AEM
Journals.ASM.org

Estimating the Prevalence of Potential Enteropathogenic *Escherichia coli* and Intimin Gene Diversity in a Human Community by Monitoring Sanitary Sewage

Kun Yang,* Eulyn Pagaling, Tao Yan

Department of Civil and Environmental Engineering, University of Hawaii at Manoa, Honolulu, Hawaii, USA

Presently, the understanding of bacterial enteric diseases in the community and their virulence factors relies almost exclusively on clinical disease reporting and examination of clinical pathogen isolates. This study aimed to investigate the feasibility of an alternative approach that monitors potential enteropathogenic *Escherichia coli* (EPEC) and enterohemorrhagic *E. coli* (EHEC) prevalence and intimin gene (*eae*) diversity in a community by directly quantifying and characterizing target virulence genes in the sanitary sewage. The quantitative PCR (qPCR) quantification of the *eae*, stx_1, and stx_2 genes in sanitary sewage samples collected over a 13-month period detected *eae* in all 13 monthly sewage samples at significantly higher abundance (93 to 7,240 calibrator cell equivalents [CCE]/100 ml) than stx_1 and stx_2, which were detected sporadically. The prevalence level of potential EPEC in the sanitary sewage was estimated by calculating the ratio of *eae* to *uidA*, which averaged 1.0% ($\sigma = 0.4\%$) over the 13-month period. Cloning and sequencing of the *eae* gene directly from the sewage samples covered the majority of the *eae* diversity in the sewage and detected 17 unique *eae* alleles belonging to 14 subtypes. Among them, *eae*-β2 was identified to be the most prevalent subtype in the sewage, with the highest detection frequency in the clone libraries (41.2%) and within the different sampling months (85.7%). Additionally, sewage and environmental *E. coli* isolates were also obtained and used to determine the detection frequencies of the virulence genes as well as *eae* genetic diversity for comparison.

Although the majority of *Escherichia coli* strains are harmless commensal microorganisms in human intestine, numerous pathotypes have been identified to cause severe human diseases, including enteropathogenic *E. coli* (EPEC) and enterohemorrhagic *E. coli* (EHEC) (1). EPEC is commonly associated with severe infant diarrhea; although large outbreaks of infant EPEC-caused diarrhea are rare in the developed world, EPEC-related diarrhea is still one of the most important causes of infant mortality in developing countries (2). EHEC strains, on the other hand, are frequently associated with food-borne outbreaks in the developed world, which are characterized by bloody diarrhea and hemolytic uremic syndrome (HUS) (2, 3). For instance, the EHEC strain O157:H7 has caused numerous food-borne outbreaks and more than 73,000 cases of disease in the United States (4–6), and the recent outbreak of EHEC O104:H4 strains in Europe was the deadliest ever recorded, infecting more than 3,900 people and causing 46 deaths (7).

As with many other pathogenic *E. coli* strains, the virulence factors of EPEC and EHEC are associated with mobile genetic elements (1). A common feature of EPEC and EHEC infection is the "attaching-and-effacing" (AE) histopathology, which is encoded by genes on a chromosomal pathogenicity island called the locus of enterocyte effacement (LEE) that was acquired through horizontal gene transfer (8–11). EHEC strains contain additional virulence factors, including the verocytotoxin Stx encoded by the stx_1 and stx_2 genes that are associated with bacteriophage (4). The stx_1 and stx_2 genes of EHEC strains confer severe disease symptoms in humans, whereas the LEE and its associated genes of EHEC and EPEC strains enable intimate attachment of *E. coli* cells to epithelial cells and the colonization of intestinal mucosa and, thus, are essential for the onset of diseases (8, 9, 12, 13). The parallel-evolution theory of pathogenic *E. coli* strains suggests that EHEC strains such as *E. coli* O157:H7 and O104:H4 evolved from EPEC-like ancestors by sequentially acquiring molecular mechanisms through horizontal gene transfer that gradually conferred increased virulence (14–16).

The most studied LEE gene is the *eae* gene that encodes intimin adhesin, an outer membrane protein essential for the formation of the characteristic AE lesion of EPEC and EHEC strains (8, 9, 12, 13). The *eae* gene has been shown to be indispensable to the infectivity of EPEC and EHEC strains. An isogenic *eae* null mutant led to the loss of infectivity of an EPEC strain in human volunteers (12), and the deletion of the *eae* gene rendered an EHEC O157:H7 strain noninfective in animal models (8, 9). The importance of *eae* to pathogenicity is also illustrated by the observation that many Stx-producing *E. coli* strains are not pathogenic because of their lack of the *eae* gene.

Previous studies using clinical EPEC and EHEC isolates have revealed extraordinary genetic diversity in the *eae* gene. To date, at least 20 different *eae* subtypes, including *eae*-α1, *eae*-α2, *eae*-β, *eae*-β2, *eae*-γ, *eae*-γ3, *eae*-γ6, *eae*-θ, *eae*-τ1, *eae*-τ2, *eae*-ο, *eae*-κ, *eae*-ε, *eae*-ε2, *eae*-ε8, *eae*-ζ, *eae*-η, *eae*-λ, *eae*-ρ, and *eae*-σ, have been reported in the literature and in the GenBank database (17–20). The different *eae* subtypes exhibit significant genetic variation among themselves (e.g., more than 15% of amino acid sequence difference was observed between the *eae*-α, *eae*-β, and *eae*-γ sub-

Received 15 August 2013 **Accepted** 8 October 2013
Published ahead of print 18 October 2013
Address correspondence to Tao Yan, taoyan@hawaii.edu.
* Present address: Kun Yang, Department of Pharmaceutical and Biological Engineering, Sichuan University, Chengdu, China.

doi:10.1128/AEM.02747-13

types [21]), and the remarkable *eae* genetic diversity is believed to be involved in host and tissue tropism (20, 22). The majority of *eae* genetic variations are observed at the 3′ end of the *eae* gene (23, 24), which encodes the highly variable, C-terminal extracellular domains responsible for receptor binding (25). The *eae* genetic diversity could also result from the variants of the *tir* gene, which is also located on the LEE and encodes the translocated intimin receptor (Tir) protein that gets anchored into epithelial cells for intimin binding (26).

Presently, our understanding of bacterial enteric diseases in the community and of their virulence factors is based primarily on clinical disease reporting and examination of clinical pathogen isolates (27–29). This study aimed to investigate the feasibility of an alternative approach that monitors enteric disease prevalence and virulence factor diversity by directly quantifying and characterizing target virulence genes in the sanitary sewage of a human community. Specifically, the study used cultivation-independent quantitative PCR (qPCR) methods to quantify the concentrations of the *eae*, stx_1, and stx_2 genes in sanitary sewage over 13 months and estimated the prevalence levels of potential EPEC and EHEC strains in the community. The genetic diversity of the intimin gene *eae* in the community was estimated by constructing *eae* clone libraries for the sanitary sewage samples directly. Additionally, sewage and environmental *E. coli* isolates were also obtained and used to determine the detection frequencies of the virulence genes as well as *eae* genetic diversity, which was compared with the cultivation-independent DNA-based approaches.

MATERIALS AND METHODS

Sample collection. Sanitary sewage samples were collected at the Sand Island Wastewater Treatment Plant (SIWTP) in Honolulu, HI, over a 13-month period (April 2010 to April 2011). The SIWTP collects and treats approximately 60% of the sewage in the City of Honolulu. Raw sewage samples (1 liter) were collected hourly using an autosampler at a location immediately before the primary clarifier, and 40 ml of the completely mixed hourly samples was mixed to make daily composite samples. The daily composite samples were stored at 4°C in the dark until the end of the week, when 100 ml of completely mixed daily samples was pooled to make weekly composite samples. After mixing, the weekly composite samples (100 ml) were immediately centrifuged at 13,000 × *g* for 15 min at 4°C to pellet suspended solids and cells, and the supernatants were vacuum filtered through 0.45-μm-pore-size cellulose ester membrane filters. The pellets and cell-bearing membranes were subsequently pooled for each wastewater sample and stored at −80°C until required for DNA extraction and subsequent analysis. Two additional grab sewage samples were collected on 18 June and 2 July 2008 and used for the isolation of sewage *E. coli* strains.

***E. coli* isolation.** *E. coli* isolates were obtained from the grab sewage samples using the standard modified membrane-thermotolerant *E. coli* (mTEC) agar method (30). Briefly, the municipal wastewater samples were diluted in phosphate-buffered saline before being spread plated. Individual colonies with typical *E. coli* characteristics were purified by repetitive streaking on LB agar, followed by the IMViC (indole, methyl red, Voges-Proskauer, citrate) tests for *E. coli* verification. A total of 236 *E. coli* isolates were obtained from the sewage samples, which herein are referred to as sewage *E. coli* isolates. *E. coli* isolates from stream water and soil samples that were collected in 2009 from the Manoa watershed in a separate study (31), which herein are referred to as environmental *E. coli* isolates, were also used in this study for comparison. A total of 467 environmental *E. coli* isolates were selected from the original collection, which includes 288 isolates from soil and 179 isolates from stream water.

DNA extraction and qPCR quantification. The weekly wastewater samples were subjected to total genomic DNA extraction using an UltraClean Soil DNA Isolation Kit (MO Bio, Carlsbad, CA) according to the manufacturer's instructions. DNA extracts from the samples from the same month were pooled to make 12 monthly composite DNA samples, which were analyzed by qPCR to quantify *eae*, stx_1, stx_2, and *uidA* genes using target-specific primers and fluorescent probes (Table 1). The 20-μl qPCR mixtures contained 10 μl of 2× iTaq Universal Probe SuperMix (Bio-Rad, Hercules, CA), 0.25 μM each primer, 0.125 μM fluorescent probe, and 0.4 μg/μl bovine serum albumin (BSA). The qPCRs were performed on an ABI 7300 System (Applied Biosystem, Foster City, CA). The thermocycler program included 5 min of initial denaturation at 95°C and 45 cycles of amplification for 15 s at 95°C, followed by 1 min at 60°C. *E. coli* O157:H7 cells were used to construct calibration curves for the *uidA*, *eae*, stx_1, and stx_2 genes. Exponential-phase *E. coli* O157:H7 cells were serially diluted to make calibration standards of known numbers of cells (10^1 to 10^6 CFU/ml), which were then subjected to DNA extraction using a GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich). Genomic DNA extracted from the sewage samples and the calibration standards were analyzed in the same batch of reactions.

Cloning of *eae* genes from sewage DNA extracts. Clone libraries of *eae* genes were constructed using PCR amplicons from the monthly sewage DNA extracts. Two rounds of PCR using the same PCR primers (eae-F1 and escD-R1) (Table 1) with an intermediate gel extraction step were carried out to enhance PCR amplification. The 25-μl PCR mixture contained 2.5 μl of 10× AmpliTaq Gold 360 buffer, 3 mM $MgCl_2$, 1 μl of 360 GC enhancer, 0.25 mM each deoxynucleoside triphosphate (dNTP), 0.1 μM each primer, 0.4 mg/ml BSA, 0.625 units of AmpliTaq Gold 360 DNA Polymerase (Invitrogen, Carlsbad, CA), and 1 μl of DNA template. The PCR was initialized at 95°C for 10 min, followed by 35 cycles of denaturing (95°C, 30s), annealing (57°C, 30s), and extension (72°C, 3 min), with a final extension at 72°C for 7 min. The first-round PCR amplicons were subjected to gel electrophoresis, and gel excision at the expected amplicon location was conducted regardless of the visibility of a DNA band. The excised gel blocks were extracted using a Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI), and the extracts were used as DNA templates in the second round of PCR amplification. This procedure successfully amplified seven monthly composite DNA samples (August 2010, October 2010, November 2010, December 2010, February 2011, March 2011, and April 2011). The PCR amplicons of the second PCR amplification were gel purified and ligated into a pGEM T-easy cloning vector (Promega), according to the manufacturer's protocol, and then transformed into *E. coli* DH10B competent cells by electroporation.

Detection of *eae*, stx_1, and stx_2 in *E. coli* isolates. Fresh single colonies of the sewage and environmental *E. coli* isolates were grown overnight in LB broth at 37°C and with constant shaking (200 rpm). One milliliter of the cell cultures was centrifuged at 10,000 × *g* for 5 min, and the cell pellets were resuspended in 50 mM NaOH solution and boiled for 10 min to release genomic DNA. After centrifugation at 10,000 × *g* for 10 min, the supernatants were used as DNA templates to amplify the *eae*, stx_1, and stx_2 genes using multiplex PCR with target-specific primers (Table 1). The 25-μl PCR mixture contained 1 mM dNTPs, 0.1 μM each primer, 2.1 mM $MgCl_2$, 1× reaction buffer (10 mM Tris-HCl and 50 mM KCl), 0.8 μg/μl BSA, 1 unit of *Taq* DNA polymerase, and 1 μl of template DNA. The hot-start technique was used to minimize nonspecific amplification. The thermocycler program included a 5-min initial denaturing step at 95°C, followed by 35 cycles of amplification (95°C for 30 s, 60°C for 30 s, and 72°C for 90 s) and a final extension step (72°C for 5 min). The PCR amplicons were then subjected to gel electrophoresis to detect the presence of target genes.

The *eae* genes from the *E. coli* isolates were PCR amplified and then grouped by restriction fragment length polymorphism (RFLP) to identify the isolates carrying unique *eae* subtypes. Total genomic DNA of the *eae*-positive *E. coli* isolates was used to amplify the variable region of the *eae* gene with primers eae-F1 and escD-R1 (Table 1) using a previously described procedure (18). The PCR amplicons were digested using three

TABLE 1 Primer pairs and probes used in qPCR, multiplex PCR, PCR-RFLP, and sequencing

Assay	Target	Primer name or probe[a]	Sequence (5′–3′)	Size (bp)	Reference(s) or source
qPCR	stx_1	Forward	TTTGTYACTGTSACAGCWGAAGCYTTACG	132	45, 46
		Reverse	CCCCAGTTCARWGTRAGRTCMACRTC		
		Probe	CTGGATGATCTCAGTGGGCGTTCTTATGTAA		
	stx_2	Forward	TTTGTYACTGTSACAGCWGAAGCYTTACG	128	45, 46
		Reverse	CCCCAGTTCARWGTRAGRTCMACRTC		
		Probe	TCGTCAGGCACTGTCTGAAACTGCTCC		
	eae	Forward	CATTGATCAGGATTTTTCTGGTGATA	102	45, 47
		Reverse	CTCATGCGGAAATAGCCGTTA		
		Probe	ATAGTCTCGCCAGTATTCGCCACCAATACC		
	uidA	Forward	GTGTGATATCTACCCGCTTCGC	83	33, 48
		Reverse	AGAACGGTTTGTGGTTAATCAGGA		
		Probe	TCGGCATCCGGTCAGTGGCAGT		
Multiplex-PCR	stx_1	Forward	CAGTTAATGTGGTGGCGAAGG	348	49
		Reverse	CACCAGACAATGTAACCGCTG		
	stx_2	Forward	ATCCTATTCCCGGGAGTTTACG	584	49
		Reverse	GCGTCATCGTATACACAGGAGC		
	eae	Forward	TCAATGCAGTTCCGTTATCAGTT	482	50
		Reverse	GTAAAGTCCGTTACCCCAACCTG		
RFLP	*eae*	eae-F1	ACTCCGATTCCTCTGGTGAC	1,800–2,100[b]	18
		escD-R1	GTATCAACATCTCCCGCCCA		
Sequencing[c]	*eae*	cesT-F9	TCAGGGAATAACATTAGAAA		18
		eae-R3	TCTTGTGCGCTTTGGCTT		
		eae-seq	GMWKMRGWTTGTKTAATCCAAG		This work
		M2eae	GTCGACCAGGTTGGGGTAA		This work

[a] All TaqMan probes all use 6-carboxyfluorescein (FAM) as the reporter dye and 6-carboxy-tetramethylrhodamine (TAMRA) as the quencher dye.
[b] The size of amplicon depends on the allele.
[c] Sequencing primers also include the two used in RFLP.

different restriction enzymes (AluI, HhaI, and RsaI) at 37°C for 8 h, and the digestion products were visualized via gel electrophoresis.

Sequencing. Plasmid extractions were conducted on seven clone libraries using the alkaline lysis method, and the inserts were sequenced using the vector primers M13F and M13R. Both were conducted by the Advanced Studies in Genomics, Proteomics, and Bioinformatics (ASGPB) sequencing facility at the University of Hawaii at Manoa, Honolulu, HI. The 3′ highly variable regions of the inserts were compared within the clone libraries to identify the unique *eae* gene inserts. The number of clones sequenced for the individual clone libraries was adjusted based on the sequencing rarefaction curves in order to exhaust the *eae* diversity in the libraries and to maximize the recovery of unique *eae* sequences. As a result, a total of 328 clones were sequenced, with some clone libraries sequenced more extensively than others. For the identified unique *eae* gene inserts, full-length sequences were then determined using the additional sequencing primers eae-seq and M2eae (Table 1). For the *eae*-positive *E. coli* isolates, the unique *eae* subtypes determined by PCR-RFLP were sequenced by amplifying the whole length of the *eae* gene with primers cesT-F9 and escD-R1 (Table 1). Similarly, additional sequencing primers, including eae-F1, eae-R3, and eae-seq, were used to assemble the full sequence length. Contigs were constructed from the sequences using SeqMan (DNASTAR, Madison, WI) until full-length *eae* genes were obtained.

Data analysis. The concentrations of *uidA*, *eae*, stx_1, and stx_2 are expressed as calibrator cell equivalents (CCE) per 100 ml of sewage sample. The calibration curves for *uidA*, *eae*, stx_1, and stx_2 using *E. coli* O157:H7 cells as calibrator cells all showed an R^2 value larger than 0.98. In calculating geometric means, 0.9 was used to mathematically represent samples with no detection of target genes (i.e., below the detection limit). All gel images of PCR-RFLP were processed using GelCompar II (Applied Maths, Austin, TX). The RFLP banding pattern for each isolate was normalized with an external DNA size marker (DNA Hyperladder I; Bioline, Taunton, MA) that was loaded into the first and last lanes of each gel. Dendrograms were created based on Pearson's correlation and the unweighted-pair group method with arithmetic mean (UPGMA). Rarefaction curves were calculated using the Analytical Rarefaction software package available from the University of Georgia Stratigraphy Lab (http://www.uga.edu/strata/software/index.html). The closest matches of the unique *eae* sequences were obtained by comparison with *eae* gene entries in the GenBank database using BLASTN. The phylogenetic relationships between the *eae* clone sequences and the *eae* genes in sewage and environmental *E. coli* isolates of this study and representative *eae* subtypes from the GenBank database were analyzed using MEGA 5 (32), where a phylogenetic tree was constructed using the maximum-likelihood method of tree inference and the Tamura-Nei nucleotide substitution model for sequence alignment.

Nucleotide sequence accession numbers. The *eae* gene sequences obtained in this study have been deposited in the GenBank database under accession numbers KF771362 to KF771382.

RESULTS AND DISCUSSION

***E. coli* virulence genes in sanitary sewage.** Concentrations of *eae*, stx_1, and stx_2 genes in the municipal wastewater samples over 13 months were quantified using qPCR (Fig. 1A). The *eae* gene was detected in all 13 (100%) sewage DNA samples, while stx_1 and stx_2 were detected in only 2 (15.4%) and 3 (23.1%) of the 13 samples, respectively. The geometric mean concentrations (± geometric standard deviation) of *eae*, stx_1, and stx_2 during the sampling period were 399 (±3.1), 1.5 (±3.5), and 2.1 (±5.3) CCE/100 ml, respectively, which corresponded to an *eae* gene abundance 266

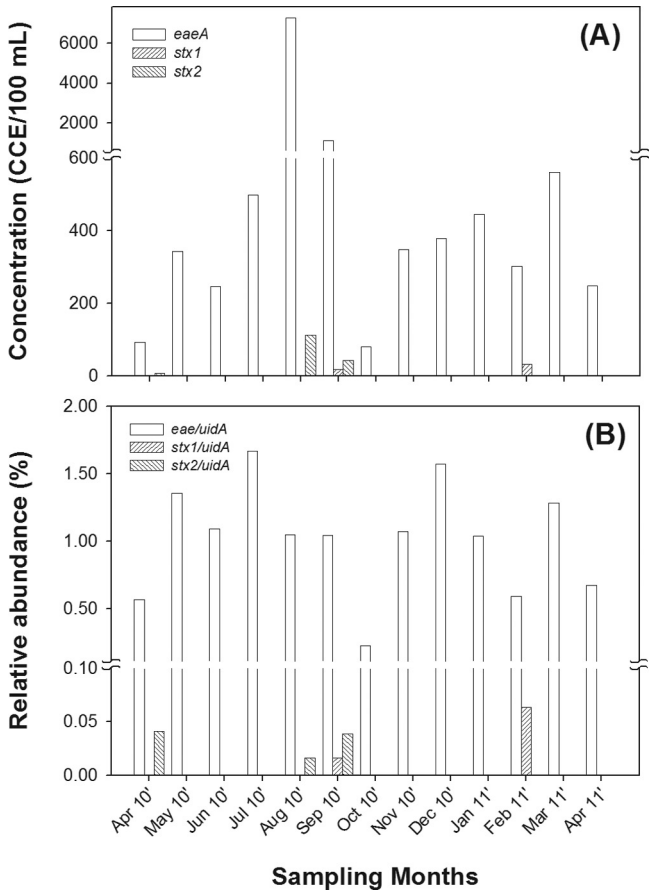

FIG 1 Concentrations of *eae*, stx_1, and stx_2 determined by qPCR (A) and relative abundances as indicated by the ratios of *eae* to *uidA*, stx_1 to *uidA*, and stx_2 to *uidA* (B) in the SIWTP sewage over 13 consecutive months.

and 190 times the abundances of stx_1 and stx_2, respectively. This indicated that the majority of *eae*-bearing *E. coli* in sewage was potential EPEC, which is defined herein as *E. coli* cells with the *eae* gene and without the stx_1 and stx_2 genes, while the abundance levels of EHEC and/or Shiga toxin-producing *E. coli* (STEC) cells were significantly lower and negligible in comparison to the potential EPEC cells. The abundance of potential EPEC cells fluctuated considerably over the sampling period (93 to 7,240 CCE/100 ml), with the two highest concentrations detected in the summer months (August and September 2010). The percentage of potential EPEC within the total *E. coli* population was estimated by calculating the ratio of *eae* to *uidA* gene copy numbers in the wastewater samples (Fig. 1B). This ratio represents the prevalence of potential EPEC in the sanitary sewage and, hence, in the human community. The ratio of *eae* abundance to *uidA* abundance over the 13-month sampling period ranged from 0.2% to 1. 7% and averaged 1.0% ($\sigma = 0.4\%$). The abundance ratios of stx_1 to *uidA* and of stx_2 to *uidA* were very small in comparison (Fig. 1B), which is similar to the pattern of their qPCR-determined concentrations (Fig. 1A).

The detection frequencies and the average concentrations of *eae*, stx_1, and stx_2 indicated significantly different prevalence levels in the sewage and, hence, in the community. Since the concentrations of stx_1 and stx_2 genes were negligible compared to the concentration of the *eae* gene, the prevalence of potential EHEC and/or STEC was expected to be very low, while potential EPEC was likely to be the dominant group of diarrheagenic *E. coli* strains in sanitary sewage. The different abundance levels of EPEC and EHEC in sanitary sewage was expected since EPEC generally causes chronic and mild diarrhea while EHEC is associated with acute and severe disease (1).

Although the *eae* gene was consistently detected in all sewage samples, its absolute concentrations determined by qPCR exhibited huge variation during the sampling months. For instance, the largest *eae* concentration was 77.8 times that of the lowest *eae* concentration (Fig. 1A). Apart from the temporal variation of potential EPEC abundance in the community, various environmental factors, such as sewage dilution by heavy rainfall, sample heterogeneity, and method limitations such as PCR inhibition, could all have contributed to the observed variation. To counter some of the variations caused by these factors, the concentration of the *uidA* gene in the sewage samples, which quantifies the overall *E. coli* population (33), was used as a common denominator. The ratio of *eae* to *uidA* exhibited much less temporal variation and gave a normalized estimation of the prevalence of potential EPEC in the sanitary sewage.

Virulence gene detection in *E. coli* isolates. The relative prevalence levels of potential EPEC and EHEC cells in sanitary sewage were also tested using sewage *E. coli* isolates. A total of 236 sewage *E. coli* isolates obtained from two sampling events were analyzed using multiplex PCR assays that detect the presence of the *eae*, stx_1, and stx_2 genes (Table 2). The stx_1 and stx_2 genes were not detected in any of the 236 sewage *E. coli* isolates, while an average of 26.3% of *E. coli* isolates were *eae* positive. Additionally, a total of 467 *E. coli* isolates that were previously obtained from soil and water samples from the Manoa watershed (31) were also analyzed for the presence of *eae*, stx_1, and stx_2. None of the environmental *E. coli* isolates contained the stx_1 and stx_2 genes, while 1.7% and 1.1% of the *E. coli* isolates from the Manoa soil and water samples were *eae* positive, respectively.

Although the detection frequencies of *eae*, stx_1, and stx_2 in the sewage isolates support the observation made by qPCR that potential EPEC strains were the dominant *eae*-bearing *E. coli* cells in sanitary sewage, the ratio of potential EPEC to the general *E. coli* population exhibited a large variation between the two sampling events, with one being 0% and the other being 44.3% (Table 2). This large variation suggests that the sample sizes (up to 140 isolates per sampling event) were still too small to be representative. The requirement of large numbers of *E. coli* isolates to achieve acceptable levels of representativeness in sewage can be attributed to both the aggregative form of microbial cells in sewage and the large concentration of *E. coli* cells in sanitary sewage (up to 10^6 cells/ml). Previous studies using limited numbers of *E. coli* isolates also reported large variations in the detection frequencies of the *eae* gene, ranging from 0.03% in a beach sand *E. coli* population (34) to 10.4% in an *E. coli* population from wildlife feces (35). In contrast, experiments using high-throughput approaches reported more modest *eae* prevalence in environmental *E. coli* populations; for example, Hamilton et al. (36) found 3.6% of 24,493 *E. coli* isolates from beach water to be *eae* positive. The qPCR approach can largely circumvent the sample size issue by using relatively large volumes of sewage samples (100 ml in this study) in the analysis, which are more likely to provide a reasonable representation of the sewage microbial community.

TABLE 2 Multiplex PCR detection of *eae*, stx_1, and stx_2 in *E. coli* isolates from sewage and environmental samples

Source	Sampling date(s) (mo/day/yr)	No. of isolates	Gene frequency (no. of positive isolates [%])		
			eae	stx_1	stx_2
Sewage	06/08/2008	96	0 (0)	0 (0)	0 (0)
	07/02/2008	140	62 (44.3)	0 (0)	0 (0)
Soil	Various[a]	288	5 (1.7)	0 (0)	0 (0)
Water	Various	179	2 (1.1)	0 (0)	0 (0)

[a] Multiple sampling sites and dates were used.

Intimin gene (*eae*) diversity. The intimin gene (*eae*) diversity in municipal wastewater was investigated by constructing *eae* clone libraries for the monthly composite sewage samples. Seven clone libraries were successfully constructed, from which a total of 328 *eae* gene clones were sequenced. The overall sequencing effort detected 17 unique *eae* sequences, which covered the majority of the *eae* diversity in the wastewater samples, as indicated by the overall rarefaction curve approaching an asymptotic state (Fig. 2A). Rarefaction analysis for the individual monthly clone libraries indicated that the sequencing effort recovered the most dominant *eae* genes within the individual monthly libraries (Fig. 2B).

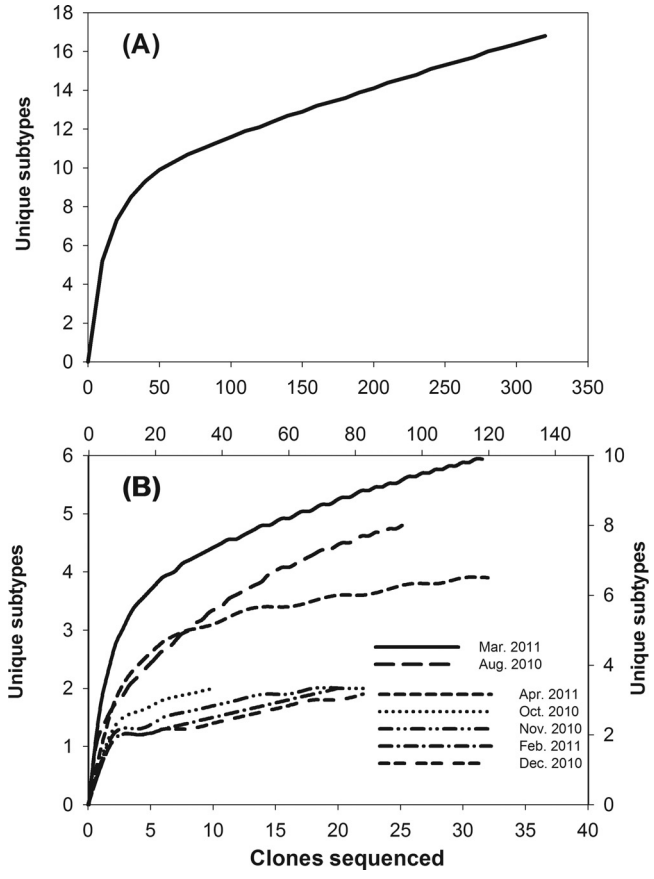

FIG 2 Rarefaction curves of *eae* sequencing in all clone libraries combined (A) and in individual monthly clone libraries (B). In panel B, rarefaction curves for the March 2011 and August 2010 clone libraries use the top and right coordinates, while rarefaction curves for other clone libraries use the bottom and left coordinates.

TABLE 3 Unique *eae* sequences and their accession numbers and closest matches in the GenBank database

Clone no.	Closest match in GenBank			
	Accession no.	Identity (no. of shared bases/ total no. of bases [%])	Subtype	Reference
1A1	AJ271407	1,678/1,695 (99)	ζ	20
1A2	DQ523605	1,919/1,919 (100)	β2	18
1A7	AJ633130	1,244/1,249 (99)	ι2	51
1A9	AP010960	2,001/2,011 (99)	θ	43
1C6	FM180568	1,941/1,952 (99)	α1	52
1D3	AJ308551	1,823/1,824 (99)	ι1	20
1E9	AY696838	1,995/2,001 (99)	ο	42
1F1	DQ523607	1,860/1,880 (99)	γ3	18
1F2	DQ523600	1,925/1,932 (99)	α2	18
1F6	AJ271407	1,901/1,948 (98)	ζ	20
1F9	AJ271407	1,926/1,946 (99)	ζ	20
1G4	CP003109	1,948/1,984 (98)	γ	53
1G8	AP010958	2,023/2,037 (99)	ε	43
2A9	AJ308552	1,926/2,011 (99)	κ	20
2D3	AB647569	1,702/1,710 (99)	ε8	54
2F7	DQ523607	1,678/1,695 (99)	γ3	18
3C11	DQ523613	1,989/1,993 (99)	ρ	18

Different levels of *eae* gene diversity were observed among the different monthly samples; samples for August 2010, March 2011, and April 2011 contained higher levels of *eae* diversity than samples from other months, as indicated by steeper rarefaction curves of these three months.

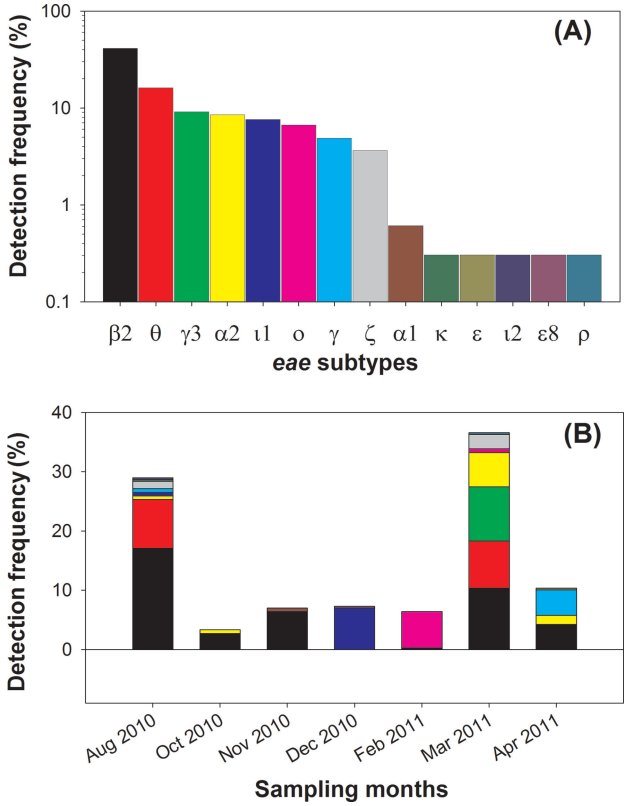

FIG 3 Frequencies of the 13 *eae* subtypes detected in the SIWTP sewage in the clones sequenced (A) and their distribution in the different monthly *eae* clone libraries (B). The same color codes are used for the two panels.

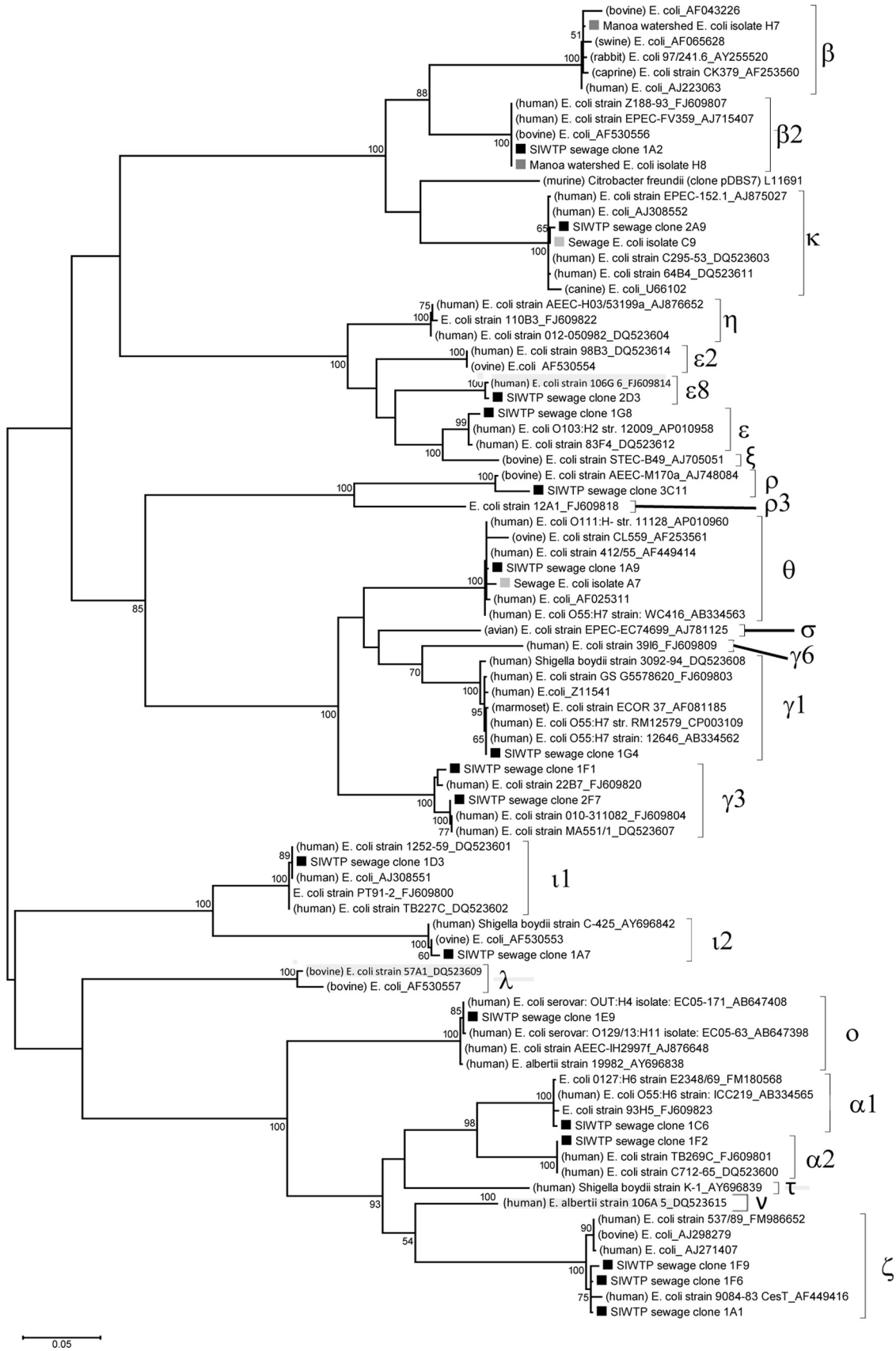

FIG 4 Phylogenetic relationship between the unique SIWTP sewage *eae* clones, unique *eae* genes from sewage and environmental *E. coli* isolates, and known *eae* subtypes from the GenBank database. Entries labeled with black and gray squares represent sequences obtained from clone library sequences and from isolated *E. coli* strains in this study, respectively. The host sources of known *eae* subtypes are provided in parentheses.

The 17 unique *eae* gene sequences were compared with entries from the GenBank database to identify their closest matches and determine their subtypes. All 17 sequences had high identity scores (98 to 100%) to known *eae* gene entries in the database and were subsequently classified to 14 different *eae* subtypes (Table 3). Most of the unique *eae* gene sequences belonged to different subtypes, except for clones 1F1 and 2F3, which were classified as *eae*-γ3, and clones 1A1, 1F6, and 1F9, which were classified as *eae*-ζ.

The highly polymorphic intimin protein corresponds to the extraordinarily high *eae* genetic diversity, which is indicated by the 27 different *eae* alleles deposited in the GenBank database (18). Previous efforts using *E. coli* isolates from various host sources, including humans (17, 19, 20, 37) and ruminants (19, 22, 38), have detected at least 20 different *eae* subtypes. Fourteen of them, including *eae*-β2, *eae*-ι1, *eae*-ο, *eae*-α2, *eae*-γ, *eae*-θ, *eae*-α1, *eae*-γ3, *eae*-κ, *eae*-ε, *eae*-ε8, *eae*-ζ, *eae*-ι2, and *eae*-ρ, were detected in this study by cloning the *eae* gene directly from the sanitary sewage samples. Six *eae* subtypes that were previously reported in the literature, including *eae*-β, *eae*-η, *eae*-ε2, *eae*-λ, *eae*-σ, and *eae*-γ6, were not detected in the SIWTP sewage samples. With more extensive sequencing efforts, these known *eae* subtypes and even new subtypes could be detected in the sanitary sewage.

Prevalence of *eae* subtypes in sanitary sewage. The prevalence of *eae* subtypes in sanitary sewage was determined by their overall detection frequencies (Fig. 3A) and their temporal variation in the monthly sewage samples (Fig. 3B). Six subtypes, including *eae*-β2 (41.2%), *eae*-θ (16.2%), *eae*-γ (9.1%), *eae*-α2 (8.5%), *eae*-ι1 (7.6%), and *eae*-ο (6.7%), were present in the clone libraries with a detection frequency of >5% and hence were considered to be the dominant subtypes in this study, while the remaining seven subtypes (*eae*-α1, *eae*-γ3, *eae*-κ, *eae*-ε, *eae*-ε8, *eae*-ζ, *eae*-ι2, and *eae*-ρ) were considered to be rare subtypes (i.e., detection frequency of <5%).

The overall detection frequencies of the *eae* subtypes (Fig. 3A) were further analyzed in the context of their temporal variation (Fig. 3B). The most prevalent subtype was *eae*-β2, which was present in six out of seven clone libraries (0.31% to 17.1%) and was represented by more than 50% of the clones in three clone libraries (August, October, and November 2010). The second most frequently detected subtype was *eae*-α2, which was detected in four of the seven clone libraries with a significantly smaller relative abundance (1.2% to 4.2%) than that of *eae*-β2. The remaining subtypes, including both the dominant and the rare ones, were detected at much lower frequencies, which indicated a strong temporal fluctuation of these serotypes in municipal sewage. In terms of the *eae* genetic diversity (i.e., the total number of different *eae* subtypes detected) in each month, the sewage samples collected in August 2010, March 2011, and April 2011 contained eight, seven, and four different *eae* subtypes, respectively, while samples from all of the other months contained just two *eae* gene subtypes each, further indicating the temporal component of *eae* genetic diversity in sanitary sewage.

The *eae*-β and *eae*-γ subtypes were the two most frequently detected subtypes in clinical isolates associated with human diarrheal diseases (17, 20, 38, 39). For example, Zhang et al. (20) found *eae*-β and *eae*-γ to be present in 34.2% and 31.5% of the EPEC strains from patients in Germany, respectively (20), while Blanco et al. (17) reported *eae*-β and *eae*-γ to be present in 28.6% and 38.6% of EHEC isolates from patients in Spain (17). However, in the sanitary sewage samples in our study, the prevalence of *eae*-γ was only 8.3%, and *eae*-β was not detected at all in the sanitary sewage samples. The low prevalence of *eae*-γ in sanitary sewage could be attributed to the low prevalence of EHEC, as observed by qPCR indicating the infrequency and low abundances of stx_1 and stx_2 genes (Fig. 1), because *eae*-γ is more likely to be associated with EHEC strains such as O157:H7 and O55:H7 (37, 40). The lack of *eae*-β in the SIWTP sanitary sewage was surprising, given that *eae*-β was frequently detected in both diarrhea-related clinical isolates (17, 20, 38, 39) and in environmental samples (35, 36). Instead, *eae*-β2 was the most dominant *eae* subtype in the SIWTP sanitary sewage. The different *eae* subtype prevalences in clinical isolates and in sanitary sewage were further illustrated by the observation that four of the five major *eae* subtypes detected in the sanitary sewage were infrequently detected in clinical *E. coli* isolates, including *eae*-β2 (17, 37, 41), *eae*-ι1 (20), *eae*-ο (42), *eae*-α2 (18, 19), and *eae*-θ (43).

The *eae* subtypes in sewage and environmental isolates. The 69 *eae*-positive *E. coli* isolates (62 from sanitary sewage and 7 from the Manoa watershed) were first screened using a PCR-RFLP procedure (data not shown), followed by sequencing to determine the *eae* subtypes. The *eae* subtypes from the sewage and environmental *E. coli* isolates were then compared with those detected by *eae* cloning and the known subtypes in the GenBank database by constructing a phylogenetic tree (Fig. 4). Nearly all of the sewage *E. coli* isolates (61/62) exhibited the same RFLP as that of *eae*-κ, which was subsequently confirmed by sequencing of the *eae* gene in the representative strain C9. The other sewage *E. coli* isolate, A7, contained the *eae*-θ gene. Two isolates from the Manoa watershed exhibited the same RFLP as *eae*-β, which was confirmed by sequencing the *eae* gene in the representative strain H7. The other five isolates from the Manoa watershed showed the same RFLP as *eae*-β2, which was also confirmed by sequencing the *eae* gene in the representative strain H8.

The sewage *eae*-positive *E. coli* isolates contained only two *eae* subtypes, κ and θ, with *eae*-κ being more dominant (61/62). Cloning methods indicated that *eae*-θ was a dominant subtype, while *eae*-κ was a rare subtype (Fig. 3). This discrepancy could be attributable to the temporal variation as the culture-dependent and culture-independent approaches were conducted on sewage samples collected in different years. However, given the large variation in *eae* detection frequencies among the sewage *E. coli* isolates (Table 2), a more probable explanation would be that the culture-based approach, due to its limited sample size, introduced significant bias in examining the *eae* diversity in sanitary sewage. Although only a very small number of *eae*-positive environmental *E. coli* isolates were examined, the detection of *eae*-β and *eae*-β2 was interesting, given the prevalence of *eae*-β2 in sanitary sewage (Fig. 3) and the frequent detection of *eae*-β in wildlife feces (19, 44) and in environmental waters (36).

ACKNOWLEDGMENTS

We acknowledge Ken Tenno of the Water Quality Laboratory, Department of Environmental Services of the City and County of Honolulu, for collecting sanitary sewage samples. We also thank Yong Li for preparing *E. coli* O157:H7 cells as qPCR calibration standards.

This material is based upon work partially supported by the National Science Foundation under grant number 0964260 and by the U.S. Environmental Protection Agency under grant number R834871 to T.Y.

REFERENCES

1. **Kaper JB, Nataro JP, Mobley HL.** 2004. Pathogenic *Escherichia coli*. Nat. Rev. Microbiol. **2:**123–140. http://dx.doi.org/10.1038/nrmicro818.
2. **Nataro JP, Kaper JB.** 1998. Diarrheagenic *Escherichia coli*. Clin. Microbiol. Rev. **11:**142–201.
3. **Muniesa M, Jofre J, Garcia-Aljaro C, Blanch AR.** 2006. Occurrence of *Escherichia coli* O157:H7 and other enterohemorrhagic *Escherichia coli* in the environment. Environ. Sci. Technol. **40:**7141. http://dx.doi.org/10.1021/es060927k.
4. **Eppinger M, Mammel MK, Leclerc JE, Ravel J, Cebula TA.** 2011. Genomic anatomy of *Escherichia coli* O157:H7 outbreaks. Proc. Natl. Acad. Sci. U. S. A. **108:**20142–20147. http://dx.doi.org/10.1073/pnas.1107176108.
5. **Mead PS, Griffin PM.** 1998. *Escherichia coli* O157:H7. Lancet **352:**1207–1212. http://dx.doi.org/10.1016/S0140-6736(98)01267-7.
6. **Mead PS, Slutsker L, Dietz V, McCaig LF, Bresee JS, Shapiro C, Griffin PM, Tauxe RV.** 1999. Food-related illness and death in the United States. Emerg. Infect. Dis. **5:**607. http://dx.doi.org/10.3201/eid0505.990502.
7. **Kupferschmidt K.** 2011. Infectious diseases. As *E. coli* outbreak recedes, new questions come to the fore. Science **333:**27. http://dx.doi.org/10.1126/science.333.6038.27.
8. **Dean-Nystrom EA, Bosworth BT, Moon HW, O'Brien AD.** 1998. *Escherichia coli* O157:H7 requires intimin for enteropathogenicity in calves. Infect. Immun. **66:**4560–4563.
9. **Donnenberg MS, Tzipori S, McKee ML, O'Brien AD, Alroy J, Kaper JB.** 1993. The role of the eae gene of enterohemorrhagic *Escherichia coli* in intimate attachment in vitro and in a porcine model. J. Clin. Invest. **92:**1418–1424. http://dx.doi.org/10.1172/JCI116718.
10. **McDaniel TK, Jarvis KG, Donnenberg MS, Kaper JB.** 1995. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. Proc. Natl. Acad. Sci. U. S. A. **92:**1664–1668. http://dx.doi.org/10.1073/pnas.92.5.1664.
11. **Wales AD, Pearson GR, Skuse AM, Roe JM, Hayes CM, Cookson AL, Woodward MJ.** 2001. Attaching and effacing lesions caused by *Escherichia coli* O157:H7 in experimentally inoculated neonatal lambs. J. Med. Microbiol. **50:**752–758.
12. **Donnenberg MS, Tacket CO, Losonsky G, Frankel G, Nataro JP, Dougan G, Levine MM.** 1998. Effect of prior experimental human enteropathogenic *Escherichia coli* infection on illness following homologous and heterologous rechallenge. Infect. Immun. **66:**52–58.
13. **Jerse AE, Yu J, Tall BD, Kaper JB.** 1990. A genetic locus of enteropathogenic *Escherichia coli* necessary for the production of attaching and effacing lesions on tissue culture cells. Proc. Natl. Acad. Sci. U. S. A. **87:**7839–7843. http://dx.doi.org/10.1073/pnas.87.20.7839.
14. **Bielaszewska M, Mellmann A, Zhang W, Kock R, Fruth A, Bauwens A, Peters G, Karch H.** 2011. Characterisation of the *Escherichia coli* strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. Lancet Infect. Dis. **11:**671–676. http://dx.doi.org/10.1016/S1473-3099(11)70165-7.
15. **Brzuszkiewicz E, Thurmer A, Schuldes J, Leimbach A, Liesegang H, Meyer FD, Boelter J, Petersen H, Gottschalk G, Daniel R.** 2011. Genome sequence analyses of two isolates from the recent *Escherichia coli* outbreak in Germany reveal the emergence of a new pathotype: entero-aggregative-haemorrhagic *Escherichia coli* (EAHEC). Arch. Microbiol. **193:**883–891. http://dx.doi.org/10.1007/s00203-011-0725-6.
16. **Reid SD, Herbelin CJ, Bumbaugh AC, Selander RK, Whittam TS.** 2000. Parallel evolution of virulence in pathogenic *Escherichia coli*. Nature **406:**64–67. http://dx.doi.org/10.1038/35017546.
17. **Blanco JE, Blanco M, Alonso MP, Mora A, Dahbi G, Coira MA, Blanco J.** 2004. Serotypes, virulence genes, and intimin types of Shiga toxin (verotoxin)-producing *Escherichia coli* isolates from human patients: prevalence in Lugo, Spain, from 1992 through 1999. J. Clin. Microbiol. **42:**311–319. http://dx.doi.org/10.1128/JCM.42.1.311-319.2004.
18. **Lacher DW, Steinsland H, Whittam TS.** 2006. Allelic subtyping of the intimin locus (eae) of pathogenic *Escherichia coli* by fluorescent RFLP. FEMS Microbiol. Lett. **261:**80–87. http://dx.doi.org/10.1111/j.1574-6968.2006.00328.x.
19. **Ramachandran V, Brett K, Hornitzky MA, Dowton M, Bettelheim KA, Walker MJ, Djordjevic SP.** 2003. Distribution of intimin subtypes among *Escherichia coli* isolates from ruminant and human sources. J. Clin. Microbiol. **41:**5022–5032. http://dx.doi.org/10.1128/JCM.41.11.5022-5032.2003.
20. **Zhang WL, Kohler B, Oswald E, Beutin L, Karch H, Morabito S, Caprioli A, Suerbaum S, Schmidt H.** 2002. Genetic diversity of intimin genes of attaching and effacing *Escherichia coli* strains. J. Clin. Microbiol. **40:**4486–4492. http://dx.doi.org/10.1128/JCM.40.12.4486-4492.2002.
21. **McGraw EA, Li J, Selander RK, Whittam TS.** 1999. Molecular evolution and mosaic structure of alpha, beta, and gamma intimins of pathogenic *Escherichia coli*. Mol. Biol. Evol. **16:**12–22. http://dx.doi.org/10.1093/oxfordjournals.molbev.a026032.
22. **Blanco M, Schumacher S, Tasara T, Zweifel C, Blanco JE, Dahbi G, Blanco J, Stephan R.** 2005. Serotypes, intimin variants and other virulence factors of *eae* positive *Escherichia coli* strains isolated from healthy cattle in Switzerland. Identification of a new intimin variant gene (*eae-eta2*). BMC Microbiol. **5:**23. http://dx.doi.org/10.1186/1471-2180-5-23.
23. **Frankel G, Candy DC, Everest P, Dougan G.** 1994. Characterization of the C-terminal domains of intimin-like proteins of enteropathogenic and enterohemorrhagic *Escherichia coli*, *Citrobacter freundii*, and *Hafnia alvei*. Infect. Immun. **62:**1835–1842.
24. **Oswald E, Schmidt H, Morabito S, Karch H, Marches O, Caprioli A.** 2000. Typing of intimin genes in human and animal enterohemorrhagic and enteropathogenic *Escherichia coli*: characterization of a new intimin variant. Infect. Immun. **68:**64–71. http://dx.doi.org/10.1128/IAI.68.1.64-71.2000.
25. **Luo Y, Frey EA, Pfuetzner RA, Creagh AL, Knoechel DG, Haynes CA, Finlay BB, Strynadka NC.** 2000. Crystal structure of enteropathogenic *Escherichia coli* intimin-receptor complex. Nature **405:**1073–1077. http://dx.doi.org/10.1038/35016618.
26. **Hartland EL, Batchelor M, Delahay RM, Hale C, Matthews S, Dougan G, Knutton S, Connerton I, Frankel G.** 1999. Binding of intimin from enteropathogenic *Escherichia coli* to Tir and to host cells. Mol. Microbiol. **32:**151–158. http://dx.doi.org/10.1046/j.1365-2958.1999.01338.x.
27. **Berger M, Shiau R, Weintraub JM.** 2006. Review of syndromic surveillance: implications for waterborne disease detection. J. Epidemiol. Community Health **60:**543–550. http://dx.doi.org/10.1136/jech.2005.038539.
28. **Bradley CA, Rolka H, Walker D, Loonsk J.** 2005. BioSense: implementation of a national early event detection and situational awareness system. MMWR Morb. Mortal. Wkly. Rep. **54**(Suppl)**:**11–19.
29. **Government Accountability Office.** 2004. Emerging infectious diseases: review of state and federal disease surveillance efforts. Document GAO-04-877. U. S. Government Printing Office, Washington, DC. http://www.gpo.gov/fdsys/pkg/GAOREPORTS-GAO-04-877/html/GAOREPORTS-GAO-04-877.htm.
30. **U. S. Environmental Protection Agency.** 2002. Method 1603: *Escherichia coli* (*E. coli*) in water by membrane filtration using modified membrane-thermotolerant *Escherichia coli* agar (modified mTEC). U.S. Environmental Protection Agency, Office of Water, Washington, DC.
31. **Goto DK, Yan T.** 2011. Genotypic diversity of *Escherichia coli* in the water and soil of tropical watersheds in Hawai'i. Appl. Environ. Microbiol. **77:**3988–3997. http://dx.doi.org/10.1128/AEM.02140-10.
32. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28:**2731–2739. http://dx.doi.org/10.1093/molbev/msr121.
33. **Frahm E, Obst U.** 2003. Application of the fluorogenic probe technique (TaqMan PCR) to the detection of *Enterococcus* spp. and *Escherichia coli* in water samples. J. Microbiol. Methods **52:**123–131. http://dx.doi.org/10.1016/S0167-7012(02)00150-1.
34. **Ishii S, Hansen DL, Hicks RE, Sadowsky MJ.** 2007. Beach sand and sediments are temporal sinks and sources of *Escherichia coli* in Lake Superior. Environ. Sci. Technol. **41:**2203–2209. http://dx.doi.org/10.1021/es0623156.
35. **Mora A, Lopez C, Dhabi G, Lopez-Beceiro AM, Fidalgo LE, Diaz EA, Martinez-Carrasco C, Mamani R, Herrera A, Blanco JE, Blanco M, Blanco J.** 2012. Seropathotypes, phylogroups, Stx subtypes, and intimin types of wildlife-carried, Shiga toxin-producing *Escherichia coli* strains with the same characteristics as human-pathogenic isolates. Appl. Environ. Microbiol. **78:**2578–2585. http://dx.doi.org/10.1128/AEM.07520-11.
36. **Hamilton MJ, Hadi AZ, Griffith JF, Ishii S, Sadowsky MJ.** 2010. Large scale analysis of virulence genes in *Escherichia coli* strains isolated from Avalon Bay, CA. Water Res. **44:**5463–5473. http://dx.doi.org/10.1016/j.watres.2010.06.058.
37. **Adu-Bobie J, Frankel G, Bain C, Goncalves AG, Trabulsi LR, Douce G, Knutton S, Dougan G.** 1998. Detection of intimins α, β, γ, and δ, four

intimin derivatives expressed by attaching and effacing microbial pathogens. J. Clin. Microbiol. **36:**662–668.
38. **Blanco M, Blanco JE, Mora A, Rey J, Alonso JM, Hermoso M, Hermoso J, Alonso MP, Dahbi G, Gonzalez EA, Bernardez MI, Blanco J.** 2003. Serotypes, virulence genes, and intimin types of Shiga toxin (verotoxin)-producing *Escherichia coli* isolates from healthy sheep in Spain. J. Clin. Microbiol. **41:**1351–1356. http://dx.doi.org/10.1128/JCM.41.4.1351-1356.2003.
39. **Jenkins C, Lawson AJ, Cheasty T, Willshaw GA, Wright P, Dougan G, Frankel G, Smith HR.** 2003. Subtyping intimin genes from enteropathogenic *Escherichia coli* associated with outbreaks and sporadic cases in the United Kingdom and Eire. Mol. Cell. Probes **17:**149–156. http://dx.doi.org/10.1016/S0890-8508(03)00046-X.
40. **Iguchi A, Ooka T, Ogura Y, Asadulghani, Nakayama K, Frankel G, Hayashi T.** 2008. Genomic comparison of the O-antigen biosynthesis gene clusters of *Escherichia coli* O55 strains belonging to three distinct lineages. Microbiology **154:**559–570. http://dx.doi.org/10.1099/mic.0.2007/013334-0.
41. **Blanco M, Blanco JE, Dahbi G, Mora A, Alonso MP, Varela G, Gadea MP, Schelotto F, Gonzalez EA, Blanco J.** 2006. Typing of intimin (*eae*) genes from enteropathogenic *Escherichia coli* (EPEC) isolated from children with diarrhoea in Montevideo, Uruguay: identification of two novel intimin variants (μB and ξR/β2B). J. Med. Microbiol. **55:**1165–1174. http://dx.doi.org/10.1099/jmm.0.46518-0.
42. **Hyma KE, Lacher DW, Nelson AM, Bumbaugh AC, Janda JM, Strockbine NA, Young VB, Whittam TS.** 2005. Evolutionary genetics of a new pathogenic *Escherichia* species: *Escherichia albertii* and related *Shigella boydii* strains. J. Bacteriol. **187:**619–628. http://dx.doi.org/10.1128/JB.187.2.619-628.2005.
43. **Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, Kodama T, Abe H, Nakayama K, Kurokawa K, Tobe T, Hattori M, Hayashi T.** 2009. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. **106:**17939–17944. http://dx.doi.org/10.1073/pnas.0903585106.
44. **Cookson AL, Bennett J, Thomson-Carter F, Attwood GT.** 2007. Intimin subtyping of *Escherichia coli*: concomitant carriage of multiple intimin subtypes from forage-fed cattle and sheep. FEMS Microbiol. Lett. **272:**163–171. http://dx.doi.org/10.1111/j.1574-6968.2007.00755.x.
45. **Kagkli DM, Weber TP, Van den Bulcke M, Folloni S, Tozzoli R, Morabito S, Ermolli M, Gribaldo L, Van den Eede G.** 2011. Application of the modular approach to an in-house validation study of real-time PCR methods for the detection and serogroup determination of verocytotoxigenic *Escherichia coli*. Appl. Environ. Microbiol. **77:**6954–6963. http://dx.doi.org/10.1128/AEM.05357-11.
46. **Perelle S, Dilasser F, Grout J, Fach P.** 2004. Detection by 5′-nuclease PCR of Shiga-toxin producing *Escherichia coli* O26, O55, O91, O103, O111, O113, O145 and O157:H7, associated with the world's most frequent clinical cases. Mol. Cell. Probes **18:**185–192. http://dx.doi.org/10.1016/j.mcp.2003.12.004.
47. **Nielsen EM, Andersen MT.** 2003. Detection and characterization of verocytotoxin-producing *Escherichia coli* by automated 5′ nuclease PCR assay. J. Clin. Microbiol. **41:**2884–2893. http://dx.doi.org/10.1128/JCM.41.7.2884-2893.2003.
48. **Schlaman HR, Risseeuw E, Franke-van Dijk ME, Hooykaas PJ.** 1994. Nucleotide sequence corrections of the *uidA* open reading frame encoding beta-glucuronidase. Gene **138:**259–260. http://dx.doi.org/10.1016/0378-1119(94)90820-6.
49. **Cebula TA, Payne WL, Feng P.** 1995. Simultaneous identification of strains of *Escherichia coli* serotype O157:H7 and their Shiga-like toxin type by mismatch amplification mutation assay-multiplex PCR. J. Clin. Microbiol. **33:**248–250.
50. **Stacy-Phipps S, Mecca JJ, Weiss JB.** 1995. Multiplex PCR assay and simple preparation method for stool specimens detect enterotoxigenic *Escherichia coli* DNA during course of infection. J. Clin. Microbiol. **33:**1054–1059.
51. **Gartner JF, Schmidt MA.** 2004. Comparative analysis of locus of enterocyte effacement pathogenicity islands of atypical enteropathogenic *Escherichia coli*. Infect. Immun. **72:**6722–6728. http://dx.doi.org/10.1128/IAI.72.11.6722-6728.2004.
52. **Iguchi A, Thomson NR, Ogura Y, Saunders D, Ooka T, Henderson IR, Harris D, Asadulghani M, Kurokawa K, Dean P, Kenny B, Quail MA, Thurston S, Dougan G, Hayashi T, Parkhill J, Frankel G.** 2009. Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. J. Bacteriol. **191:**347–354. http://dx.doi.org/10.1128/JB.01238-08.
53. **Kyle JL, Cummings CA, Parker CT, Quinones B, Vatta P, Newton E, Huynh S, Swimley M, Degoricija L, Barker M, Fontanoz S, Nguyen K, Patel R, Fang R, Tebbs R, Petrauskene O, Furtado M, Mandrell RE.** 2012. *Escherichia coli* serotype O55:H7 diversity supports parallel acquisition of bacteriophage at Shiga toxin phage insertion sites during evolution of the O157:H7 lineage. J. Bacteriol. **194:**1885–1896. http://dx.doi.org/10.1128/JB.00120-12.
54. **Ooka T, Seto K, Kawano K, Kobayashi H, Etoh Y, Ichihara S, Kaneko A, Isobe J, Yamaguchi K, Horikawa K, Gomes TA, Linden A, Bardiau M, Mainil JG, Beutin L, Ogura Y, Hayashi T.** 2012. Clinical significance of *Escherichia albertii*. Emerg. Infect. Dis. **18:**488–492. http://dx.doi.org/10.3201/eid1803.111401.