Identification serologically, chemically and genetically of two *Escherichia coli* strains as candidates for new O serogroups

Min Chen, Anna M. Shpirt, Xi Guo, Alexander S. Shashkov, Yuan Zhuang, Lei Wang, Yuriy A. Knirel and Bin Liu

1Shanghai Municipal Center for Disease Control and Prevention, No. 1380, Zhong Shan Xi Road, Shanghai 200336, PR China
2N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Leninskii Prospekt 47, Moscow 119991, Russia
3TEDA Institute of Biological Sciences and Biotechnology, Nankai University, 23 Hongda Street, TEDA, Tianjin 300457, PR China

Correspondence
Bin Liu
liubin1981@nankai.edu.cn

Escherichia coli strains are normally identified by the combination of their O and H (and sometimes K) antigens, and serotyping based on the antigens is believed to be crucial for clinical detection and epidemiological investigation. Two *E. coli* strains, G5413 and G5287, were isolated from faecal samples of female patients with diarrhoea and were not agglutinated with any antisera that cover the well-known O serogroups of *E. coli*. We elucidated the O-polysaccharide (OPS) structures and analysed the O-antigen gene clusters of these bacteria. The OPS structure of G5413 established by monosaccharide analysis and NMR spectroscopy was found to be unique amongst known bacterial polysaccharide structures. The O-antigen gene cluster of this strain was sequenced and did not match sequence data with any of the 184 O serogroups that have been recognized internationally. Gene functions were tentatively assigned and were appropriate to the OPS structure. Based on these data, we suggest G5413 as a candidate for a new *E. coli* O serogroup. Both the OPS structure and O-antigen gene cluster of G5287 were identical to those of *E. coli* L-19, a candidate for another new O serogroup characterized by us recently. Recognition of these two provisional O serogroups increases the number of known O-antigen forms of *E. coli* to 186.

**INTRODUCTION**

*Escherichia coli* is one of the most widely studied organisms in the microbial world. Non-pathogenic strains of *E. coli* are part of the normal flora of the gastrointestinal tract; pathogenic strains have been implicated in causing a wide range of intestinal and extra-intestinal diseases in humans and animals (Kaper *et al.*, 2004). The O-antigen, or O-polysaccharide (OPS), which contains a number of oligosaccharide repeats (O-units), forms part of the lipopolysaccharide in the outer membrane of Gram-negative bacteria. The variation in sugar composition, their arrangement and linkages, makes the O-antigen one of the most variable constituents on the cell surface (Reeves & Wang, 2002).

O-antigen variability has been used as the basis of serotyping schemes established for many Gram-negative bacteria, the first being for *E. coli*, established by Fritz Kauffmann in the 1940s. In 2004, Scheutz added eight O serogroups to the existing scheme (Scheutz *et al.*, 2004). Furthermore, six new *E. coli* O serogroups (O182–O187) were internationally recognized recently (http://www.ssi.dk/English/SSI%20Diagnostica/Products%20from%20SSI%20Diagnostica/Antisera_antibodies/), meaning that 184 *E. coli* O serogroups are now known (Iguchi *et al.*, 2015). An *E. coli* strain, L-19, has been previously suggested by us as a candidate for a new O serogroup (Zdorovenko *et al.*, 2014).
Genes for O-antigen synthesis are normally located on the chromosome as an O-antigen gene cluster and genetic variations in the cluster are the major basis for the diversity of the O-antigen forms (Reeves & Wang, 2002). In most E. coli strains, the O-antigen gene cluster maps between the housekeeping genes galP and gnd (Bastin & Reeves, 1995). O-antigen synthesis genes commonly fall into three main classes: (1) nucleotide sugar precursor synthesis genes for sugars specific to the particular O-antigen; (2) sugar transferase genes associated with O-unit assembly that are specific for the donor and acceptor sugars, and generate a specific linkage between them; and (3) genes for O-unit processing and polymerization to the O-antigen (Liu et al., 2014).

In this study, we characterized serologically, chemically and genetically two non-typable E. coli strains, G5413 and G5284, from faecal samples from two female patients with diarrhoea. Strain G5413 was found not to be related to any of the known E. coli clones and is suggested to be a candidate for another new O serogroup. Strain G5284 shared the OPS structure and the O-antigen gene cluster with E. coli L-19, showing they are members of the same O serogroup. Recognition of these two provisional O serogroups increases the number of known O-antigen forms of E. coli to 186.

METHODS

Bacterial strains. Two E. coli strains (laboratory stock numbers G5413 and G5287) were isolated from faecal samples from two female patients (27 and 50 years old, respectively) with diarrhoea. Bacteria were grown to late exponential phase in 8 l Luria–Bertani broth using a 10 l BIOSTAT C-10 fermenter (B. Braun Biotech) under constant aeration at 37°C and pH 7.0. Bacterial cells were washed and dried as described previously (Robbins & Uchida, 1962).

Serotyping. Twenty-one sets of polyvalent antisera containing antisera against 174 E. coli O serogroups (not including O18ab, O28ac, O112ac, O93 and O182–O187) were purchased from Tianjin Biochip Corporation (Tianjin, China). Agglutination tests were performed according to the manufacturer’s instructions.

Isolation of LPS. LPS was isolated from bacterial cells of E. coli G5413 and G5287 by the phenol/water method (Westphal & Jann, 1965). The crude extract was dialysed without separation of the layers, and freed from nucleic acids and proteins by treatment with 50 % aqueous CCl3CO2H to pH 2 at 4°C. The supernatant was dialysed and freed from nucleic acids and proteins by treatment with 50 % aqueous CCl3CO2H to pH 2 at 4°C. The supernatant was dialysed and lyophilised. The yields of LPS were 8.3 and 6.4 % of dried cells mass for G5413 and G5287, respectively.

Isolation of OPS. Delipidation of LPS (163 mg from G5413 and 103 mg from G5287) was performed with 2 % aqueous HOAc at 100°C until precipitation of the lipids. The precipitate was removed by centrifugation (13 000 g, 20 min) and the supernatant was fractionated by gel-permeation chromatography on a Sephadex G-50 Superfine column (Amersham Biosciences) in 0.05 M pyridinium acetate buffer, pH 5.5, and monitored with a differential refractometer (Knauer). OPS was obtained in yields of 19.5 and 37 % of the LPS mass for G5413 and G5287, respectively.

Monosaccharide analysis. A sample of the OPS from G5413 (0.4 mg) was hydrolysed with 2 M CF3CO2H (120°C, 2 h). Neutral monosaccharides were identified by GLC of the alditol acetates on a Maestro 7820 GC instrument (Interlab) equipped with an HP-5ms column using a temperature program of 160°C (1 min) to 290°C at 7 °C min⁻¹. The absolute configurations of the monosaccharides were determined by GLC of the acetylated (S)-2-octyl glycosides as described previously (Leontein & Lönngren, 1993).

NMR spectroscopy. Samples were deuterium-exchanged by freeze-drying from 99.9 % D2O and then examined as solutions in 99.95 % D2O. NMR spectra were recorded on a Bruker Avance II 600 spectrometer at 55°C using internal sodium 3-(trimethylsilyl)propionate-2,2,3,3-d4 (δH 0, δC –1.6) as reference for calibration. 2D NMR spectra were obtained using standard Bruker software, and the Bruker TopSpin 2.1 program was used to acquire and process the NMR data. Mixing times of 100 and 150 ms were used in total correlation spectroscopy (TOCSY) and rotating-frame Nuclear Overhauser effect spectroscopy (ROESY) experiments, respectively.

Sequencing and analysis of genes. Genomic DNA was extracted from 1.5 ml overnight bacterial culture (~10⁹ c.f.u. ml⁻¹) using a DNA extraction kit according to the manufacturer’s instructions (Tiangen). Genomic DNA from G5413 and G5287 was sheared, polished and prepared using an Illumina Sample Preparation kit. Genomic libraries were constructed containing 500 bp paired-end inserts and sequencing was then performed with Solexa sequencing technologies (Illumina) to produce ~100-fold coverage. The reads obtained were assembled using the de novo genome assembly program Velvet to generate a multi-contig draft genome. The sequences of the O-antigen gene cluster, located between galP and gnd, were retrieved from the genome and analysed in silico as described previously (Feng et al., 2004).

RESULTS

Structure elucidation of the OPS of E. coli G5413. Mild acid degradation of the LPS of E. coli G5413 isolated from bacterial cells by the phenol/water procedure resulted in a high-molecular-mass polysaccharide, which was isolated by gel-permeation chromatography on Sephadex G-50 Superfine. Sugar analysis by GLC of the alditol acetates derived after full acid hydrolysis of the OPS revealed rhamnose (Rha), glucose (Glc) and 2-acetamido-2-deoxy-glucose (GlcNAc) in the ratio 1.4 : 1 : 2.5 (detector response). GLC of the acetylated (S)-2-octyl glycosides indicated that Rha had the l-configuration, and Glc and GlcNAc had the d-configuration.

The 1H-NMR and 13C-NMR spectra of the OPS from E. coli G5413 are shown in Fig. 1(a, b). The 13C-NMR spectrum of the OPS (Fig. 1b) showed signals of five anomic carbons in the region δ 97.0–104.4, three C-CH3OH groups (C-6 of Glc and GlcNAc) at δ 61.8–62.4, two C-CH3 groups (C-6 of Rha) at δ 17.8 and 17.9, two nitrogen-bearing carbons (C-2 of GlcNAc) at δ 54.4 and 57.3, 18 oxygen-bearing non-anomeric sugar ring carbons in the region δ 70.4–83.2, and two N-acetyl groups at δ 23.6, 23.9 (both CH3), 174.7 and 175.2 (both CO). The 1H-NMR spectrum of the OPS (Fig. 1a) contained signals of five anomic protons at δ 4.60–5.45, two C-CH3 groups (H-6 of Rha) at δ 1.24 and 1.31, other sugar protons in the region δ 3.38–4.21, and two N-acetyl groups at δ 2.02.
and 2.11. Therefore, the OPS had a pentasaccharide O-unit containing two residues each of D-GlcNAc (units A and D) and L-Rha (units B and E), and one D-Glc residue (unit C).

The $^1$H- and $^{13}$C-NMR spectra of the OPS were assigned using 2D $^1$H,$^1$H-correlation spectroscopy (COSY), TOCSY, ROESY, $^1$H,$^{13}$C-HSQC (heteronuclear single-quantum coherence), HSQC-nuclear Overhauser effect spectroscopy and heteronuclear multiple-bond correlation (HMBC) experiments (Table 1, Figs S1–S6, available in the online Supplementary Material). Based on intra-residue $^1$H,$^1$H correlations and $^3$J$_{HH}$ coupling constants, five spin systems were recognized for units A–E, all being in the pyranose form. The spin systems for amino sugars A and D were distinguished by correlations between the protons at the nitrogen-bearing carbons (H-2) and the corresponding carbons (C-2) at δ 3.67/57.3 and 4.05/54.4, respectively. The spin systems for Rha residues B and E were identified by the presence of the CH$_3$ groups (H-6).

The $\alpha$ configuration of units B and D and $\beta$ configuration of units A, C, and E were established by characteristic C-5 chemical shifts compared to published data (Lipkind et al., 1988).

The signals for C-2 of unit C, C-3 of units A and B, and C-3 and C-4 of unit D were shifted significantly downfield, as compared with their positions in the corresponding non-substituted monosaccharides, whereas the C-2–C-4 chemical shifts of unit E were close to those of free $\beta$-rhamnopyranose (Lipkind et al., 1988). These data showed that the OPS was branched with 3,4-disubstituted $\alpha$-GlcNAc D at the branching point and the side-chain $\beta$-Rha E. Finally, the monosaccharide sequence in the O-unit was determined by inter-residue correlations between the anomic protons and protons at the linkage carbons in the 2D ROESY spectrum, as well as between the anomic protons and linkage carbons, and vice versa, in the $^1$H,$^{13}$C-HMBC spectrum of the OPS (Table 2).

Based on the data obtained, it was concluded that the OPS of E. coli G5413 had the structure shown in Fig. 2. As judged by the Bacterial Carbohydrate Structure Database (www.csdb.glycoscience.ru/bacterial), this structure is unique amongst known structures of bacterial polysaccharides.

### Characterization of the O-antigen gene cluster of E. coli G5413

Ten ORFs were found in the O-antigen gene cluster of E. coli G5413. All ORFs mapped between the housekeeping genes galF and gnd, and had the same transcriptional direction from the former to the latter (Table 3, Fig. 3).

L-Rha is widely distributed in O-antigens of bacteria, and the four $rml$ genes involved in its biosynthesis are usually grouped together and easily identified in a range of species. In E. coli and Salmonella, the genes are generally located in the order $rmlB$, $rmlD$, $rmlA$ and $rmlC$ at the 5' end of the

### Table 1. $^1$H- and $^{13}$C-NMR chemical shifts (δ, ppm) of the OPS from E. coli G5413

The chemical shifts for the N-acetyl groups are δ$_H$ 2.02 and 2.11; δ$_C$ 23.6, 23.9 (both Me), 174.7 and 175.2 (both CO).

<table>
<thead>
<tr>
<th>Sugar residue</th>
<th>H-1</th>
<th>H-2</th>
<th>H-3</th>
<th>H-4</th>
<th>H-5</th>
<th>H-6 (6a, 6b)</th>
<th>C-1</th>
<th>C-2</th>
<th>C-3</th>
<th>C-4</th>
<th>C-5</th>
<th>C-6</th>
</tr>
</thead>
<tbody>
<tr>
<td>→3)-β-D-GlcNAc-(1→ A</td>
<td>4.60</td>
<td>3.67</td>
<td>3.67</td>
<td>3.42</td>
<td>3.41</td>
<td>3.75, 3.97</td>
<td>101.1</td>
<td>57.3</td>
<td>83.2</td>
<td>70.4</td>
<td>77.1</td>
<td>62.4</td>
</tr>
<tr>
<td>→3)-α-L-Rhap-(1→ B</td>
<td>4.89</td>
<td>3.99</td>
<td>3.75</td>
<td>3.48</td>
<td>3.95</td>
<td>1.24</td>
<td>102.3</td>
<td>71.3</td>
<td>81.4</td>
<td>71.8</td>
<td>70.4</td>
<td>17.8</td>
</tr>
<tr>
<td>→2)-β-D-Glc-(1→ C</td>
<td>4.75</td>
<td>3.47</td>
<td>3.59</td>
<td>3.46</td>
<td>3.44</td>
<td>3.76, 3.88</td>
<td>104.4</td>
<td>77.8</td>
<td>76.0</td>
<td>70.3</td>
<td>76.8</td>
<td>61.8</td>
</tr>
<tr>
<td>→3,4)-α-L-GlcNAc-(1→ D</td>
<td>5.45</td>
<td>4.05</td>
<td>4.21</td>
<td>3.73</td>
<td>4.07</td>
<td>3.85, 3.88</td>
<td>97.0</td>
<td>54.4</td>
<td>76.3</td>
<td>76.6</td>
<td>72.0</td>
<td>61.8</td>
</tr>
<tr>
<td>β-L-Rhap-(1→ E</td>
<td>4.87</td>
<td>4.19</td>
<td>3.57</td>
<td>3.38</td>
<td>3.39</td>
<td>1.31</td>
<td>102.4</td>
<td>71.6</td>
<td>74.1</td>
<td>73.3</td>
<td>73.5</td>
<td>17.9</td>
</tr>
</tbody>
</table>
Table 2. Correlations for H-1 and C-1 in the 2D ROESY and $^1$H,$^1$C-HMBC spectra of the OPS of E. coli G5413

Sugar abbreviations: A, $\beta$-GlcNAc; B, $\alpha$-Rhap; C, $\beta$-GlcP; D, $\alpha$-GlcPNAc; E, $\beta$-Rhap.

<table>
<thead>
<tr>
<th>Anomeric atom in sugar residue ($\delta$)</th>
<th>ROESY</th>
<th>HMBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>A H-1 (4.60)</td>
<td>D H-3 (4.21), A H-3 (3.67), H-5 (3.41)</td>
<td>D C-3 (76.3)</td>
</tr>
<tr>
<td>A C-1 (101.1)</td>
<td>A H-3 (3.67)</td>
<td>D H-3 (4.21), A H-2 (3.67)</td>
</tr>
<tr>
<td>B H-1 (4.89)</td>
<td>B H-2 (3.99), H-3 (3.75), C H-3 (3.59), H-5 (3.44)</td>
<td>A C-3 (83.2), B C-3 (81.4), C-5 (70.4)</td>
</tr>
<tr>
<td>C H-1 (4.75)</td>
<td>B H-3 (3.75), C H-3 (3.59), H-5 (3.44)</td>
<td>B C-3 (81.4)</td>
</tr>
<tr>
<td>C C-1 (104.4)</td>
<td>B H-3 (3.75), C H-2 (3.47)</td>
<td>B C-3 (81.4)</td>
</tr>
<tr>
<td>D H-1 (5.45)</td>
<td>C H-2 (3.47), D H-2 (4.05)</td>
<td>C C-2 (77.8), D C-5 (72.0)</td>
</tr>
<tr>
<td>D C-1 (97.0)</td>
<td>C H-2 (3.47)</td>
<td>C H-2 (3.47)</td>
</tr>
<tr>
<td>E H-1 (4.87)</td>
<td>D H-4 (3.73), E H-2 (4.21), H-3 (3.57), H-5 (3.39)</td>
<td>D C-4 (76.6)</td>
</tr>
<tr>
<td>E C-1 (102.4)</td>
<td>D H-4 (3.73)</td>
<td>D H-4 (3.73)</td>
</tr>
</tbody>
</table>

O-antigen gene cluster (Li & Reeves, 2000). Four proteins encoded by orf1–orf4 shared high identity levels (83–100 %) with known RmlB, RmlD, RmlA and RmlC of other *E. coli* strains. Therefore, orf1–orf4 were named *rmlB*, *rmlD*, *rmlA* and *rmlC*, respectively.

In addition to $l$-Rha, the G5413 O-unit included one Glc and two GlcNAc residues, which are common sugars in bacteria. Genes for the synthesis of their nucleotide precursors are usually located outside O-antigen gene clusters (Liu *et al.*, 2008). The gene *wecA*, whose product mediates transfer of GlcNAc-1-phosphate to the undecaprenyl phosphate carrier to initiate O-antigen synthesis, is present in the enterobacterial common antigen gene cluster (Rush *et al.*, 2010) and is not duplicated in the O-antigen gene cluster.

orf5 and orf7 were putative O-antigen processing genes in G5413. Orf5 shared 28–50 % identity with Wzx of other strains, which is involved in flipping the O-unit across the cytoplasmic membrane. Furthermore, Orf5 had 11 well-proportioned transmembrane domains, which is a typical topological characteristic of Wzx (Liu *et al.*, 1996). Orf7 shared 23–31 % identity with Wzy, the O-antigen polymerase of other strains, and had 10 well-proportioned transmembrane domains, suggesting the classical topological characteristics of Wzy (Feng *et al.*, 2004). Therefore, orf5 and orf7 were named *wzx* and *wzy*, respectively.

Four genes, orf6, orf8, orf9 and orf10, were predicted to be glycosyltransferase genes involved in the assembly of the G5413 O-unit. Proteins encoded by orf6, orf8 and orf9 belonged to glycosyltransferase family 1 (Pfam ID: PF00534) and shared 49, 39 and 42 % identity with glycosyltransferases of other strains, respectively. However, their particular functions could not be inferred by homology comparison. Orf10 belonged to glycosyltransferase family 2 (Pfam ID: PF00535) and shared 51 % identity with dTDP-rhamnosyltransferase of a strain of *Shigella flexneri* – a species closely related to *E. coli* (Liu *et al.*, 2008). Therefore, orf10 putatively encoded one of two dTDP-rhamnosyltransferases involved in G5413 O-unit assembly.

---

Fig. 2. Structures of the OPSs of *E. coli* G5413 and G5287. The latter structure is identical to that of *E. coli* L-19 (Zdorovenko *et al.*, 2014). d-GlcA6(GroN), 2-(d-glucuronoylamino)-1,3-propanediol.

---

E. coli G5413

\[
\begin{align*}
\rightarrow3)-\alpha-d-Glc\text{-}N\text{Ac}-(1\rightarrow2)-\beta-d-Glc\text{-}p-(1\rightarrow3)-\alpha-l-Rhap-(1\rightarrow3)-\beta-d-Glc\text{-}N\text{Ac}-(1\rightarrow4) \\
\rightarrow1)-\beta-d-Glc\text{-}pA6(\text{GroN})-(1\rightarrow4)-\beta-d-Glc\text{-}p-(1\rightarrow4)-\beta-d-Glc\text{-}pA-(1\rightarrow3)-\beta-d-Glc\text{-}pN\text{Ac}-(1\rightarrow3)
\end{align*}
\]

E. coli G5287 and L-19

\[
\begin{align*}
\rightarrow3)-\beta-d-Glc\text{-}pA6(\text{GroN})-(1\rightarrow4)-\beta-d-Glc\text{-}p-(1\rightarrow4)-\beta-d-Glc\text{-}pA-(1\rightarrow3)-\beta-d-Glc\text{-}pN\text{Ac}-(1\rightarrow3)
\end{align*}
\]
<table>
<thead>
<tr>
<th>Gene</th>
<th>Position of gene</th>
<th>G + C content (%)</th>
<th>Conserved domain(s) (Pfam ID), BLAST E value</th>
<th>Similar protein(s), strain(s) (GenBank accession no.)</th>
<th>Identical aa (%)/Similar aa (%) (no. of aa overlap)</th>
<th>Putative function of protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>rmlB</td>
<td>1279–2364</td>
<td>43.37</td>
<td>NAD-dependent epimerase/dehydratase (PF01370), $E=8.9 \times 10^{-81}$</td>
<td>dTDP-glucose 4,6-dehydratase 1, <em>E. coli</em> (WP_001549387)</td>
<td>99/100 (361)</td>
<td>dTDP-glucose 4,6-dehydratase</td>
</tr>
<tr>
<td>rmlD</td>
<td>2364–3263</td>
<td>46.77</td>
<td>dTDP-4-dehydrorhamnose reductase (PF04321), $E=2.9 \times 10^{-110}$</td>
<td>dTDP-4-dehydrorhamnose reductase, <em>E. coli</em> (WP_001459897)</td>
<td>100/100 (299)</td>
<td>dTDP-4-dehydrorhamnose reductase</td>
</tr>
<tr>
<td>rmlA</td>
<td>3321–4199</td>
<td>43.23</td>
<td>Nucleotidyltransferase (PF00483), $E=3.6 \times 10^{-23}$</td>
<td>Glucose 1-phosphate thymidylyltransferase, <em>E. coli</em> (WP_001459896)</td>
<td>100/100 (292)</td>
<td>Glucose 1-phosphate thymidylyltransferase</td>
</tr>
<tr>
<td>rmlC</td>
<td>4204–4734</td>
<td>36.34</td>
<td>dTDP-4-dehydrorhamnose 3,5-epimerase-related (PF00908), $E=2.4 \times 10^{-80}$</td>
<td>dTDP-4-dehydrorhamnose 3,5-epimerase, <em>E. coli</em> (WP_001767433)</td>
<td>83/88 (175)</td>
<td>dTDP-4-dehydrorhamnose 3,5-epimerase</td>
</tr>
<tr>
<td>wzx</td>
<td>4764–5987</td>
<td>30.22</td>
<td>Polysaccharide biosynthesis protein (PF01943), $E=3.2 \times 10^{-49}$</td>
<td>Wzx, <em>E. coli</em> (ACA24819)</td>
<td>50/71 (400)</td>
<td>O-antigen flippase</td>
</tr>
<tr>
<td>orf6</td>
<td>6164–7342</td>
<td>30.46</td>
<td>Glycosyltransferase WbsX family protein, <em>E. coli</em> (WP_001666533)</td>
<td>Glycosyltransferase</td>
<td>49/66 (351)</td>
<td>Glycosyltransferase</td>
</tr>
<tr>
<td>wzy</td>
<td>7305–8339</td>
<td>27.63</td>
<td>O-antigen polymerase, <em>Pseudoalteromonas piscicida</em> (WP_010368537)</td>
<td>O-antigen polymerase</td>
<td>31/54 (323)</td>
<td>O-antigen polymerase</td>
</tr>
<tr>
<td>orf8</td>
<td>8398–9432</td>
<td>31.2</td>
<td>Glycosyltransferase family 1 (PF00534), $E=9.2 \times 10^{-24}$</td>
<td>Group 1 glycosyltransferase, <em>Shewanella baltica</em> OS155 (YP_001051234)</td>
<td>39/58 (339)</td>
<td>Glycosyltransferase family 1</td>
</tr>
<tr>
<td>or9</td>
<td>9518–10258</td>
<td>32.11</td>
<td>Glycosyltransferase family 2 (PF00535), $E=8.2 \times 10^{-29}$</td>
<td>Glycosyltransferase family 2, <em>Acinetobacter baumannii</em> (WP_029754861)</td>
<td>42/61 (235)</td>
<td>Putative glycosyltransferase</td>
</tr>
<tr>
<td>orf10</td>
<td>10283–11149</td>
<td>31.25</td>
<td>Glycosyltransferase family 2 (PF00535), $E=1.1 \times 10^{-16}$</td>
<td>dTDP-rhamnosyltransferase, <em>Shigella flexneri</em> (WP_005049197)</td>
<td>51/69 (291)</td>
<td>dTDP-rhamnosyltransferase</td>
</tr>
</tbody>
</table>
**DISCUSSION**

Strains of *E. coli*, both commensal and pathogenic, are normally identified by a combination of their O and H (and sometimes K) antigens. In spite of the recent development of molecular typing methods, serotyping based on the O- and H-antigens remains the ‘gold standard’ for detecting pathogenic *E. coli* strains in clinical specimens, foods and environmental samples, as well as for understanding the epidemiology of the pathogen (Wang et al., 2010). For example, O157 is a leading O serogroup associated with enterohaemorrhagic *E. coli* and is a significant food-borne pathogen worldwide (Armstrong et al., 1996). *E. coli* strains are also being identified as agents of emerging infectious diseases, such as a novel shiga toxin-producing *E. coli* strain (O104:H4), which caused a widespread and severe food-borne disease outbreak in Germany in 2011 (Muniesa et al., 2012).

In the 1940s, the O-antigenic scheme of *E. coli* strains was first presented by Kauffmann. Orskov *et al.* established a comprehensive serotyping scheme for 164 O serogroups in 1977, which was the basis for O classification for epidemiological surveillance. Since then, *E. coli* strains with different O-antigen forms have been detected and 184 O serogroups of *E. coli* have been recognized to date (DebRoy *et al.*, 2011; Iguchi *et al.*, 2015). These observations suggest that *E. coli* clones with more O-antigen types may exist and further O serogroups will emerge.

![Diagram of the O-antigen gene cluster of *E. coli* G5413](image)

**Fig. 3.** Organization of the O-antigen gene cluster of *E. coli* G5413.

**ACKNOWLEDGEMENTS**

Work in the laboratory of L. W. and B. L. was supported by the International Science & Technology Cooperation Program of China (2012DFG31680 and 2013DFR30640), National Key Program for Infectious Diseases of China (2013ZX10004216-001 and 2013ZX10004221-003), National Natural Science Foundation of China Program (31270003, 81471904, 31571259, 31270133 and 31470194) and Research Project of Chinese Ministry of Education (NO. 113015A). Work in the laboratory of Y. A. K. was supported by the Russian Science Foundation (project 14-14-01042).

**REFERENCES**


Edited by: M. Stevens