulence gene *tir* was detected in water samples at a significantly greater frequency than *stx1/stx2*. The location of the sampling sites may in part explain the lack of STEC isolates, as the common reservoirs for these strains, ruminants and swine (Djordjevic *et al.* 2004, Fratamico *et al.* 2004, Gyles 2007, Ishii *et al.* 2007b), are likely not contributing to the fecal load in the water we examined. Potential input sources into Avalon Bay include only humans, pets, and wildlife. Also, our lack of detection of ETEC strains was not unexpected; these strains are often a major cause of traveler's diarrhea and childhood diarrhea in developing countries (Nataro and Kaper 1998, Kaper *et al.* 2004).

Potential EPEC strains identified in this study were found in every sample examined, but were present at a greater frequency at site B compared to site C (5.3 and 1.9%, respectively). Statistical analysis indicated the EPEC strains were also not evenly distributed among samples collected at common dates and times at different sites ($P < 0.001$). The reason for the uneven distribution is unknown, but

may be related to the proximity of the sampling sites to point sources of fecal contamination, ocean currents present at the sampling sites, and wind and wave action. It is also possible that leaky sewer pipes or run-off originating from humans, pets, and wildlife may have impacted site B to a greater extent than site C.

Results of this study also showed that temporal distribution of the *E. coli* varied considerably. More than 99% of the *E. coli* isolates screened in this study were collected from samples obtained at 8 a.m., and only 3 of the 875 potential aEPEC strains were isolated from the 12 p.m. samples. Additional isolates from 12:00 p.m. samples were not collected due to low *E. coli* counts in samples. Inactivation of fecal indicator bacteria as result of exposure to sunlight has been shown to be a major factor in the persistence of this bacterium in the environment (Davies-Colley *et al.* 1994), and may help to explain the reduced number of culturable organisms in samples collected at 12 p.m.

Phylogenetic analyses, intimin subtyping assays, and HFERP DNA fingerprint analyses revealed a diverse population of potential aEPEC strains. The strains mainly belonged to phylogenetic groups B1 and B2, with frequencies of about 70 and 25%, respectively. In contrast, while strains belonging to phylogenetic groups A and D were also found, they were present at much lower frequencies. Nine intimin subtypes (α -2, β , ζ , η , θ , ι , κ , ν , and ξ) were identified among the potential EPEC isolates by PCR-RFLP analysis. The β , ι , and ζ subtypes were most represented and present in frequencies of 87.5, 6.2, and 3.1%, respectively. Strains possessing the β intimin subtype were assigned to all four phylogenetic groups, although the majority of the strains (77%) were assigned to phylogenetic group B1. This result is consistent with previous reports by Reid *et al.* (2000) and Ramachandran *et al.* (2003) who showed that the majority of strains comprising the β intimin subtype belonged to phylogenetic group B1, and that the β intimin subtype was found at higher frequencies than other subtypes in clinical isolates from humans. The ι intimin subtype was found only in B1 and B2 strains, with approximately 75% of the isolates belonging to phylogenetic group B2. The other intimin subtypes were detected at frequencies of <1.0%. Ishii *et al.* (2007b) reported on the presence of the ι subtype in isolates obtained from ducks and geese, and the κ subtype among *E. coli* from domestic dogs and cats. Humans, pets, and birds are

likely the main contributors to fecal loading at the sampling sites described in this study and, in light of the studies described above, may explain the prevalence of the β , ι , and κ subtypes. Additional studies are necessary to conclusively determine the sources of the potential aEPEC strains in Avalon Bay.

The HFERP DNA fingerprint data showed that the potential EPEC strains belonged to several groups, consisting of both clonal and closely related strains. While 875 of the potential EPEC isolates were comprised of 128 different genotypes, as defined by Johnson *et al.* (Johnson 2004), some were genetically diverse: 650 strains belonged to clonal groups consisting of 10 or more strains. Moreover, 40% of the potential EPEC isolates (348 of 875) belonged to 1 of the 2 clonal groups and were comprised of strains obtained from both sites and across sampling dates. The remaining 60% of the isolates were comprised of 126 different genotypes, which were found only once or in smaller groups and were closely related to other strains in most cases. Since some of the clonal potential aEPEC strains were isolated during different years, our results suggest that there is likely a reservoir of *E. coli* contributing to fecal contamination at Avalon Bay.

Cluster analysis done using MANOVA confirmed results obtained using the correlation analyses represented by dendrograms and showed that isolates cluster relatively well by year into overlapping groups of isolates from each site. Taken together, these results suggest that a few different potential EPEC strains are predominantly isolated from both sites, at least during the summer months, and that most other strains exist transiently. Also, the population of potential EPEC strains may shift in successive years. The reason(s) why some strains were detected at much greater frequencies over a range of dates than other strains is not clear, but may be due to continual deposition as the result of an unknown reservoir or through persistence in the environment.

While the presence and detection of potential EPEC strains in environmental samples has been previously reported, the majority of these studies examined freshwater environments and much fewer isolates were examined (Lauber *et al.*, 2003, Hamelin *et al.* 2006, Ishii *et al.* 2007a). To our knowledge, this present study represents the largest examination of virulence genes among *E. coli* strains obtained from marine recreational water completed to date. Although the presence of virulence determinants alone cannot be used to determine the patho-

genicity of strains, results from this study show that potential EPEC strains can be found in marine water at Avalon Bay, adjacent to some of the most highly contaminated beaches in CA. While the ability of the strains described in this study to cause human disease has yet to be determined, potential EPEC strains have previously been implicated in human diarrheal diseases. Screening of the potential EPEC strains for other virulence factor genes, serotype testing, and other assays may provide further evidence to support this hypothesis. A quantitative measure of the health risk associated with exposure to contaminated water containing these strains also needs to be established through epidemiological studies.

LITERATURE CITED

- Afset, J.E., E. Anderssen, G. Bruant, J. Harel, L. Wieler and K. Bergh. 2008. Phylogenetic backgrounds and virulence profiles of atypical enteropathogenic *Escherichia coli* strains from a case-control study using multilocus sequence typing and DNA microarray analysis. *Journal of Clinical Microbiology* 46:2280-2290.
- Afset, J.E., K. Bergh and L. Bevanger. 2003. High prevalence of atypical enteropathogenic *Escherichia coli* (EPEC) in Norwegian children with diarrhea. *Journal of Medical Microbiology* 52:1015-1019.
- Ahmed, W., J. Tucker, K.A. Bettelheim, R. Neller and M. Katouli. 2007. Detection of virulence genes in *Escherichia coli* of an existing metabolic database to predict the sources of pathogenic *E. coli* in surface waters. *Water Research* 41:3785-3791.
- Anderson, I.C., M. Rhodes and H.I. Kator. 1983. Seasonal variations in survival of *Escherichia coli* exposed in situ in membrane diffusion chambers containing filtered and nonfiltered estuarine water. *Applied and Environmental Microbiology* 45:1877-1883.
- Bielaszewska, M., B. Middendorf, R. Köck, A.W. Friedrich, A. Fruth, H. Karch, M.A. Schmidt and A. Mellmann. 2008. Shiga toxin-negative attaching and effacing *Escherichia coli*: Distinct clinical associations with bacterial phylogeny and virulence traits and inferred in-host pathogen evolution. *Clinical Infectious Diseases* 47:208-217.
- Black, R.E. 1990. Epidemiology of traveler's diarrhea and relative importance of various pathogens. *Reviews of Infectious Diseases* 12:S73-S79.
- Bokete, T.N., T.S. Wittam, R.A. Wilson, C.R. Clausen, C.M. O'Callahan, S.L. Moseley, T.R. Fritsche and P.I. Tarr. 1997. Genetic and phenotypic analysis of *Escherichia coli* with enteropathogenic characteristics isolated from Seattle children. *Journal of Infectious Diseases* 175:1382-1389.
- Byappanahalli, M.N. and R.S. Fujioka. 1998. Evidence that tropical soil environment can support the growth of *Escherichia coli*. *Water, Science and Technology* 38:171-174.
- Byappanahalli, M.N., R.L. Whitman, D.A. Shively, M.J. Sadowsky and S. Ishii. 2006. Population structure, persistence, and seasonality of autochthonous *Escherichia coli* in temperate, coastal forest soil from a Great Lakes watershed. *Environmental Microbiology* 8:504-513.
- Cabelli, V.J., A.P. Dufour, L.J. McCabe and M.A. Levin. 1982. Swimming-associated gastroenteritis and water quality. *American Journal of Epidemiology* 115:606-616.
- Clermont, O., S. Bonacorsi and E. Bingen. 2000. Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Applied and Environmental Microbiology* 66:4555-4558.
- Davies, C.M., J.A. Long, M. Donald and N.J. Ashbolt. 1995. Survival of fecal microorganisms in marine and freshwater sediments. *Applied and Environmental Microbiology* 61:1888-1896.
- Davies-Colley, R.J., R.G. Bell and A.M. Donnison. 1994. Sunlight inactivation of enterococci and fecal coliforms within sewage effluent diluted in seawater. *Applied and Environmental Microbiology* 60:2049-2058.
- Desmarais, T.R., H.M. Solo-Gabriele and C.J. Palmer. 2002. Influence of soil on fecal indicator organisms in a tidally influenced subtropical environment. *Applied Environmental Microbiology* 68:1165-1172.
- Dietzman, D.E., G.W. Fischer and F.D. Schoenkecht. 1974. Neonatal *Escherichia coli* septicemia-bacterial counts in blood. *Journal of Pediatrics* 85:128-130.
- Djordjevic, S.P., V. Ramachandran, K.A. Bettelheim, B.A. Vanselow, P. Holst, G. Bailey and M.A. Hornitzky. 2004. Serotypes and virulence gene profiles of Shiga toxin-producing *Escherichia coli*

- strains isolated from feces of pasture-fed and lot-fed sheep. *Applied and Environmental Microbiology* 70:3910-3917.
- Dombek, P.E., L.K. Johnson, S.T. Zimmerley and M.J. Sadowsky. 2000. Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. *Applied and Environmental Microbiology* 66:2572-2577.
- Dorfman, M. and K.S. Rosselot. 2008. Testing the Waters: A Guide to Water Quality at Vacation Beaches. Natural Resources Defense Council. New York, NY.
- Dufour, A.P. 1984. Health Effects Criteria for Fresh Recreational Waters. United States Environmental Protection Agency. Research Triangle Park, NC.
- Escobar-Páramo, P., O. Clermont, A.B. Blanc-Potard, H. Bui, C.L. Bouguenec and E. Denamur. 2004. A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. *Molecular Biology and Evolution* 21:1085-1094.
- Franke, J., S. Franke, H. Schmidt, A. Schwarzkopf, L.H. Wieler, G. Baljer, L. Beutin and H. Karch. 1994. Nucleotide sequence analysis of enteropathogenic *Escherichia coli* (EPEC) adherence factor probe and development of PCR for rapid detection of EPEC harboring virulence plasmids. *Journal of Clinical Microbiology* 32:2460-2463.
- Fratamico, P.M., L.K. Bagi, E.J. Bush and B.T. Solow. 2004. Prevalence and characterization of Shiga toxin-producing *Escherichia coli* in swine feces recovered in the National Animal Health Monitoring System's Swine 2000 Study. *Applied and Environmental Microbiology* 70:7173-7178.
- Gordon, D.M., O. Clermont, H. Tolley and E. Denamur. 2008. Assigning *Escherichia coli* strains to phylogenetic groups: multi-locus sequence typing versus the PCR triplex method. *Environmental Microbiology* 10:2484-2496.
- Gyles, C.L. 2007. Shiga toxin-producing *Escherichia coli*: An overview. *Journal of Animal Science* 85:E45-E62.
- Hamelin, K., G. Bruant, A. El-Shaarawi, S. Hill, T.A. Edge, S. Bekal, M.J. Fairbrother, J. Harel, C. Maynard, L. Masson and R. Brousseau. 2006. A virulence and antimicrobial resistance DNA microarray detects a high frequency of virulence genes in *Escherichia coli* isolates from Great Lakes recreational waters. *Applied and Environmental Microbiology* 72:4200-4206.
- Hamilton, M.J., T. Yan and M.J. Sadowsky. 2006. Development of goose- and duck-specific DNA markers to determine sources of *Escherichia coli* in waterways. *Applied and Environmental Microbiology* 72:4012-4019.
- Hanes, B. and C. Fragala. 1967. Effect of seawater concentration on the survival of indicator bacteria. *Water Pollution Control Federation Journal* 39:97-104.
- Hardina, C.M. and R.S. Fujioka. 1991. Soil: The environmental source of *E. coli* and enterococci in Hawaii's streams. *Environmental Toxicology and Water Quality* 6:185-195.
- Hedberg, C.W., S.J. Savarino, J.M. Besser, C.J. Paulus, V.M. Thelen, L.J. Myers, D.N. Cameron, T.J. Barrett, J.B. Kaper and M.T. Osterholm. 1997. An outbreak of foodborne illness caused by *Escherichia coli* O39:NM, an agent not fitting into the existing scheme for classifying diarrheogenic *E. coli*. *Journal of Infectious Diseases* 176:1625-1628.
- Higgins, J.A., K.T. Belt, J.S. Karns, J. Russell-Anelli and D.R. Shelton. 2005. *tir*- and *stx*-positive *Escherichia coli* in stream waters in a metropolitan area. *Applied and Environmental Microbiology* 71:2511-2519.
- Ishii, S., D.L. Hansen, R.E. Hicks and M.J. Sadowsky. 2007a. Beach sand and sediments are temporal sinks and sources of *Escherichia coli* in Lake Superior. *Environmental Science & Technology* 41:2203-2209.
- Ishii, S., W.B. Ksoll, R.E. Hicks and M.J. Sadowsky. 2006a. Presence and growth of naturalized *Escherichia coli* in temperate soils from Lake Superior watersheds. *Applied and Environmental Microbiology* 72:612-621.
- Ishii, S., K.P. Meyer and M.J. Sadowsky. 2007b. Relationship between phylogenetic groups, genotypic clusters, and virulence gene profiles of *Escherichia coli* strains from diverse human and animal sources. *Applied and Environmental Microbiology* 73:5703-5710.

- Ishii, S., T. Yan, D.A. Shively, M.N. Byappanahalli, R.L. Whitman and M.J. Sadowsky. 2006b. *Cladophora* (Chlorophyta) spp. harbor human bacterial pathogens in nearshore water of Lake Michigan. *Applied and Environmental Microbiology* 72:4545-4553.
- Johnson, L.K., M.B. Brown, E.A. Carruthers, J.A. Ferguson, P.E. Dombek and M.J. Sadowsky. 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. *Applied and Environmental Microbiology* 70:4478-4485.
- Kaper, J.B., J.P. Nataro and H.L.T. Mobley. 2004. Pathogenic *Escherichia coli*. *Nature Reviews Microbiology* 2:123-140.
- Lang, A. L., Y.L. Tsai, C. L. Mayer, K.C. Patton and C.J. Palmer. 1994. Multiplex PCR for detection of the heat-labile toxin gene and shiga-like toxin I and II genes in *Escherichia coli* isolated from natural waters. *Applied and Environmental Microbiology* 60:3145-3149.
- Lauber, C.L., L. Glatzer and R.L. Sinsabaugh. 2003. Prevalence of pathogenic *Escherichia coli* in recreational waters. *Journal of Great Lakes Research* 29:301-306.
- Levine, M.M. 1987. *Escherichia coli* that cause diarrhea: Enterotoxigenic, enteropathogenic, enteroinvasive, enterohemorrhagic, and enteroadherent. *Journal of Infectious Diseases* 155:377-389.
- Levine, M.M. and R. Edelman. 1984. Enteropathogenic *Escherichia coli* of classic serotypes associated with infant diarrhea: Epidemiology and pathogenesis. *Epidemiologic Reviews* 6:31-51.
- Manges, A.R., J.R. Johnson, B. Foxman, T.T. O'Bryan, K.E. Fullerton. and L.W. Riley. 2001. Widespread distribution of urinary tract infections caused by a multidrug-resistant *Escherichia coli* clonal group. *New England Journal of Medicine* 345:1007-1013.
- Martinez, J., J. Garcia-Lara and J. Vives-Rego. 1989. Estimation of *Escherichia coli* mortality in seawater by the decrease in ³H-label and electron transport system activity. *Microbial Ecology* 17:219-225.
- McCambridge, J. and T.A. McMeekin. 1981. Effect of solar radiation and predacious microorganisms on survival of fecal and other bacteria. *Environmental Microbiology* 41:1083-1087.
- Nataro, J.P. 2006. Atypical enteropathogenic *Escherichia coli*: typical pathogens? *Emerging Infectious Diseases* 12:696.
- Nataro, J.P. and J.B. Kaper. 1998. Diarrheagenic *Escherichia coli*. *Clinical Microbiology Reviews* 11:142-201.
- Nguyen, R.N., L.S. Taylor, M. Tauschek and R.M. Robins-Browne. 2006. Atypical enteropathogenic *Escherichia coli* infection and prolonged diarrhea in children. *Emerging Infectious Diseases* 12:597-603.
- Paton, A.W. and J.C. Paton. 1998. Detection and characterization of Shiga toxigenic *Escherichia coli* by using multiplex PCR assays for *stx1*, *stx2*, *eaeA*, enterohemorrhagic *E. coli* *hlyA*, *rfbO111*, and *rfbO157*. *Journal of Clinical Microbiology* 36:598-602.
- Prüss, A. 1998. Review of epidemiological studies on health effects from exposure to recreational water. *International Journal of Epidemiology* 27:1-9.
- Ramachandran, V., K. Brett, M.A. Hornitzky, M. Dowton, K.A. Bettelheim, M.J. Walker and S.P. Djordjevic. 2003. Distribution of intimin subtypes among *Escherichia coli* isolates from ruminant and human sources. *Journal of Clinical Microbiology* 41:5022-5032.
- Ram, S., P. Vajpayee and R. Shanker. 2007. Prevalence of multi-antimicrobial-agent resistant, shiga toxin and enterotoxin producing *Escherichia coli* in surface waters of River Ganga. *Environmental Science & Technology* 41:7383-7388.
- Reid, S.D., C.J. Herbelin, A.C. Bumbaugh, R.K. Selander and T.S. Whittam. 2000. Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature* 406:64-67.
- Stacy-Phipps, S., J.J. Mecca and J.B. Weiss. 1995. Multiplex PCR assay and simple preparation method for stool specimens to detect enterotoxigenic *Escherichia coli* DNA during course of infection. *Journal of Clinical Microbiology* 33:1054-1059.
- Tennant, S.M., M. Tauschek, K. Azzopardi, A. Bigham, V. Bennett-Wood, E.L. Hartland, W. Qi, T.S. Whittam and R.M./ Robins-Browne. 2009.

Characterisation of atypical enteropathogenic *E. coli* strains of clinical origin. *BMC Microbiology* 9:117-128.

Trabulsi, L.R., R. Keller and T.A.T. Gomes. 2002. Typical and atypical enteropathogenic *Escherichia coli*. *Emerging Infectious Diseases* 8:508-513.

United States Environmental Protection Agency (USEPA). 1986. Bacteriological Water Quality Criteria for Marine and Fresh Waters. USEPA. Washington, DC.

USEPA. 2002. Method 1603: *Escherichia coli* (*E. coli*) in water by membrane filtration using modified membrane-thermotolerant *Escherichia coli* agar. USEPA. Washington, DC.

Yan, T., M.J. Hamilton and M.J. Sadowsky. 2007. High-throughput and quantitative procedure for determining sources of *Escherichia coli* in waterways by using host-specific DNA marker genes. *Applied and Environmental Microbiology* 73:890-896.

Yan, T. and M.J. Sadowsky. 2007. Determining sources of fecal bacteria in waterways. *Environmental Monitoring and Assessment* 129:97-106.

Yatsuyanagi, J., S. Shioko, Y. Miyajima, K. Amano and K. Enomoto. 2002. Characterization of atypical enteropathogenic *Escherichia coli* strains harboring the *astA* gene that were associated with a waterborne outbreak of diarrhea in Japan. *Journal of Clinical Microbiology* 40:294-297.

ACKNOWLEDGEMENTS

The authors would like to thank Yiping Cao, Ben Ferraro, and Nick Miller from the Southern California Coastal Water Research Project for their assistance with sample collection and processing. They would also like to thank Nick Hahn and the High-Throughput Biological Analysis facility at the University of Minnesota for assistance with automated arraying, Daniel Norat and Chris Brandsey for help with colony-picking, and John Ferguson for help with cluster analyses. This work was funded, in part, by a grant from the Minnesota Agricultural Research Station (to MJS), and by training grant 2T32-GM008347 from the National Institutes of Health (to MJH).