Sequencing and analysis of the *Escherichia coli* serogroup O117, O126, and O146 O-antigen gene clusters and development of PCR assays targeting serogroup O117-, O126-, and O146-specific DNA sequences

Yanhong Liua, Chitrita DebRoyb, Pina Fratamicoa,*

aEastern Regional Research Center, US Department of Agriculture, Agricultural Research Service, Wyndmoor, PA 19038, USA
bGastroenteric Disease Center, Department of Veterinary Science, The Pennsylvania State University, University Park, PA 16802, USA

Received 30 October 2006; accepted 8 March 2007
Available online 14 March 2007

Abstract

The O-antigen gene clusters of *Escherichia coli* serogroups O117, O126, and O146 were sequenced, and 11, 10, and 11 open reading frames (ORFs) were identified, respectively. Genes required for O-antigen sugar biosynthesis, sugar transfer, and sugar processing were identified. Multiplex polymerase chain reaction (PCR) assays were developed targeting the *wzx* and *wzy* genes present in the O-antigen gene cluster of these serogroups. The assays were highly serogroup specific when tested against strains belonging to serogroups that were isolated from food, humans, animals, and environmental sources, as well as against representative strains belonging to ca. 165 different *E. coli* O serogroups and a number of non-*E. coli* bacteria. Thus, the results demonstrate that the *wzx* and *wzy* gene sequences were specific to *E. coli* O117, O126, and O146 and can be used as diagnostic markers for rapid identification and detection of these serogroups. Published by Elsevier Ltd.

Keywords: *wzx*; *wzy*; Serotyping; *E. coli* O117; *E. coli* O126; *E. coli* O146; Serogroup; O-antigen; PCR

1. Introduction

*Escherichia coli*, which includes both commensal and pathogenic strains, is the most thoroughly studied bacterial species in the microbial world. The six major categories of *E. coli* strains that can cause enteric diseases in humans are (1) enterohemorrhagic *E. coli* (EHEC), (2) enterootoxigenic *E. coli* (ETEC), (3) enteropathogenic *E. coli* (EPEC), (4) enteroggregative *E. coli* (EAggEC), (5) enteroinvasive *E. coli* (EIIE), and (6) diffusely adherent *E. coli* (DAEC) [1].

ETEC strains are a major cause of traveler’s diarrhea. This disease is characterized by adherence to and colonization of the intestinal mucosa by the bacteria and the production of the enterotoxins, heat-labile toxin (LT) and/or heat-stable toxin (ST). Strains of *E. coli* serogroup O117 have been identified as ETEC due to the presence of heat-stable enterotoxin genes and their ability to adhere to the brush border of human enterocytes [2]. *E. coli* O117 strains have been associated with acute diarrhea in infants in Africa [3]. In addition to enteric disease, *E. coli* O117 was also implicated in urinary tract infection through sexual transmission [4].

EPEC strains are a major cause of diarrhea in children living in developing countries and are characterized by the presence of two virulence markers: a chromosomal *eae* gene involved in intimate attachment to intestinal cells and a plasmid harboring genes involved in localized adherence. *E. coli* serogroup O126 is a class I EPEC [5], even though some strains of this serogroup do not possess the virulence markers described above [6]. *E. coli* O126 strains have been associated with sporadic cases and outbreaks of infantile diarrhea [6]. Although strains of *E. coli* O126 are human pathogens, they are also found in healthy animals, such as dairy goats [7].

All EHEC strains produce Shiga toxins (Stx), also known as verotoxins or verocytotoxins. Thus, EHEC strains are also called Shiga toxin-producing *E. coli* (STEC) [5]. STEC are recognized as important pathogens and have...
E. coli serotyping is typically performed by agglutination reactions using antisera raised in rabbits against the ca. 165 different O standard reference strains. However, traditional serotyping is both laborious and time consuming, and it often generates equivocal results due to cross-reactions between different serogroups. Furthermore, the antisera used for serotyping can only be generated by specialized laboratories with animal facilities. As such, rapid, more specific molecular methods for identifying different E. coli serogroups are needed.

The O-antigen, which contains many repeats of an oligosaccharide unit (O unit), is present in the outer membrane of Gram-negative bacteria and contributes the major antigenic variability to the cell surface. The genes involved in the biosynthesis of O-antigens in E. coli are located in the O-antigen gene cluster and are flanked by the galF and gnd genes on the E. coli chromosome. The genes that encode proteins within the E. coli O-antigen gene clusters consist of the following three categories: nucleotide sugar biosynthesis, glycosyltransferase, and O-antigen processing genes [14]. Nucleotide sugar biosynthesis gene products are involved in the biosynthesis of the nucleotide sugar precursors in the O-antigen, which occurs in the cytoplasm. Glycosyltransferases, usually have narrow substrate specificity and are responsible for transferring the various precursor sugars to form an oligosaccharide on a carrier lipid, undecaprenyl phosphate (UndP), which is located on the inner membrane facing the cytoplasmic side. O-antigen processing proteins include a flippase (Wzx) and the O-antigen polymerase (Wzy). The Wzx flips the O-unit across the inner membrane. After the UndPP-linked O-unit is translocated across the cytoplasmic membrane, the O-units are linked together by Wzy through a glycosidic linkage. Although both Wzx and Wzy are membrane proteins usually with high variation among different microorganisms, the action of Wzx is not very specific while the function of Wzy is specific [14].

A number of E. coli O-antigen gene clusters have been sequenced and the genes were annotated [8,15–21]. Several genes in the O-antigen gene clusters, in particular, the wzx (O-antigen flippase) and wzy (O-antigen polymerase) genes, show relatively low similarity among different E. coli serogroups, and PCR primers targeting the wzx and wzy genes have been used to develop serogroup-specific PCR assays [15–21]. Our objective was to sequence and characterize the O-antigen gene clusters of E. coli O117, O126, and O146 serogroups and identify specific genes that can be used as diagnostic markers for these serogroups.

2. Materials and methods

2.1. Bacterial strains

E. coli O117:H7 strain 97-3039, E. coli O126:H27 strain 89-3506, and E. coli O146:H21 strain 90-3158 used for DNA sequencing of the O-antigen gene clusters were obtained from The Centers for Disease Control and Prevention (CDC), Atlanta, GA. Bacteria used to test for specificity of the PCR included the following: 90 E. coli O117 strains, 77 E. coli O126 strains, 98 E. coli O146 strains, 47 non-O117, non-O126, and non-O146 E. coli strains isolated from humans, animals, food, and water, and 165 E. coli reference standard strains belonging to serogroups O1–O175 used for serotyping, except for O14, O31, O47, O67, O72, O93, O94 and O122 strains that have not been designated [22]. In addition, strains representative of other bacterial genera (n = 20) used to test the PCR specificity included Bacillus cereus, Citrobacter freundii, Enterobacter cloacae, Enterococcus aerogenes, Enterococcus faecalis, Hafnia alvei, Klebsiella pneumonia, Listeria monocytogenes, Pseudomonas aeruginosa, Proteus vulgaris, Salmonella Anatum, Salmonella Arizona, Salmonella Choleraesuis, Salmonella Enteritidis, Salmonella Typhimurium, Serratia marcescens, Shigella boydii, Staphylococcus aureus, Vibrio cholerae, and Yersinia enterocolitica. All bacteria were grown in Luria Bertani (LB) broth or on LB agar plates at 37°C.

2.2. Construction of a DNaseI shotgun library, DNA sequencing, and gene annotation

Genomic DNA was isolated using the DNeasy Tissue Kit (Qiagen, Inc., Valencia, CA) according to the manufacturer’s instructions. Long PCR assays were performed to amplify the O-antigen gene clusters using the Expand Long Template PCR System (Roche Applied Science, Mannheim, Germany) and JUMPSTART (named for Just Upstream of Many Polysaccharide-associated gene STARTs) and GND (6-phosphogluconate dehydrogenase gene) primers that flank the E. coli O-antigen gene clusters. The sequence of the JUMPSTART sense primer was 5'-ATTGGTAGCTGTAAAGCAGGCGGTACG-3', and the antisense GND primer sequence was 5’-CAGTCGCATAACGCAGACCGCGATCTGTGCTTG-3’ (In vitrogen Life Technologies, Inc., Carlsbad, CA). The long PCR conditions were as described previously [17]. The long PCR products were verified on 0.8% agarose gels and purified according to the instructions in the QIAquick PCR Purification Kit (Qiagen, Inc., Valencia, CA). DNase I digestion, shotgun cloning, and DNA sequencing were performed as described previously [17]. To confirm the sequence of each O-antigen gene cluster, ten individual long PCR products were pooled together and the DNA was resequenced using primers designed from different regions of the O-antigen gene clusters. Sequencing data were assembled using Sequencher software, and gene
2.3. Testing for specificity by the PCR

Bacterial DNA used as template for the PCR assays was isolated as described previously [17]. PCR primers were designed from the wzx and wzy genes found in the O-antigen gene clusters of each serogroup using the Primer3 software program. Multiplex PCR reactions were performed and analyzed as described [17].

2.4. Nucleotide sequence accession numbers

The DNA sequences of the O-antigen gene clusters of E. coli O117, O126, and O146 were deposited into GenBank under the accession numbers DQ465247, DQ465248, and DQ465249, respectively.

3. Results and discussion

DNA sequences of 10,886, 11,783, and 11,888 bases were obtained from the E. coli O117, O126, and O146 O-antigen gene clusters, respectively (Tables 1–3). All of the genes in each of the clusters had the same transcriptional direction from galF to gnd. The deduced amino acid sequences from these ORFs were used to search the NCBI database for indication of their possible functions. Gene names were assigned on the basis of the bacterial polysaccharide gene nomenclature system (http://www.microbio.usyd.edu.au/BPGD/big-paper.pdf).

### 3.1. Sequence analysis of the E. coli O117 O-antigen gene cluster

#### 3.1.1. Sugar biosynthetic pathway genes

The structure of the O-unit of E. coli O117 has been characterized, and it contains four residues: N-acetyl galactosamine (GalpNAc), L-Rhamnose (L-Pha), glucose (Glcp), and galactose (Galp) [24] (Fig. 1). Since genes for the synthesis of common sugars, including glucose and galactose are located outside the O-antigen gene cluster, only genes involved in the synthesis of L-Rhamnose were expected in the E. coli O117 O-antigen gene cluster.

As shown in Table 1, ORF1, ORF2, ORF3, and ORF4 of the O117 O-antigen gene cluster showed between 77% and 98% identity to known Rml B, D, A, and C proteins, respectively. L-Rhamnose synthesis requires rml genes and these genes have been well characterized [14]. Recently, L-Rhamnose was synthesized in vitro using the enzymes found in the L-Rhamnose synthetic pathway [25]. Therefore, ORF1–4 were identified as rmlB, rmlD, rmlA, and rmlC, respectively, and named accordingly. The rml genes are usually found together at the 5’ end of the O-antigen gene clusters in the order of rmlBDAC [14]. Our results are consistent with these findings; However, there are exceptions. For example, in the E. coli O177 and Shigella boydii type 9 O-antigen gene clusters the rmlC gene is separated from other rml genes by wzx [15,20].

ORF11 shows 80% similarity to a putative UDP-galacto-4-epimerase found in E. coli. The UDP-galacto-4-epimerase is the product of galE, and it catalyzes the conversion of UDP galactose to UDP glucose and the reverse reaction. Even though the galE gene is a housekeeping gene that is located elsewhere in the chromosome...

---

### Table 1

<table>
<thead>
<tr>
<th>ORF Proposed gene name</th>
<th>Location</th>
<th>No. of amino acids</th>
<th>Putative function</th>
<th>Most significant homology (accession no.)</th>
<th>% Identity/% similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 rmlB 300–1193 297</td>
<td>dTDP-glucose 4,6-dehydratase [E. coli] (AAN60454.1)</td>
<td>98/98</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 rmlD 1193–2092 299</td>
<td>dTDP-6-deoxy-a-glucose-3,5 epimerase [E. coli] (AAZ5715.1)</td>
<td>98/99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 rmlA 2150–3028 292</td>
<td>Glucose-1-phosphate thymidylyltransferase [Shigella boydii] (AAL27324.1)</td>
<td>99/100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 rmlC 3033–3575 180</td>
<td>dTDP-6-deoxy-a-glucose-3,5 epimerase</td>
<td>Putative protein [E. coli] (AAZ65384.1)</td>
<td>77/86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 wzx 3575–4792 405</td>
<td>O-antigen flippase</td>
<td>Polysaccharide biosynthesis protein [Shewanella baltica OS155] (ZP_00581317.1)</td>
<td>26/45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 wzy 4785–6107 440</td>
<td>O-antigen polymerase</td>
<td>Secreted polysaccharide polymerase [Bacillus cereus ATCC 14579] (AAP12132.1)</td>
<td>22/40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 wbeA 6100–6900 266</td>
<td>Glycosyl transferase, family 2</td>
<td>Glycosyl transferase, family 2 [Pseudomonas aeruginosa DSM 2379] (ZP_00676473)</td>
<td>42/58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 wbeB 6887–7831 314</td>
<td>Glycosyl transferase, family 2</td>
<td>Glycosyl transferase, family 2 [Shewanella putrefaciens CN-32] (ZP_00813359.1)</td>
<td>35/55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 wbeC 7828–8889 353</td>
<td>Glycosyltransferase</td>
<td>Putative LPS biosynthesis related glycosyltransferase [Bacteroides fragilis NCTC 9343] (CAH09146.1)</td>
<td>33/51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 wbeD 8886–9695 269</td>
<td>Glycosyl transferase</td>
<td>WecG [E. coli] (AAY74552.1)</td>
<td>50/67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 galE 9709–10728 339</td>
<td>UDP-galactose-4-epimerase</td>
<td>Putative UDP-galactose-4-epimerase [E. coli] (AAH373174.1)</td>
<td>69/80</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2
Open reading frames (ORFs) located in the E. coli O126 O-antigen gene cluster

<table>
<thead>
<tr>
<th>ORF #</th>
<th>Gene name</th>
<th>Location</th>
<th>No. of amino acids</th>
<th>Putative function</th>
<th>Most significant homology (accession no.)</th>
<th>% Identity/% similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>wzy</td>
<td>43–1086</td>
<td>347</td>
<td>O-antigen polymerase</td>
<td>Wzy [Yersinia enterocolitica (type 0:3)] (AAC60768.1)</td>
<td>34/57</td>
</tr>
<tr>
<td>2</td>
<td>wbgK</td>
<td>1083–2084</td>
<td>333</td>
<td>Glycosyltransferase</td>
<td>WbcF [Yersinia enterocolitica (type 0:3)] (AAC60769.1)</td>
<td>42/65</td>
</tr>
<tr>
<td>3</td>
<td>wbgL</td>
<td>2094–2987</td>
<td>297</td>
<td>Glycosyltransferase</td>
<td>Glycosyltransferase [Submonella enterica subsp. salamae serovar Greenside] (AYA34522.1)</td>
<td>42/57</td>
</tr>
<tr>
<td>4</td>
<td>wbsT</td>
<td>2994–3764</td>
<td>356</td>
<td>Glycosyltransferase</td>
<td>WblT protein [Photorhabdus luminescens] (CAE17191.1)</td>
<td>55/73</td>
</tr>
<tr>
<td>5</td>
<td>wbgL</td>
<td>3768–4898</td>
<td>376</td>
<td>Glycosyltransferase</td>
<td>WblU protein [Photorhabdus luminescens] (CAE17192.1)</td>
<td>39/60</td>
</tr>
<tr>
<td>6</td>
<td>gmd</td>
<td>4972–6093</td>
<td>313</td>
<td>GDP-α-mannose dehydratase</td>
<td>GDP-α-mannose dehydratase [Yersinia mollaretii ATCC 43969]</td>
<td>81/90</td>
</tr>
<tr>
<td>7</td>
<td>fcl</td>
<td>6059</td>
<td>334</td>
<td>GDP-α-l-fucose synthetase</td>
<td>GDP-α-l-fucose synthetase [Yersinia pseudotuberculosis (type O:1b)] (CAB63301.1)</td>
<td>97/98</td>
</tr>
<tr>
<td>8</td>
<td>manC</td>
<td>7150–8547</td>
<td>465</td>
<td>Mannose-1-phosphate guanylyltransferase</td>
<td>Mannose-1-phosphate guanylyltransferase [Aeromonas hydrophila] (AAM74484.1)</td>
<td>56/75</td>
</tr>
<tr>
<td>9</td>
<td>manB</td>
<td>8611–10050</td>
<td>479</td>
<td>Phosphomannomutase</td>
<td>ManB [E. coli] (AAY74380.1)</td>
<td>54/73</td>
</tr>
<tr>
<td>10</td>
<td>wzx</td>
<td>10154–11638</td>
<td>494</td>
<td>Flippase</td>
<td>Polysaccharide transport protein, putative [Bacillus cereus ATCC 10987] (AAS44286.1)</td>
<td>30/53</td>
</tr>
</tbody>
</table>

Table 3
Open reading frames (ORFs) in the O-antigen gene cluster of E. coli serogroup O146

<table>
<thead>
<tr>
<th>ORF</th>
<th>Gene name</th>
<th>Location</th>
<th>No. of amino acids</th>
<th>Putative function</th>
<th>Most significant homology (accession no.)</th>
<th>% Identity/% similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rmlB</td>
<td>301–1194</td>
<td>297</td>
<td>dTDP-glucose 4,6-dehydratase</td>
<td>dTDP-glucose 4,6-dehydratase [E. coli] (AAZ85703.1)</td>
<td>98/99</td>
</tr>
<tr>
<td>2</td>
<td>rmlD</td>
<td>1194–2093</td>
<td>299</td>
<td>dTDP-4-dehydrorhamnose reductase</td>
<td>dTDP-4-dehydrorhamnose reductase [E. coli] (AAZ85702.1)</td>
<td>97/98</td>
</tr>
<tr>
<td>3</td>
<td>rmlA</td>
<td>2151–3029</td>
<td>292</td>
<td>Glucose-1-phosphate thymidylyltransferase</td>
<td>Glucose-1-phosphate thymidylyltransferase [E. coli] (AAZ85705.1)</td>
<td>98/100</td>
</tr>
<tr>
<td>4</td>
<td>rmlC</td>
<td>3034–3567</td>
<td>177</td>
<td>dTDP-6-deoxy-α-glucose-3,5 epimerase</td>
<td>dTDP-6-deoxy-α-glucose-3,5 epimerase [E. coli] (AAZ85706.1)</td>
<td>66/80</td>
</tr>
<tr>
<td>5</td>
<td>wbuD</td>
<td>3582–4574</td>
<td>330</td>
<td>Nitroreductase</td>
<td>Nitroreductase [Aeromonas hydrophila] (EAN02140.1)</td>
<td>29/47</td>
</tr>
<tr>
<td>6</td>
<td>wzx</td>
<td>4571–6109</td>
<td>512</td>
<td>Unknown</td>
<td>Hypothetical protein BT_0040 [Bacteroides thetaiotaomicron VPI-5482] (AAO75147.1)</td>
<td>32/50</td>
</tr>
<tr>
<td>7</td>
<td>wbuE</td>
<td>6102–7205</td>
<td>367</td>
<td>Unknown</td>
<td>Hypothetical protein [Bacteroides fragilis NCTC 9343] (CAH08488.1)</td>
<td>32/50</td>
</tr>
<tr>
<td>8</td>
<td>wbuU</td>
<td>7552–8352</td>
<td>266</td>
<td>Glycosyltransferase</td>
<td>Eps4L [Streptococcus thermophilus] (AAN63688.1)</td>
<td>29/52</td>
</tr>
<tr>
<td>9</td>
<td>wzy</td>
<td>8472–9650</td>
<td>392</td>
<td>Polymerase</td>
<td>putative polysaccharide polymerase [Staphylococcus saprophyticus] (ZP_00762371.1)</td>
<td>21/40</td>
</tr>
<tr>
<td>10</td>
<td>wbtA</td>
<td>9613–10,500</td>
<td>295</td>
<td>Glycosyltransferase</td>
<td>Glycosyl transferase, family 2 [Pseudoalteromonas atlantica] (ZP_00762371.1)</td>
<td>46/62</td>
</tr>
<tr>
<td>11</td>
<td>wbtB</td>
<td>10,550–11,362</td>
<td>270</td>
<td>Glycosyltransferase</td>
<td>WbcZ [E. coli] (AAY28251.1)</td>
<td>51/67</td>
</tr>
</tbody>
</table>

Fig. 1. The structures of the E. coli O117 (a) [24] and O126 (b) [29] O-antigen oligosaccharides.
gene cluster. ORF6 shows 90% similarity to GDP-D- and 55% similarity to proteins in the glycosyltransferase family 2 of Pelobacter propionicus DSM 2379 and Shewanella putrefaciens CN-32, respectively. Therefore, ORF7 and ORF8 were proposed to be glycosyltransferase genes and named ORF7 and 8 share 58% similarity to WclG in E. coli W32, respectively. Therefore, ORF7 and ORF8 were proposed to be glycosyltransferase genes and named ORF7 and ORF8. Likewise, ORF9 and ORF10 share 51% and 67% similarity with putative glycosyltransferases in Bacteroides fragilis NCTC 9343 and WeG in E. coli, respectively. Based on high levels of similarity, ORF9 and ORF10 were proposed to be glycosyltransferases and were named wbcA and wbcB, respectively.

3.1.2. Sugar transferase genes
Glycosyltransferases are specific to different sugar donors, acceptors, and linkages between two sugars. Since the structure of the E. coli O117 O-unit is a pentasaccharide (Fig. 1), four glycosyltransferases were expected in the O-antigen gene cluster. In fact, ORF 7 and 8 share 58% and 55% similarity to proteins in the glycosyltransferase family 2 of Pelobacter propionicus DSM 2379 and Shewanella putrefaciens CN-32, respectively. Therefore, ORF7 and ORF8 were proposed to be glycosyltransferase genes and named wbeA and wbeB, respectively.

3.1.3. O-antigen processing genes
The ORF5 shows low homology with a number of Wzx proteins that have been identified previously, and ORF6 shows 40% similarity to a secreted polysaccharide polymerase in B. cereus. Thus, ORF5 and ORF6 were named wzx and wzy, respectively. As expected, both Wzx and Wzy proteins were hydrophilic and contained 12 and 13 transmembrane domains, respectively, as analyzed by the HMMTOP program [23]. Since these genes are not very conserved among different E. coli serogroups, wzx and wzy are suitable targets for serogroup-specific PCR assay development.

3.2. Sequence analysis of the E. coli O126 O-antigen gene cluster

3.2.1. Sugar biosynthetic pathway genes
The structure of the E. coli O126 O-antigen contains the following pentasaccharide repeats: mannose, galactose, N-acetylglucosamine (two), and L-fucose [29] (Fig. 1). Since galactose and N-acetylglucosamine are synthesized elsewhere in the genome, only genes responsible for mannose and L-fucose are expected to be present in the O-antigen gene cluster. ORF6 shows 90% similarity to GDP-mannose dehydratase in Yersinia mollaretii, and ORF7 shows 80% similarity to GDP-L-fucose synthetase in Yersinia pseudotuberculosis (Table 2). Both GDP-mannose dehydratase (gmd) and GDP-L-fucose (fcl) were involved in the biosynthesis of L-fucose [18,19]. Therefore, ORF6 and ORF7 were named gmd and fcl, respectively. ORF8 displays 75% similarity to mannone-1-phosphate guanylyltransferase (ManC), and ORF9 shows 73% similarity to phosphomannomutase (ManB) in E. coli. Based on these similarities, ORF6 and ORF7 were named manC and manB, respectively. Both ManC and ManB have been well characterized and were involved in the biosynthesis of GDP-mannose [14]. The gmd, fcl, manC, and manB genes are also present in a number of other O-antigen gene clusters, including those of E. coli serogroups O86, O111, and O128 [16,18,19].

3.2.2. Sugar transferase genes
Since the O-unit of E. coli O126 has a pentasaccharide structure (Fig. 1), four glycosyltransferases were expected to be present in the O-antigen gene cluster of this E. coli serogroup. Consistent with this finding, ORF2-5 show different levels of similarities to glycosyltransferases; therefore, they were proposed to have glycosyltransferase activity and named wbgK, wbgL, wbgT, and wbgU, respectively. As for other O-antigen genes, ORF2-5 may be responsible for linking the sugars together to form a pentasaccharide.

3.2.3. O-antigen processing genes
The O-antigen processing genes, wzx and wzy, are also present in the O-antigen gene cluster of E. coli O126, although their locations are different than in E. coli O117. ORF1 shows 57% similarity to Wzy in Y. enterocolitica (Table 2), and in addition, ORF1 has 10 transmembrane helices as revealed by the HMMTOP program. Therefore, ORF1 was named wzy. ORF10 shows 53% similarity to a putative polysaccharide transport protein in Bacillus cereus (Table 2) and also shows high homology to a number of flippases (data not shown). The HMMTOP program revealed that ORF10 has 14 transmembrane helices. Based on these results, ORF10 was expected to encode a flippase and named wzx, accordingly.

3.3. Sequence analysis of the E. coli O146 O-antigen gene cluster

3.3.1. Sugar biosynthetic pathway genes
The structure of the E. coli O146 polysaccharide has not yet been determined. However, ORF1, ORF2, ORF3, and ORF4 showed between 66% and 98% identity to known Rml B, D, A, and C proteins, respectively. This indicates that the E. coli O146 polysaccharide may contain Rhamnose similar to E. coli O117. ORF5 shows 47% similarity to a nitroreductase in Methylobacillus flagellatus KT. This gene has never been reported to be involved in O-antigen polysaccharide biosynthesis. ORF7 shows 50% similarity to an unknown protein. The presence of the unknown protein and the nitroreductase in the O-antigen gene cluster may indicate a novel polysaccharide structure in the O-unit of E. coli O146.

3.3.2. O-antigen processing genes
ORF6 shows 50% similarity to a hypothetical membrane protein in Bacteroides thetaiotaomicron speculated to be involved in the export of O-antigen and teichoic acid.
Analysis using HMMTOP revealed that ORF6 contains 14 transmembrane helices. In addition, ORF6 also shows homology to a number of flippases in other bacteria such as *Vibrio parahaemolyticus*, *B. fragilis*, and *Streptococcus pneumoniae*. Therefore, ORF6 was speculated to encode a flippase and named \textit{wzx}. ORF9 shows 40% similarity to a putative polysaccharide polymerase in \textit{Staphylococcus saprophyticus}. ORF9 was predicted to encode a polymerase and was named \textit{wzy}. ORF9 also has 11 transmembrane helices as indicated by HMMTOP analysis.

### 3.3.3. Sugar transferase genes

ORF8 shows 52% similarity to Eps4L in *Streptococcus thermophilus*, a glycosyltransferase in group 1. ORF9 shows 62% similarity to WbcZ in *E. coli*, which is a putative glycosyltransferase. Based on the homologies to glycosyltransferases, ORF8, 10, and 11 were proposed to be glycosyltransferases and named \textit{wbuU}, \textit{wbuV}, and \textit{wbuW}, respectively. The fact that three glycosyltransferases are present in the O-antigen gene cluster indicates that the polysaccharide unit of *E. coli* O146 may be tetrasaccharide.

### 3.4. Identification of *E. coli* O117, O126, and O146 serogroup-specific genes and specificity testing

Sequencing analysis revealed that the \textit{wzx} and \textit{wzy} genes of the O-antigen gene clusters of *E. coli* O117, O126, and O146 share the least similarities with known \textit{wzx} and \textit{wzy} genes, and this is consistent with previous studies that demonstrated that the \textit{wzx} and \textit{wzy} genes were serogroup specific. Primers targeting the \textit{wzx} gene of *E. coli* O117, O126, and O146 were designed (Table 4) and used in multiplex PCR assays to determine their specificities for these serogroups. Multiplex PCR assays were performed, and the results are shown in Fig. 2. In this study, PCR reactions were performed to test the specificities against the ca. 165 *E. coli* standard strains, 47 strains of non-O117, O126, and O146 *E. coli*, as well as 20 non-*E. coli* bacterial strains. Moreover, 90 strains belonging to *E. coli* O117 (collected between 1976 and 2005), 77 strains belonging to O126 (collected between 1982 and 2005), and 98 strains belonging to *E. coli* O146 (collected between 1976 and 2006) isolated from humans, animals, food, and water were also included for specificity analysis. PCR assays targeting the \textit{wzx} and \textit{wzy} genes of *E. coli* O126 and O146 were highly specific for their respective serogroups. Cross-reactions did not occur with other *E. coli* strains or with the non-*E. coli* strains tested. Therefore, the \textit{wzx} and \textit{wzy} primers designed in this study could potentially be used in rapid diagnostic assays for *E. coli* serogroups O126 and O146. Interestingly, primers targeting the \textit{wzx} and \textit{wzy} genes of *E. coli* O117 also amplified the \textit{wzx} and \textit{wzy} genes in *E. coli* O107 (Fig. 2). Restriction fragment length polymorphism (RFLP) analyses were performed on the long PCR products of the O-antigen gene clusters of *E. coli* O117 and O107. Since both *E. coli* O117 and O107 had indistinguishable RFLP patterns (data not shown), we speculated that O117 and O107 have identical or at least very similar O-antigen gene cluster sequences. The fact that O117 and O107 antisera have also been found to cross-react supports this notion [30].

In the current study, we determined the DNA sequences of the O-antigen gene clusters of strains representing *E. coli* O117, O126, and O146.
serogroups O117, O126, and O146, identified the genes in the O-antigen gene clusters, and developed specific PCR primers for each serogroup targeting the \( wzx \) and \( wzy \) genes. Since strains of \( E. coli \) O117, O126, and O146 are human pathogens, identifying diagnostic markers will help in the clinical diagnosis of diseases caused by these \( E. coli \) serogroups. Sequence information of the O-antigen gene clusters will also assist in developing microarray-based assays for \( E. coli \) serotyping. Recently, DNA microarrays were developed for rapid identification of different serogroups of \( E. coli \) in a single platform [31]. Unique sequences in the O-antigen gene clusters of several \( E. coli \) serogroups were selected and placed onto the microarrays, and specific signals were generated for each serogroup tested. Sequencing of additional \( E. coli \) O-antigen gene clusters will allow us to expand the DNA array to contain unique target sequences for all the \( E. coli \) serogroups.

The O-antigen lipopolysaccharide has been associated with a number of biological phenomena including stress responses [32], swarming motility [33], virulence [34], and flagellum biogenesis [35]. Elucidation of O-antigen gene sequences will provide additional insight on these biological phenomena. In addition, characterization of the O-antigen gene clusters will aid in designing diagnostic markers for specific serogroups and in the development of \( E. coli \) O-antigen-based vaccines.

Acknowledgments

We are grateful to Drs. John Luchansky and Shu-I Tu (Eastern Regional Research Center, Wyndmoor, PA) for their support. We thank Mr. Paul Pierlott (Eastern Regional Research Center, Wyndmoor, PA) for processing the images. We also thank Dr. Connie Briggs, Dr. David Needleman, and Ms. Laurie Fortis (Eastern Regional Research Center, Wyndmoor, PA) for assisting in sequencing the O-antigen gene clusters of \( E. coli \) O117, O126, and O146, and Ms. Elizabeth Roberts at the Pennsylvania State University for performing the PCR assays.

References


[26] Fratamico PM, DebRoy C, Strobaugh Jr TP, Chen CY. DNA sequence of the \( Escherichia coli \) O103 O antigen gene cluster and


