Genetic Differentiation of *Escherichia coli* O157:H7 Clades Associated with Human Disease by Real-Time PCR[∇]

James T. Riordan, Sandeep B. Viswanath, Shannon D. Manning, and Thomas S. Whittam*

Microbial Evolution Laboratory, National Food Safety and Toxicology Center, Michigan State University, East Lansing, Michigan 48824

The rapid and accurate identification of *Escherichia coli* O157:H7 strains is central to reducing the impact of outbreaks. A real-time PCR-based approach to differentiating major outbreak lineages of O157 with novel single-nucleotide polymorphisms is described. The utility of this method is in detection of hypervirulent strains in cases of clinical disease.

Escherichia coli O157:H7 is a gastrointestinal pathogen associated with severe cases of hemorrhagic colitis and life-threatening sequelae, such as hemolytic-uremic syndrome (HUS) (5, 9). Variations in disease severity among O157 outbreaks are evident, as measured by the frequencies of hospitalization and HUS. For example, the 1993 multistate outbreak in North America (2) and the large 1996 outbreak in Japan (10) had low rates of hospitalization and HUS (4, 7). By comparison, two recent outbreaks in the United States, caused by contaminated spinach and lettuce, resulted in much higher frequencies of both hospitalization (average, 63%) and HUS (average, 13%) (1). The strains that caused the latter outbreak, recently characterized for up to 96 single-nucleotide polymorphisms (SNPs), were part of a distinct *E. coli* O157 lineage (clade 8) that was associated with HUS, that has increased in frequency, and that contains a distinct combination of Shiga toxin genes (8). Consequently, it was suggested that a more virulent O157 subpopulation or genetic clade has emerged (8). Such findings emphasize the importance of the rapid identification of emergent O157 clades in an effort to reduce the disease burden associated with future outbreaks.

In this report we describe real-time PCR (RT-PCR) assays that are based on hairpin (HP) primers targeting specific SNPs and that differentiate the four predominant O157 clades described previously (8). More importantly, this method rapidly (within 24 h) and accurately differentiates O157 strains belonging to clade 8.

HP primers were constructed for SNP targets (Table 1) in four open reading frames (ORFs), as described previously (6): ECs2357 (hypothetical protein), ECs2521 (*para*-aminobenzoate synthase), ECs3881 (*hybA*, hydrogenase 2), and ECs4130 (*panF*, sodium/panthothenate symporter). Two SNPs involved nonsynonymous polymorphisms and two involved synonymous substitutions (Table 1). Three primers were designed for each SNP: two HP primers, each of which was specific for either the wild type (reference) SNP or the mutant (diagnostic) SNP, and a conserved non-HP primer (Table 2). The HP primers were linked at the 5′ ends to form 6- to 10-bp stem structures with the corresponding 3′ end. The 3′-end terminal base of the HP primers was complementary to either the reference SNP or the diagnostic SNP. The stem structures for the HP primers were designed to have melting temperatures between 65°C and 69°C to enhance priming specificity. These primer sets were used to differentiate O157 clades 1 to 3 and 8, each of which has a distinct SNP profile (Table 3). The O157 clades were previously constructed on the basis of SNP genotyping data for 538 O157 strains (8).

All primer combinations specific for the four SNPs were initially tested with at least three strains from each of clades 1 to 3 and 8. In addition, five strains representing other clades (clades 4 to 7 and 9) were examined with one non-O157 *E. coli* strain (negative control) (Table 3). The prototypical strain for each clade included strains Sakai (clade 1), 93-111 (clade

TABLE 1. Four genes with SNPs

Gene (ORF)	SNP target	Strain Sakai nucleotide	Amino acid position	Codon polymorphism[a]	Nonsynonymous or synonymous[b]	Function
ECs2357	539	C	180	T c/a C	NON	Hypothetical protein
ECs2521	1060	T	354	g/t CC	NON	*p*-Aminobenzoate synthetase component I
ECs3881	438	T	146	A C t/c	SYN	Hydrogenase-2 small subunit
ECs4130	630	T	210	G C t/c	SYN	Sodium/pantothenate symporter

[a] Lowercase letters indicate positions of reference versus diagnostic amino acids in the SNP codon.
[b] NON, nonsynonymous; SYN, synonymous.

* Corresponding author. Mailing address: National Food Safety & Toxicology Center, 165 Food Safety & Toxicology Building, Michigan State University, East Lansing, MI 48824. Phone: (517) 884-2033. Fax: (517) 432-2310. E-mail: whittam@msu.edu.

TABLE 2. Hairpin primers and SNP targets

ID[a] and primer	Target [clade(s)][b]	Sequence (5′→3′)[c]
ECs2357		
539C-RHP	R (1–7, 9)	<u>CCGAGC</u>GTTTTCCAGTG<u>GC</u> <u>TCGG</u>
539A-RHP	D (8)	<u>ACGAGC</u>GTTTTCCAGTG<u>GC</u> <u>TCGT</u>
539-F	Common	GAATCTGCAGGCCAAAATTTC
ECs2521		
1060G-FHP	R (2–9)	<u>CGTGTAACTGCGCAACTGCCA</u> <u>GAACAGTTACACG</u>
1060T-FHP	D (1)	<u>AGTGTAACTGCGCAACTGCCA</u> <u>GAACAGTTACACT</u>
1060-R	Common	TTCGGAGCCCCGGTTATT
ECs3881		
438C-FHP	R (3–9)	<u>GGTGCACATTACGACTAAGA</u> <u>CGTGTGCACC</u>
438T-FHP	D (1, 2)	<u>AGTGCACATTACGACTAAGAC</u> <u>GTGTGCACT</u>
438-R	Common	GGACAGGCGACCATGCAG
ECs4130		
630C-RHP	R (4–9)	<u>CGGCTTAATC</u>TGTACTGCGTT <u>GATTAAGCCG</u>
630T-RHP	D (1, 2, 3)	<u>TGGCTTAATC</u>TGTACTGCGTT <u>GATTAAGCCA</u>
630-F	Common	GGCACCGTTGTGCTGCTTAT

[a] The locus identifier (ID) is based on the published genome of *E. coli* O157:H7 Sakai (GenBank accession no. BA000007).

[b] The target of the HP primer is for the reference (R) or the diagnostic (D) sequence; the clade(s) for which the HP primers are specific is given in parentheses. Common primers are nonspecific.

[c] Primer sequences are based on the sequence of *E. coli* O157:H7 Sakai; HP regions are underlined, and SNPs are in boldface.

2), EDL933 (clade 3), and Spinach (strain MI06-63; clade 8), whereas *E. coli* K-12 (strain MG1655) was used as the negative control (Table 3). The RT-PCR assays were performed with DNA (Puregene DNA isolation kits; Gentra, Minneapolis, MN) and colony picks from sorbitol MacConkey (SMAC) agar to enhance the usefulness of the assays in the clinical labora-

tory setting. The method was validated by tests with a blinded set of 62 strains representing 20 strains from clades 2, 3, and 8 and two non-O157 strains. Strains from clade 1 were not included in the blinded study, as there were too few such strains in the collection.

HP-primed RT-PCR was performed with an iQ5 system (Bio-Rad, Hercules, CA) by using previously described cycling conditions (6). The reaction mixtures contained 1× SYBR green supermix (Bio-Rad) and 0.25 μM each HP primer and common primer in a 25-μl final volume. The experiments used 1 ng total DNA per reaction mixture or cell lysates from colony picks. To prepare cell lysates, cultures were grown 18 to 20 h at 37°C on SMAC, one colony was transferred into 50 μl sterile water and vortexed, and 5 μl was added to each reaction mixture. The assay results were analyzed with the iQ5 optical systems software (version 2; Bio-Rad).

Critical threshold (C_T) values were transformed to cycle threshold differences (ΔC_T) by using the equation $C_{T\text{-}R} - C_{T\text{-}D}$, where $C_{T\text{-}R}$ is the C_T value for the reference HP primers and $C_{T\text{-}D}$ is the C_T value for the diagnostic HP primers. ΔC_T values reflect the difference in the amplification efficiencies of matched HP primers versus those of mismatched HP primers at each SNP locus of a target DNA sequence, with larger differences representing more robust reactions (6). For all experiments, an average ΔC_T value of <0 denoted specificity of the reaction for the reference SNP, and an average ΔC_T value of >0 denoted specificity for the diagnostic SNP. At least three biological replicates were used for the development of the assay, whereas two replicates were used for the validation experiment. Significant differences in average ΔC_T values among strains of distinct clades for each SNP target were inferred by using a generalized linear model, (α = 0.01), and posteriori multiple contrasts were performed by Fisher's least-significant-difference test (α = 0.05) for comparison of treatment group means.

In the assay development stage, RT-PCR with HP primers for SNPs 539, 1060, 438, and 630 was diagnostic for all 21 strains representing the nine O157 phylogenetic clades identified previously (8). Strains of clades 8 and 1 could be distin-

TABLE 3. SNP profiles of strains

Clade[a] or strain	SNP profile[b]				Strain(s)[c]
	539 (C/A)	1060 (G/C)	438 (C/T)	630 (C/T)	
1	C	C*	T*	T*	Sakai (TW08264), EK4 (TW08612), MLVA-24 (TW10022)
2	C	G	T*	T*	93–111 (TW04863), MLVA-47 (TW10045), MI04-43 (TW11110), 96M1006 (TW11308)
3	C	G	C	T*	EDL932 (TW02299), EDL933 (TW02302), MLVA-10 (TW10008), MLVA-26 (TW10024)
4	C	G	C	C	MI02–53 (TW11039)
5	C	G	C	C	MI02–99 (TW11116)
6	C	G	C	C	MI03–48 (TW09109)
7	C	G	C	C	MI03–9 (TW09178)
8	A*	G	C	C	E32511 (TW02883), MI06-31 (TW14313), MI06-63 (TW14359), 64645 (TW14585), 64086 (TW14586)
9	C	G	C	C	2664/91 (TW07763)
K-12	C	G	C	C	MG1655

[a] Based on the work of Manning et al. (8).

[b] The SNP profile associated with the respective clade. The SNP is described in terms of the nucleotide position within the ORF, and the polymorphism is given in parentheses. Asterisks denote diagnostic SNPs.

[c] All strains were acquired from the collection of the STEC Center at Michigan State University.
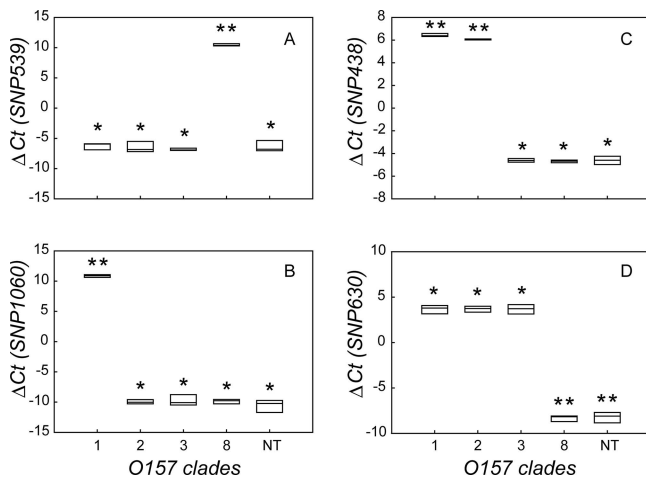
FIG. 1. Results for optimization of rapid RT-PCR method. Box plots of mean ΔC_T values for four SNPs (A to D) as a function of O157 clade are shown, as follows: SNP 539 (A), SNP 1060 (B), SNP 438 (C), and SNP 630 (D). Plot boundaries represent the 25th and 75th percentiles; the median is given by the line. The number of asterisks above each plot denotes statistical difference among clades ($P < 0.05$).
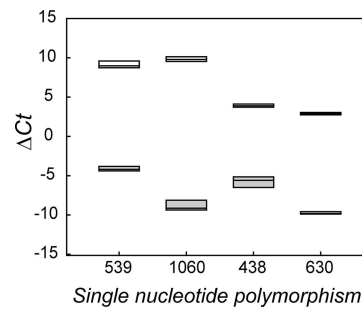

Single nucleotide polymorphism

FIG. 2. Box plot of mean ΔC_T as a function of SNP for colony picks from SMAC. Open plots represent the mean ΔC_T values for positive control strains for each SNP, including strains Spinach (SNP 539), Sakai (SNP 1060), 93-111 (SNP 438), and EDL933 (SNP 630); filled plots represent the mean ΔC_T values for the negative control strain, strain MG1655.

guished from strains of all other clades with HP primers for SNPs 539 and 1060, respectively ($P < 0.001$) (Fig. 1A and B). The average SNP 539 ΔC_T was 10.45 ± 0.20, and that for all other strains ($n = 16$) was −6.49 ± 0.75. Likewise, for clade 1 strains ($n = 3$), the average SNP 1060 ΔC_T was 10.86 ± 0.23, and for all other strains ($n = 19$) it was −10.08 ± 0.89. Strains from clades 2 and 3 were further differentiated with HP primers for SNP 438 and 630 by a process of elimination ($P < 0.001$) (Fig. 1C and D). The average SNP 438 and 630 ΔC_T values for clade 2 strains ($n = 5$) were 6.06 ± 0.05 and 3.68 ± 0.38, respectively, and for all other strains except clade 1 strains ($n = 14$), the ΔC_T values averaged −4.64 ± 0.24. For clade 3 strains ($n = 4$), the average SNP 438 and 630 ΔC_T values were −4.61 ± 0.17 and 3.68 ± 0.54, respectively. As predicted, all strains from clades 4 to 7 and 9 ($n = 5$) had ΔC_T values of <0 for all SNPs (Fig. 1A to D).

Similar results were obtained when SMAC colony picks of the four prototype strains were evaluated. Prototype strains representing clades 1 to 3 and 8 were differentiated from negative control strain MG1655 for all SNPs ($P < 0.001$) (Fig. 2). The average SNP 539 ΔC_T values were 9.10 ± 0.47 for strain Spinach (clade 8) and −4.13 ± 0.29 for strain MG1655. For SNP 1060, the average ΔC_T values were 9.82 ± 0.31 for strain Sakai (clade 1) and −8.87 ± 0.63 for MG1655. For SNP 438 and SNP 630, the average ΔC_T values were 3.91 ± 0.21 and 2.89 ± 0.15, respectively, for strain 93-111 (clade 2) and strain EDL933 (clade 3) and −5.74 ± 0.68 and −9.76 ± 0.16, respectively, for MG1655.

In the RT-PCR assay validation stage, colony picks from a blinded sample ($n = 62$) of strains representing clades 2, 3 and 8 were evaluated. A total of 50 (81%) of the 62 strains were correctly assigned to their respective clades. Within this proportion, all 20 clade 8 strains were correctly identified, as were 18 of 20 clade 3 strains and both non-O157 negative control strains. The average SNP 538 ΔC_T was 10.24 ± 0.41 for clade 8 strains, and the average SNP 630 ΔC_T was 2.84 ± 0.59 for

clade 3 strains. Only 10/20 clade 2 strains were properly identified; the remaining strains were identified as members of the adjacent clade 1 group, for which there were no representative strains in the blinded study. For strains identified as clade 1, the average SNP 1060 ΔC_T was 7.69 ± 1.29, whereas for the remaining clade 2 strains, the average SNP 1060 ΔC_T was −7.01 ± 0.87. As suspected, a review of our comparative genomics sequencing data revealed that SNP 1060 for this subset of strains identified as clade 1 contained the T polymorphism in place of G. Interestingly, of the 192 clade 2 strains whose SNP genotypes were determined in this study, the 10 identified as members of clade 1 represented a minority within this sample (18.8%).

In conclusion, rapid differentiation of four novel SNPs by RT-PCR methods can detect major outbreak strains and clades of *E. coli* O157:H7 from SMAC differential medium within 24 h. In particular, clade 8 strains, representative of the 2006 spinach outbreak lineage, can be differentiated from all other clades in a single reaction. Such rapid detection of strains of distinct clades, such as those in clade 8, represents a useful method to detect emergent lineages that are associated with more severe diseases. This assay could therefore be used as a means of detecting hypervirulent O157 strains in a preventative, risk assessment capacity. Clade 1 and clade 2 strains are very closely related and could not be consistently identified by the SNP 1060 assay. Use of the HP primer technology reduces the cost normally equated with the use of the specific primer-probe chemistries without jeopardizing reaction accuracy. It will be insightful in future applications to compare the SNP application to other methods, such as pulsed-field gel electrophoresis and multilocus variable-number tandem repeat analysis (3, 11). It is possible that SNP typing will provide a means of rapid identification before full diagnostic analysis and that it may help to curb the impact of *E. coli* O157:H7 outbreaks.

REFERENCES

1. **CDC.** 2006. Ongoing multistate outbreak of Escherichia coli serotype O157:H7 infections associated with consumption of fresh spinach—United States, September 2006. MMWR Morb. Mortal. Wkly. Rep. **55:**1045–1046.

2. **CDC.** 1993. Update: multistate outbreak of *Escherichia coli* O157:H7 infections from hamburgers—western United States, 1992–1993. MMWR Morb. Mortal. Wkly. Rep. **42:**258–263.

3. **Cooley, M., D. Carychao, L. Crawford-Miksza, M. T. Jay, C. Myers, C. Rose, C. Keys, J. Farrar, and R. E. Mandrell.** 2007. Incidence and tracking of *Escherichia coli* O157:H7 in a major produce production region in California. PLoS ONE **2:**e1159.

4. **Fukushima, H., T. Hashizume, Y. Morita, J. Tanaka, K. Azuma, Y. Mizumoto, M. Kaneno, M. Matsuura, K. Konma, and T. Kitani.** 1999. Clinical experiences in Sakai City Hospital during the massive outbreak of enterohemorrhagic *Escherichia coli* O157 infections in Sakai City, 1996. Pediatr. Int. **41:**213–217.

5. **Griffin, P. M., and R. V. Tauxe.** 1991. The epidemiology of infections caused by *Escherichia coli* O157:H7, other enterohemorrhagic *E. coli*, and the associated hemolytic uremic syndrome. Epidemiol. Rev. **13:**60–98.

6. **Hazbon, M. H., and D. Alland.** 2004. Hairpin primers for simplified single-nucleotide polymorphism analysis of *Mycobacterium tuberculosis* and other organisms. J. Clin. Microbiol. **42:**1236–1242.

7. **Higami, S., K. Nishimoto, T. Kawamura, T. Tsuruhara, G. Isshiki, and A. Ookita.** 1998. Retrospective analysis of the relationship between HUS incidence and antibiotics among patients with *Escherichia coli* O157 enterocolitis in the Sakai outbreak. Kansenshogaku Zasshi **72:**266–272. (In Japanese.)

8. **Manning, S. D., A. S. Motiwala, A. C. Springman, W. Qi, D. W. Lacher, L. M. Ouellette, J. M. Mladonicky, P. Somsel, J. T. Rudrik, S. D. Dietrich, W. Zhang, D. Alland, and T. S. Whittam.** 2008. Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. Proc. Natl. Acad. Sci. USA **105:**4868–4873.

9. **Mead, P. S., and P. M. Griffin.** 1998. *Escherichia coli* O157:H7. Lancet **352:**1207–1212.

10. **Michino, H., K. Araki, S. Minami, S. Takaya, N. Sakai, M. Miyazaki, A. Ono, and H. Yanagawa.** 1999. Massive outbreak of *Escherichia coli* O157:H7 infection in schoolchildren in Sakai City, Japan, associated with consumption of white radish sprouts. Am. J. Epidemiol. **150:**787–796.

11. **Noller, A. C., M. C. McEllistrem, K. A. Shutt, and L. H. Harrison.** 2006. Locus-specific mutational events in a multilocus variable-number tandem repeat analysis of *Escherichia coli* O157:H7. J. Clin. Microbiol. **44:**374–377.