FOODBORNE PATHOGENS AND DISEASE
Volume 5, Number 4, 2008

DOI: 10.1089/fpd.2008.0096

DNA Sequencing and Identification of Serogroup-Specific Genes in the *Escherichia coli* O118 O Antigen Gene Cluster and Demonstration of Antigenic Diversity But Only Minor Variation in DNA Sequence of the O Antigen Clusters of *E. coli* O118 and O151

Yanhong Liu,[1] Pina Fratamico,[1] Chitrita Debroy,[2] Alyssa C. Bumbaugh,[3] and John W. Allen[4]

Abstract

The DNA sequence of the O antigen gene cluster of an *Escherichia coli* serogroup O118 strain was determined, and 13 open reading frames (ORFs) were identified, encoding genes required for O antigen sugar biosynthesis, transfer, and processing. Polymerase chain reaction (PCR) assays targeting the *wzx* (O antigen flippase) and *wzy* (O antigen polymerase) genes in the O antigen gene cluster of *E. coli* O118 were designed for identification of these serogroups. Specificity testing using strains belonging to *E. coli* O118 isolated from various sources, representative strains of 167 other *E. coli* O serogroups, and 20 non–*E. coli* bacteria revealed that the PCR assays were specific for *E. coli* O118. Thus, the PCR assays can be used for rapid identification of *E. coli* O118 as an alterative to typing using antisera. However, the PCR assays targeting the *E. coli* O118 *wzx* and *wzy* genes were also positive using *E. coli* serogroup O151 DNA. Therefore, the sequence of the O antigen gene cluster of *E. coli* O151 was determined, and it was very similar to that of *E. coli* O118, with only three nucleotide differences. Although the lipopolysaccharide profiles of O118 and O151 showed differences, multilocus sequence typing of *E. coli* O118 and O151 strains only revealed minor variation at the nucleotide level. Since *E. coli* O118 strains are more frequently isolated from humans, animals, and the environment than *E. coli* O151, serogroup O151 may likely be a minor variant of *E. coli* O118. Further studies are needed to elucidate this possibility.

Introduction

Escherichia coli is the predominant facultative anaerobe in the human gastrointestinal tract, and pathogenic *E. coli* strains are frequent causes of urinary tract infections, sepsis and meningitis, and enteric and diarrheal disease. The six major categories of human diarrheagenic strains include: (1) enterohemorrhagic *E. coli* (EHEC), (2) enterotoxigenic *E. coli*, (3) enteropathogenic *E. coli*, (4) enteroaggregative *E. coli*, (5) enteroinvasive *E. coli*, and (6) diffusely adherent *E. coli* (Smith and Fratamico, 2005).

E. coli O118 strains have been identified as EHEC and thus also as Shiga toxin–producing *E. coli* (STEC). STEC O118 strains cause diarrhea in calves and have been associated with human cases of bloody diarrhea and hemolytic uremic

[1]U.S. Department of Agriculture, Agricultural Research Service, Eastern Regional Research Center, Wyndmoor, Pennsylvania.
[2]*E. coli* Reference Center, Department of Veterinary and Biomedical Sciences, The Pennsylvania State University, University Park, Pennsylvania.
[3]Division of Mathematics & Natural Sciences, Altoona College, The Pennsylvania State University, Altoona, Pennsylvania.
[4]Department of Animal Sciences, North Carolina A & T State University, Greensboro, North Carolina.

syndrome (HUS) (Beutin *et al*., 1998; Wieler *et al*., 1998). An outbreak of gastrointestinal illness caused by a Shiga toxin 1–producing *E. coli* O118:H2 strain occurred at a junior high school in Japan in 1996 (Hashimoto *et al*., 1999). Of the 131 individuals from whom STEC O118:H2 was isolated, 57 had diarrhea (4 of 57 had bloody diarrhea); 25 had abdominal pain, vomiting, and/or headache without diarrhea; and 49 were asymptomatic. The pathogen was isolated from a food utensil used at the school. There were no cases of HUS.

E. coli serogroup O151 strains cause human illness infrequently; however, a STEC O151:H12 strain was isolated from beef in the United Kingdom, and a STEC O151:H−strain was isolated from asymptomatic individuals in Germany (Willshaw *et al*., 1993; Friedrich *et al*., 2002). *E. coli* O151:H10 was associated with diarrhea in individuals in Japan and the former Soviet Union (Steinruck *et al*., 1980). Cross-reactivity of antisera against *E. coli* O151 can occur with *Citrobacter* O1 and O14 and *Hafnia* spp., and biochemical tests may be needed to differentiate *E. coli* O151 strains from the other bacterial genera (Kiseleva *et al*., 1981). Thus, detection and identification of this serogroup may be problematic.

The O antigen, which contains many repeats of an oligosaccharide unit (O unit), is present in the outer membrane of gram-negative bacteria and is the major contributor to the antigenic variability of the cell surface. *E. coli* serotyping is conventionally performed by agglutination reactions using antisera raised in rabbits against the approximately 165 different O standard references strains. However, traditional serotyping is laborious and time consuming and often generates equivocal results due to cross-reactions among different serogroups. Furthermore, the antisera used for serotyping can only be generated by specialized laboratories with animal facilities. Rapid, more specific molecular methods for identifying different *E. coli* serogroups are needed.

The genes involved in the biosynthesis of O antigens in *E. coli* are located in the O antigen gene cluster, which is flanked by the *galF* and *gnd* genes on the *E. coli* chromosome. A number of *E. coli* O antigen gene clusters have been sequenced, and the genes were annotated (D'Souza *et al*., 2002; Wang *et al*., 2002; Fratamico *et al*., 2003; DebRoy *et al*., 2004, 2005; Beutin *et al*., 2005a, 2005b; Feng *et al*., 2005). Several genes in the clusters, in particular the *wzx* (O antigen flippase) and *wzy* (O antigen polymerase) genes, show relatively low similarity among different *E. coli* serogroups, and polymerase chain reaction (PCR) primers targeting the *wzx* and *wzy* genes have been used to develop serogroup-specific PCR assays (D'Souza *et al*., 2002; Wang *et al*., 2002; Fratamico, *et al*., 2003; DebRoy *et al*., 2004, 2005; Beutin *et al*., 2005a, 2005b; Feng *et al*., 2005).

The objective of the current study was to sequence and characterize the O antigen gene clusters of strains representative *E. coli* serogroups O118 and O151 and to identify specific genes that can be used as diagnostic markers for these serogroups. Since strains of *E. coli* O118 and O151 are human pathogens, identifying diagnostic markers will be helpful for detection and clinical diagnosis of diseases caused by these *E. coli* serogroups.

Materials and Methods

Bacterial strains and culture conditions

E. coli O118:H16 strain 31w and *E. coli* O151:H10 strain 880-67 (Ørskov *et al*., 1977), serotyping standard strains, were obtained from the World Health Organization (WHO). These strains were used for DNA sequencing of the O antigen gene clusters. *E. coli* O118:H16 98.0556 (isolated from a cow) and *E. coli* O151:H26 99.1227 (isolated from a rattlesnake) strains, as well as other strains used to validate the sensitivity and specificity of the PCR assays were from the culture collection of the *E. coli* Reference Center at the Pennsylvania State University. Bacteria used to test for specificity of the PCR included the following: 65 *E. coli* O118 strains isolated from humans, animals, food, and water; 47 non–O118 field isolates; and 167 *E. coli* standard reference strains used for serotyping at the *E. coli* Reference Center. The 167 *E. coli* reference standard strains used in this study belonged to serogroups O1-O175, but excluding O14, O31, O47, O67, O72, O93, O94, and O122 strains, since these serogroups have not been designated (Ørskov *et al*., 1977). In addition, 20 strains representative of other bac-

terial genera used to test the specificity of the PCR assays included *Bacillus cereus*, *Citrobacter freundii*, *Enterobacter cloacae*, *Enterococcus aerogenes*, *Enterococcus faecalis*, *Hafnia alvei*, *Klebsiella pneumonia*, *Listeria monocytogenes*, *Pseudomonas aeroginosa*, *Proteus vulgaris*, *Salmonella* Anatum, *Salmonella* Arizona, *Salmonella* Choleraesuis, *Salmonella* Enteriditis, *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.

Preparation of lipopolysaccharide (LPS) extracts and analysis

Whole cell lysates of the reference standard strains, *E. coli* O118:H16 31w and *E. coli* O151:H10 880-67, as well as field strains O151:H26 99.1227 and O118:H16 98.0556, were used as the source of LPS for analysis by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The bacteria were grown overnight in LB broth at 37°C, and a 5-mL aliquot of a suspension of bacteria were used to isolate LPS according to the instructions of the manufacturer of LPS extraction Kit (iNtRON Biotechnology, Boca Raton, FL). Purified LPS from *E. coli* 0111:B4 (Sigma L2630, St. Louis, MO) and *Salmonella enterica* serotype Enteritidis (Sigma L6011) were used as LPS standards. Discontinuous SDS-PAGE was performed in a 0.75-mm vertical slab gel with 4.5% stacking and 15% separating gels. Sample volume ranged from 3 to 5 μL of LPS extract. Electrophoresis was carried out at 70 mA with two gels run simultaneously for approximately 2 to 3 hours. The temperature was maintained at 8°C with a circulating cooling system. Following electrophoresis, the gels were stained using the Bio-Rad Silver Stain Kit, following the manufacturer's instructions (Bio-Rad Laboratories, Hercules, CA).

Construction of 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 the JUMPSTART (named for Just Upstream of Many Polysaccharide-associated gene STARTs) and GND (6-phosphogluconate dehydrogenase gene) primer set targeting sequences that flank the *E. coli* O antigen gene clusters (Fratamico *et al*., 2003). The sequence of the JUMPSTART sense primer was 5′-ATTG GTAGCTGTAAGCCAAGGGCGGTAGCGT-3′, and the antisense GND primer sequence was 5′-CACTGCCATACCGACGACGCCGATCTG TTGCTTGG-3′ (Invitrogen Life Technologies, Inc., Carlsbad, CA). The long PCR conditions were as described previously (Fratamico *et al*., 2003). The long PCR products were verified on 0.8% agarose gels and purified according to instructions in the QIAquick PCR Purification Kit (Qiagen, Inc.). DNaseI digestion, shotgun cloning, and DNA sequencing were performed as described previously (Fratamico *et al*., 2003). To confirm the sequences of each of the O antigen gene clusters, 10 individual long PCR products were pooled together, and the DNA was resequenced using primers designed from different regions of the gene clusters. Sequencing data were assembled using Sequencher software (Gene Codes Corporation, Ann Arbor, MI), and gene annotation was performed as described previously (Fratamico *et al*., 2003). The HMMTOP program (Tusnády *et al*., 2001) was used to identify potential transmembrane segments from the amino acid sequences.

PCR specificity testing

Bacterial DNA used as template for the PCR assays was isolated as described previously (DebRoy *et al*., 2005). PCR primers (Table 1) were designed from the *E. coli* O118 *wzx* and *wzy* genes using the Primer3 software program. Multiplex PCR reactions using primers targeting the *E. coli* O118 *wzx* and *wzy* genes were performed in a RapidCycler (Idaho Technologies, Salt Lake City, UT) and analyzed as described previously (DebRoy *et al*., 2005). The cycling conditions were the following: denaturation at 94°C, annealing at 60°C, and extension at 72°C for 12 seconds.

Nucleotide sequence accession numbers

DNA sequences of the O antigen gene clusters of *E. coli* O118 and O151 (see below) were

Table 1. Polymerase Chain Reaction (PCR) Primers and Conditions for Amplifying the *wzx* and *wzy* Genes of *E. coli* O118 by Multiplex PCR

Primers target gene	*Primer sequence*	*Size of PCR product (bp)*	*PCR components*
O118wzx, *wzx*	5′-GTG GGA GTC TGA ATC AAG TTG CGA-3′ 5′-AGC AAC CTT ACC CAA TCC TAA GGG-3′	344	DebRoy *et al.*, 2005
O118wzy, *wzy*	5′-TGC AAG AGA TGG TAT TGA GCT GGG-3′ 5′-TCC TGA GCC AAT TTC TGT AGG TCG-3′	517	2mM $MgCl_2$

deposited into GenBank under the accession numbers of DQ990684 and DQ990685, respectively.

Multilocus sequence typing (MLST) of O118 and O151 strains

E. coli O118:H16 strain 98.0556, O118:H16 strain 31w, *E. coli* O151:H26 strain 99.1227, and O151:H10 strain 880-67 were analyzed by multilocus sequencing in an attempt to determine the overall relatedness of the two serogroups. An *E. coli* MLST scheme using seven conserved housekeeping genes (*asp*C, *clp*X, *fad*D, *icd*A, *lys*P, *mdh*, and *uid*A) was used to explore the genetic diversity and relatedness among these isolates. All isolates were grown overnight in LB broth at 37°C. In preparation for sequencing, genomic DNA was isolated from 1 mL of overnight culture using the Puregene DNA isolation kit (Gentra Systems Inc., Minneapolis, MN). DNA preparations were quantified and diluted to a final concentration of 100 ng/μL and stored at 4°C. Polymerase chain reaction amplification was performed using PCR Master Mix, 2×(Promega Corp., Madison, WI) with upstream and downstream primers at a final concentration of 0.5 mM, and 100 ng of template DNA. PCR was performed for 35 cycles under the following conditions: 1 minute of denaturation at 94°C, 1 minute of primer annealing at 57°C, and 30 seconds of extension at 72°C, with an initial denaturing step of 94°C for 2 minutes. Amplicons were electrophoresed on a 0.8% agarose gel and visualized. PCR products were purified using the QIAquick PCR Purification Kit. Cycle sequencing reactions were performed with CEQ dye terminator cycle sequencing kits (Beckman-Coulter, Fullerton, CA) with approximately 50 fmol of template and a final primer concentration of 2 μM. The thermal cycle was run for 30 cycles with the following parameters: 96°C for 20 seconds, 50°C for 20 seconds, and 60°C for 4 minutes. Reactions were purified using Sephadex columns and dried under vacuum centrifugation at room temperature. The samples were then rehydrated in 30 μL of formamide and sequenced using a Beckman CEQ 8000 (Beckman-Coulter) capillary sequencer. Sequences were concatenated and aligned with the SeqMan module in the DNAStar Lasergene (Lasergene, Madison, WI) computer software package. Consensus sequences were aligned with ClustalX (Thompson *et al.*, 1997) and the output files were modified for use in MEGA3 (Kumar *et al.*, 2004).

Results and Discussion

The genes that encode proteins within the *E. coli* O antigen gene clusters consist of three categories: nucleotide sugar biosynthesis, glycosyl transferase, and O antigen processing genes (Samuel and Reeves, 2003). Nucleotide sugar biosynthesis gene products are involved in the synthesis of the nucleotide sugar precursors in the O antigen, which occurs in the cytoplasm. Glycosyl transferases 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. The 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 has been 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 showing a high degree of

Table 2. Open Reading Frames (ORFs) in the O Antigen Gene Cluster of *E. coli* Serogroup O118

ORF	Proposed gene name	Location	No. of amino acids	Putative function	Most significant homology (accession no.)	% Identity/ % similarity
1	*wbhU*	143–1501	452	Pyrophosphorylase	Pyrophosphorylase [*Bacteroides thetaiotaomicron* VPI-5482] (AAO76758.1)	49/70
2	*wzx*	1506–2726	406	O-antigen flippase	Membrane protein involved in the export of O-antigen and teichoic acid [*Actinobacillus succinogenes* 130Z] (ZP_00731935.1)	31/50
3	*wbhV*	2719–3537	272	LicD-family phosphotransferase	Hypothetical protein CbeiDRAFT_1437 [*Clostridium beijerincki* NCIMB 8052] (ZP_00910509.1)	38/57
4	*wzy*	3544–4644	366	O-antigen polymerase	Wzy [*Escherichia coli*] (AAK64372.1)	32/53
5	*wbhW*	4637–5746	369	Glycosyl transferase	Glycosyltransferase [Leptospira borgpetersenii serovar Hardjo-bovis L550] (YP_797647.1)	42/61
6	*wbuX*	5743–6879	378	Amidotransferase	WbuX [*Escherichia coli*] (AAV74532.1)	84/92
7	*wbuY*	6876–7502	208	Glutamine amidotransferase	Glutamine amidotransferase [*Escherichia coli*] (AAX58765.1)	86/94
8	*wbuZ*	7495–8283	262	Imidazoleglycerol phosphate synthase	WbuZ [*Escherichia coli*] (AAV74534.1)	91/96
9	*fnlA*	8293–9330	345	L-fucosamine synthetase	FnlA [*Escherichia coli*] (AAV74544.1)	98/98
10	*fnlB*	9332–10,435	367	epimerase	FnlB [*Escherichia coli*] (AAV74545.1)	98/99
11	*fnlC*	10,435–11,565	376	C-2 epimerase	FnlC [*Escherichia coli*] (AAV74537.1)	99/99
12	*wbuB*	11,565–12,776	403	L-fucosamine transferase	L-fucosamine transferase [*Escherichia coli*] (AAX58770.1)	98/99
13	*wbuC*	12,763–13,161	132	unknown	WbuC [*Escherichia coli*] (AAV74539.1)	95/96

variation in sequence among different microorganisms, the action of Wzx is not very specific, while Wzy function is specific (Samuel and Reeves, 2003).

DNA sequences of 13,283 bases were obtained from the *E. coli* O118 O antigen gene cluster of strain 31w, which contained 13 ORFs (Table 2), all in the same transcriptional direction from *galF* to *gnd.* The deduced amino acid sequences from these ORFs were used to search the NCBI database for an indication of their possible functions. Gene names were assigned on the basis of the Bacterial Polysaccharide Gene Nomenclature system (http://www.mmb.usyd.edu.au/BPGD/).

Sequence analysis of the E. coli *O118 O antigen gene cluster*

Sugar biosynthetic pathway genes. The structure of the *E. coli* O118 O antigen polysaccharide has not been determined. ORF9, 10, and 11 were almost identical (98%, 99%, and 99% similarity) to FnlA, FnlB, and FnlC in the O antigen gene cluster of *E. coli* O145 (Feng *et al.*, 2005). These genes were also present in other O antigen gene clusters and were proposed to be involved in synthesis of UDP-L-FucNAc (D'Souza *et al.*, 2002; Feng *et al.*, 2005). The presence of these genes indicates that the O antigen polysaccharide of *E. coli* O118 may contain FucNAc.

Sugar transferase genes. Six transferases were identified based on their homology to known sugar transferases. ORF1 shared 70% similarity with the pyrophosphorylase gene in *Bacteroides thetaiotaomicron* VPI-5482. ORF3 showed 55% similarity to a hypothetical protein in *Clostridium beijerincki* NCIMB 8052. In addition, ORF3 showed homology to a number of putative LicD-family phosphotransferases; therefore, we propose that ORF3 may have phosphotransferase

activity. ORF5 shared 61% similarity with a glycosyltransferase in *Leptospira borgpetersenii*. ORF6 showed high homology with WbuX in *E. coli* O145, which encodes for an amidotransferase involved in the amination of L-FucNAc to synthesize L-FucNAm; therefore, ORF6 was named WbuX. ORF7 showed 94% homology to a glutamine amidotransferase (WbuY) in *E. coli* O145, which may serve as an ammonia tunnel to convey ammonia to another protein (Feng *et al*., 2005). ORF8 shared 96% similarity to the *wbuZ* gene in the O antigen gene cluster of *E. coli* O145. Even though this gene product is similar to imidazole glycerol phosphate synthase, its function is unknown. It was also proposed to be involved in the ammonia tunnel in *E. coli* (Feng *et al*., 2005). ORF12 shared 99% homology with L-fucosamine transferase, which is proposed to transfer L-FucNAm in *E. coli* O145 and in other O antigen gene clusters.

O antigen processing genes. ORF2 showed 50% similarity to a membrane protein involved in the export of O antigen and teichoic acid in *Actinobacillus succinogenes* 130Z. In addition, ORF2 showed moderate homology to a number of O antigen flippases (Wzx); therefore, it was named *wzx*. ORF4 showed 47% homology to the Wzy protein in *E. coli* O104; therefore, it was named *wzy*.

Identification of E. coli *O118 serogroup-specific genes and specificity testing*

Sequence analysis revealed that the *wzx* and *wzy* genes of the O antigen gene cluster of *E. coli* O118 shared the least similarity with other genes in GenBank, and this is consistent with previous studies that demonstrated that the *wzx* and *wzy* genes were serogroup specific. Primers targeting the *E. coli* O118 *wzx* and *wzy* genes were designed (Table 1) and used in PCR assays to determine the specificity for this serogroup. PCR reactions were performed to test the specificity against 167 *E. coli* standard strains, 47 non-O118 strains, as well as *E. coli* serogroup O118 (65 strains) strains isolated from humans, animals, food, or water. The PCR assays targeting the *wzx* and *wzy* genes of *E. coli* O118 were specific for this serogroup and could be potentially used for rapid diagnostic screening for *E. coli* O118. However, the PCR assays targeting the *E. coli* O118 *wzx* and *wzy* genes were also positive for products of the expected sizes using DNA from *E. coli* serogroup O151 strains, although no cross-reactivity was observed between these two serogroups by serotyping. This led us to compare the similarity in the DNA sequence of the O antigen gene clusters of *E. coli* serogroups O118 and O151. PCR primers were designed based on regions in the sequence of the O118 O antigen gene cluster sequence and were used to determine the DNA sequence of the *E. coli* O151 O antigen gene cluster using the long PCR product.

Nucleotide sequence analysis and antigenic diversity between O118 and O151 serotypes

A comparison of the *E. coli* O118 and O151 O antigen gene clusters revealed little variation at the nucleotide level. This is consistent with the finding that the restriction fragment length polymorphism patterns of *E. coli* O118 and O151 are indistinguishable (Coimbra *et al*., 2000). Nucleotide analysis of the O antigen gene cluster revealed three differing nucleotide sites with two being nonsynonymous substitutions. The nonsynonymous substitutions give the following amino acid changes: V → A (O118 → O151) at residue 292 in ORF4 and P → S (O118 → O151) at residue 88 in ORF10. Further analysis of ORFs 4 and 10 using the Jameson–Wolf antigenic index (Jameson and Wolf, 1988) predicted differences in the antigenic determinants near the sites of residue replacement. In ORF4 (Wzy), the antigenic index differed in the O118 and O151 proteins between residues 290 (threonine) and 294 (serine) that includes the variable site at 292. There is an example that a single nucleotide change in the *wzy* gene changed the LPS phenotype of *E. coli* O6 (Grozdanov *et al*., 2002). The LPS side chain was changed by a truncated Wzy protein caused by a point mutation in the *wzy* gene. Analysis of ORF10 predicts an antigenic index difference between two serine residues located at 82 and 92 with the variable amino acid occurring at residue 88. It is possible that these amino acid differences display sufficient antigenic difference to account for the difference in serotype. The fact that a single amino acid substitution in a mannosyl transferase converted the *E. coli* O9 polysaccha-

Table 3. Allelic Profiles of *E. coli* O118 and O151 Strains

Strain	*aspC*	*clpX*	*fadD*	*icdA*	*lysP*	*mdh*	*uidA*	*Clonal Group*
E. coli O118:H1698.0556	7	8	2	4	1	5	New allele	14
E. coli O118:H1631w	4	9	1	15	1	53	New allele	Not identified
E. coli O151:H2699.1227	3	3	1	4	1	1	New allele	Not identified
E. coli O151:H10880-67	4	12	79	15	New allele	8	New allele	Not identified

ride into O9a supports this notion (Kido and Kobayashi, 2000).

The MLST analysis of two O118 and two O151 isolates again revealed minor variation at the nucleotide level ranging from 0.2% variability in the *lysP* locus to 3.5% in the *clpX* locus. Over a total of 3753 bp, 61 sites (1.6%) differed between the four isolates. When translated, *aspC*, *clpX*, *fadD*, *icdA*, *lysP*, and *mdh* displayed no variability at the amino acid level; however, beta-D-glucuronidase encoded by *uidA* had three residue differences in a total of 192 amino acids. The nucleotide sequences for each locus were compared with the EcMLST database (Qi *et al*. 2004), and allele profiles were assigned for each isolate (Table 3). The allelic profile of the *E. coli* O118 strain 98.0556 matched six of the seven loci for clonal group 14, which is comprised of EHEC2 isolates. Three other O118 isolates within the EcMLST database are also in clonal group 14. The fact that O118 and O151 are similar at the nucleotide level for the 7 MLST genes and the O antigen gene clusters and almost identical with only five changes at the amino acid levels suggests a shared common ancestor for the two serogroups. Since the O118 serogroup is more prevalent taken together with the limited sequence diversity and well conserved amino acid sequences, it is possible that the O151 serogroup is a minor variant of the predominant O118 serogroup.

Analysis of E. coli *O118 and O151 LPS*

The LPS profiles of the two *E. coli* O118 strains, 31w and 98.0556, were the same (Fig. 1). However, the *E. coli* O118 banding pattern differed from that of the *E. coli* O151 strains (880-67 and 99.1227), although the differences between the O118 and O151 strains were not major. There were shifts in two bands indicated by the arrows in Fig. 1 between the O118 and O151 LPS profiles.

Conclusions

The O antigen gene clusters of an *E. coli* O118 strain and an *E. coli* O151 strain were sequenced, the genes were putatively named, and PCR assays targeting the *wzx* (O antigen flippase) and *wzy* (O antigen polymerase) genes were designed. These genes can be used as diagnostic markers for rapid identification of *E. coli* serogroups O118 and O151 by the PCR as an alternative to serotyping using antisera. However, additional testing using specific antisera would need to be performed to distinguish these two serogroups. Alternatively, a single nucleotide polymorphism–specific method targeting the base differences within the O antigen gene clusters of O118 and O151 can be designed to discriminate between these two serogroups. Recently, we developed a microarray-based *E. coli* typing assay for rapid identification of different serogroups of *E. coli* in a single platform (Liu and Fratamico, 2006). Different sequences

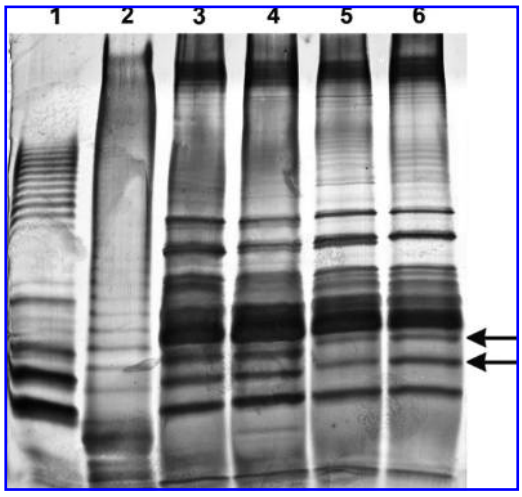

FIG. 1. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) of lipopolysaccharide (LPS) extracts. (**1**) LPS standard for *E. coli*. (**2**) LPS standard for *Salmonella*. (**3**) *E. coli* O118:H16 31w, (**4**) *E. coli* O118:H16 98.0556, (**5**) *E. coli* O151:H10 880-67, (**6**) *E. coli* O151:H26 99.1227.

in the O antigen gene clusters of several *E. coli* serogroups were spotted onto the microarrays, and specific signals were generated for each serogroup tested. Sequence information of additional O antigen gene clusters, including that from *E. coli* O118 will help us to expand our research on development of DNA microarrays for *E. coli* typing. Although there was no cross-reactivity between *E. coli* O118 and O151 using specific antisera, we found minor variation in the O antigen gene cluster sequences of both serogroups. *E. coli* O118 and O151 may share a common ancestor, and *E. coli* O151 may actually be a minor variant of *E. coli* O118. Further studies are needed to elucidate this possibility. The O antigen LPS has been associated with a number of biological phenomena including stress responses (Vines *et al*., 2005), swarming motility (Toguchi *et al*., 2000), virulence (Canals *et al*., 2006), and flagellum biogenesis (Abeyrathne *et al*., 2005). 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 thank Dr. Connie Briggs, Dr. David Needleman, and Ms. Laurie Fortis (U.S. Department of Agriculture, Agricultural Research Service, Eastern Regional Research Center) for their assistance in sequencing the O antigen gene clusters of *E. coli* O118 and O151 and Ms. Elisabeth Roberts at the *E. coli* Reference Center, Penn State for performing PCR testing. We also thank Mr. Paul Pierlott for photography and Mr. Robert Downs for word processing. Mention of trade names or commercial products is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

References

Abeyrathne PD, Daniels C, Poon KK, Matewish MJ, and Lam JS. Functional characterization of WaaL, a ligase associated with linking O antigen polysaccharide to the core of *Pseudomonas aeruginosa* lipopolysaccharide. J. Bacteriol. 2005;187:3002–3012.

Beutin L, Long Q, Feng L, Wang Q, Krause G, Leomil L, Jin Q, and Wang L. Development of PCR assays targeting the gene involved in symthesis and assembly of the new *Escherichia coli* O174 and O177 O antigens. J. Clin. Microbiol. 2005a;43:5143–5149.

Beutin L, Tao J, Feng L, Krause G, Zimmermann S, Gleier K, Xia Q, and Wang L. Sequence analysis of the *Escherichia coli* O15 antigen gene cluster and development of a PCR assay for rapid detection of intestinal and extraintestinal pathogenic *E. coli* O15 strains. J. Clin. Microbiol. 2005b;43:703–710.

Beutin L, Zimmermann S, and Gleier K. Human infections with Shiga toxin-producing *Escherichia coli* other than serogroup O157 in Germany. Emerg. Infect. Dis. 1998; 4:635–639.

Canals R, Jimenez N, Vilches S, Regue M, Merino S, and Tomas JM. The UDP N-acetylgalactosamine 4-epimerase gene is essential for mesophilic *Aeromonas hydrophila* serotype O34 virulence. Infect. Immun. 2006; 74:537–548.

Coimbra RS, Grimont F, Grimont F, Lenormand P, Burguiere P, Beutin L, and Grimont PA. Identification of *Escherichia coli* O-serogroups by restriction of the amplified O antigen gene cluster (rfb-RFLP). Res. Microbiol. 2000;151:639–654.

DebRoy C, Fratamico PM, Roberts E, Davis MA, and Liu Y. Development of PCR assays targeting genes in O antigen gene clusters for detection and identification of *Escherichia coli* O45 and O55 serogroups. Appl. Environ. Microbiol. 2005;71:4919–4924.

DebRoy C, Roberts E, Kundrat J, Davis MA, Briggs CE, and Fratamico PM. Detection of *Escherichia coli* serogroups O26 and O113 by PCR amplification of the *wzx* and *wzy* genes. Appl. Environ. Microbiol. 2004;70:1830–1832.

D'Souza JM, Wang L, and Reeves P. Sequence of the *Escherichia coli* O26 O antigen gene cluster and identification of O26 specific genes. Gene 2002;297:123–127.

Feng L, Senchenkova SN, Tao J, Shashkov AS, Liu B, Shevelev SD, Reeves PR, Xu J, Knirel YA, and Wang L. Structural and genetic characterization of enterohemorrhagic *Escherichia coli* O145 O antigen and development of an O145 serogroup-specific PCR assay. J. Bacteriol. 2005;187:758–764.

Fratamico PM, Briggs CE, Needle D, Chen CY, and DebRoy C. Sequence of the *Escherichia coli* O121 O antigen gene cluster and detection of enterohemorrhagic *E. coli* O121 by PCR amplification of the *wzx* and *wzy* genes. J. Clin. Microbiol. 2003;41:3379–3383.

Friedrich AW, Bielaszewska M, Zhang W-L, Pulz M, Kuczius T, Ammon A, and Karch H. *Escherichia coli* harboring Shiga toxin 2 gene variants: frequency and association with clinical symptoms. J. Infect. Dis. 2002; 185:74–84.

Grozdanov L, Zähringer U, Blum-Oehler G, Brade L, Henne A, Knirel YA, Schombel U, Schulze J, Sonnenborn U, Gottschalk G, Hacker J, Rietschel ET, and Dobrindt U. A single nucleotide exchange in the *wzy* gene is responsible for the semirough O6 lipopolysaccharide phenotype and serum sensitivity of *Escherichia coli* strain Nissle 1917. J. Bacteriol. 2002;184:5912–5925.

Hashimoto H, Mizukoshi K, Nishi M, Kawakita T, Hasui S, Kato Y, Ueno Y, Takeya R, Okuda N, and Takeda T. Epidemic of gastrointestinal tract infection including hemorrhagic colitis attributable to Shiga toxin 1-producing *Escherichia coli* O118:H2 at a junior high school in Japan. Pediatrics 1999;103:E2.

Jameson BA and Wolf H. The antigenic index: a novel algorithm for predicting antigenic determinants. CABIOS 1988;4:181–186.

Kido N and Kobayashi H. A single amino acid substitution in a mannosyltransferase, WbdA, converts the *Escherichia coli* O9 polysaccharide into O9a: generation of a new O-serotype group. J. Bacteriol. 2000;182:2567–2573.

Kiseleva BS, Raginskaia VP, Baturo AP, Lifshits MB, and Ratiner IUA. Antigenic relations between enteropathogenic *Escherichia coli* O151:K, *Citrobacter*, and *Hafnia*. Zh. Mikrobiol. Epidemiol. Immunobiol. 1981;4:85–88.

Kumar S, Tamura K, and Nei M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics 2004;5:150–163.

Liu Y and Fratamico PM. *Escherichia coli* O antigen typing using DNA microarrays. Mol. Cell. Probes 2006;20:239–244.

Ørskov I, Ørskov F, Jann B, and Jann K. Serology, chemistry and genetics of O and K antigens of *Escherichia coli*. Bacteriol. Rev. 1977;41:667–710.

Qi W, Lacher DW, Bumbaugh AC, Hyma KE, Ouellette LM, Large TM, Tarr CL, and Whittam TS. EcMLST: An Online Database for Multi Locus Sequence Typing of Pathogenic *Escherichia coli*. *csb*. In: 2004 IEEE Computational Systems Bioinformatics Conference (CSB'04), 2004, pp. 520–521.

Samuel G and Reeves P. Biosynthesis of O antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O antigen assembly. Carbohydrate Res. 2003;338:2503–2519.

Smith JL and Fratamico PM. Diarrhea-inducing *Escherichia coli*. In: *Foodborne Pathogens Microbiology and Molecular Biology*. Fratamico PM, Bhunia AK, Smith JL (eds). Norfolk, UK: Caister Academic Press, 2005, pp. 357–382.

Steinruck H, Mochmann H, Kontny I, and Werner U. *Escherichia coli* O151:K-:H10 ("krim") as etiologic agent of diarrhoea. Zentralbl. Bakteriol. A. 1980;248:335–344.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, and Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 1997;25:4876–4882.

Toguchi A, Siano M, Burkart M, and Harshey RM. Genetics of swarming motility in *Salmonella enterica* serovar typhimurium: critical role for lipopolysaccharide. J. Bacteriol. 2000;182:6308–6321.

Tusnády GE and Simon I. The HMMTOP transmembrane topology prediction server. Bioinformatics 2001;17:849–850.

Vines ED, Marolda CL, Balachandran A, and Valvano MA. Defective O antigen polymerization in tolA and pal mutants of *Escherichia coli* in response to extracytoplasmic stress. J. Bacteriol. 2005;187:3359–3368.

Wang L, Huskic S, Cisterne A, Rothemund D, and Reeves PR. The O antigen gene cluster of *Escherichia coli* O55:H7 and identification of a new UDP-GlcNAc C4 epimerase gene. J. Bacteriol. 2002;184:2620–2625.

Wieler LH, Schwanitz A, Vieler E, Busse B, Steinrück H, Kaper JB, and Baljer G. Virulence properties of Shiga toxin-producing *Escherichia coli* (STEC) strains of serogroup O118, a major group of STEC pathogens in calves. J. Clin. Microbiol. 1998;36:1604–1607.

Willshaw GA, Smith HR, Roberts D, Thirlwell J, Cheasty T, and Rowe B. Examination of raw beef products for the presence of Vero cytotoxin producing *Escherichia coli*, particularly those of serogroup O157. J. Appl. Bacteriol. 1993;75:420–426.

Address reprint requests to:
Yanhong Liu, Ph.D.
U.S. Department of Agriculture
Agricultural Research Service
Eastern Regional Research Center
600 East Mermaid Lane
Wyndmoor, PA 19038

E-mail: yanhong.liu@ars.usda.gov