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

Published Ahead of Print 9 April 2008.

Updated information and services can be found at:
http://jcm.asm.org/content/46/6/2070

**REFERENCES**

This article cites 11 articles, 5 of which can be accessed free at:
http://jcm.asm.org/content/46/6/2070#ref-list-1

**CONTENT ALERTS**

Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), more»

Information about commercial reprint orders: http://journals.asm.org/site/misc/reprints.xhtml
To subscribe to another ASM Journal go to: http://journals.asm.org/site/subscriptions/
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

Received 1 February 2008/Returned for modification 10 March 2008/Accepted 26 March 2008

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 multisate 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/pantothenate 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 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 primer 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 2), and accurately differentiates O157 strains belonging to clade 8.

<table>
<thead>
<tr>
<th>Gene (ORF)</th>
<th>SNP target</th>
<th>Strain Sakai nucleotide</th>
<th>Amino acid position</th>
<th>Codon polymorphism</th>
<th>Nonsynonymous or synonymous</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECs2357</td>
<td>539</td>
<td>C</td>
<td>180</td>
<td>T c/a C</td>
<td>NON</td>
<td>Hypothetical protein</td>
</tr>
<tr>
<td>ECs2521</td>
<td>1060</td>
<td>T</td>
<td>354</td>
<td>g/t CC</td>
<td>NON</td>
<td><em>p</em>-Aminobenzoate synthase component I</td>
</tr>
<tr>
<td>ECs3881</td>
<td>438</td>
<td>T</td>
<td>146</td>
<td>A C t/c</td>
<td>SYN</td>
<td>Hydrogenase-2 small subunit</td>
</tr>
<tr>
<td>ECs4130</td>
<td>630</td>
<td>T</td>
<td>210</td>
<td>G C t/c</td>
<td>SYN</td>
<td>Sodium/pantothenate symporter</td>
</tr>
</tbody>
</table>

* 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.

Published ahead of print on 9 April 2008.
and colony picks from sorbitol MacConkey (SMAC) agar. DNA (Puregene DNA isolation kits; Gentra, Minneapolis, MN) was used as the negative control (Table 3). The RT-PCR assays were performed with E. coli O157:H7 Sakai (GenBank accession no. BA000007).

Critical threshold (C_T) values were transformed to cycle threshold differences (ΔC_T) by using the equation ΔC_T = C_T(R) − C_T(D), where C_T(R) is the C_T value for the reference HP primers and C_T(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 distinguished by using 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 ΔC_T = C_T(R) − C_T(D), where C_T(R) is the C_T value for the reference HP primers and C_T(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 distinguished by using the iQ5 optical systems software (version 2; Bio-Rad).
strains (0.38, respectively, and for all other strains except clade 1 from clades 2 and 3 were further differentiated with HP primer and for all other strains (n)

The number of asterisks above each plot denotes statistical difference among clades (P < 0.05).

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 \( \Delta C_T \) 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 \( \Delta C_T \) 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 \( \Delta C_T \) 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 \( \Delta C_T \) values averaged −4.64 ± 0.24. For clade 3 strains (n = 4), the average SNP 438 and 630 \( \Delta C_T \) 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 \( \Delta C_T \) 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 \( \Delta C_T \) values were 9.10 ± 0.47 for strain Spinach (clade 8) and −4.13 ± 0.29 for strain MG1655. For SNP 1060, the average \( \Delta C_T \) values were 9.82 ± 0.31 for strain Sakai (clade 1) and −8.87 ± 0.63 for strain MG1655. For SNP 438 and SNP 630, the average \( \Delta C_T \) 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 \( \Delta C_T \) was 10.24 ± 0.41 for clade 8 strains, and the average SNP 630 \( \Delta C_T \) 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 \( \Delta C_T \) was 7.69 ± 1.29, whereas for the remaining clade 2 strains, the average SNP 1060 \( \Delta C_T \) 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.

We thank Galeb Abu-Ali for critically reviewing earlier versions of the manuscript.

The STEC Center is supported with funds from the NIAID, NIH, DHHS, under NIH research contract N01-AI-30058 (to T.S.W.).

REFERENCES