

# Structural and genetic evidence that the *Escherichia coli* O148 O antigen is the precursor of the *Shigella dysenteriae* type 1 O antigen and identification of a glucosyltransferase gene

Lu Feng,<sup>1,2†</sup> Andrei V. Perepelov,<sup>3†</sup> Guang Zhao,<sup>1,2†</sup> Sergei D. Shevelev,<sup>3</sup> Quan Wang,<sup>1,2</sup> Sof'ya N. Senchenkova,<sup>3</sup> Alexander S. Shashkov,<sup>3</sup> Yunqi Geng,<sup>4</sup> Peter R. Reeves,<sup>5</sup> Yuriy A. Knirel<sup>3</sup> and Lei Wang<sup>1,2</sup>

## Correspondence

Lei Wang

wanglei@nankai.edu.cn

<sup>1</sup>TEDA School of Biological Sciences and Biotechnology, Nankai University, 23 HongDa Street, TEDA, Tianjin 300457, P. R. China

<sup>2</sup>Tianjin Key Laboratory for Microbial Functional Genomics, Nankai University, 23 HongDa Street, TEDA, Tianjin 300457, P. R. China

<sup>3</sup>N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 119991 Moscow, Russian Federation

<sup>4</sup>College of Life Sciences, Nankai University, Tianjin 300071, P. R. China

<sup>5</sup>School of Molecular and Microbial Biosciences (G08), University of Sydney, Sydney, NSW 2006, Australia

*Shigella dysenteriae* type 1 is the most virulent serotype of *Shigella*. Enterotoxigenic *Escherichia coli* O148 is pathogenic and can cause diarrhoea. The following structure was established for the tetrasaccharide repeating unit of the *E. coli* O148 O antigen:  $\rightarrow 3\text{-}\alpha\text{-L-Rhap-(1}\rightarrow 3\text{-}\alpha\text{-L-Rhap-(1}\rightarrow 2\text{-}\alpha\text{-D-Glcp-(1}\rightarrow 3\text{-}\alpha\text{-D-GlcpNAc-(1}\rightarrow$ . This differs from the structure reported earlier for *S. dysenteriae* type 1 by having a glucose (Glc) residue in place of a galactose (Gal) residue. The two bacteria also have the same genes for O antigen synthesis, with the same organization and high level of DNA identity, except that in *S. dysenteriae* type 1 *wbbG* is interrupted by a deletion, and a galactosyltransferase gene *wbbP* located on a plasmid is responsible for the transfer of galactose to make a novel antigenic epitope of the O antigen. The *S. dysenteriae* type 1 O antigen was reconstructed by replacing the *E. coli* O148 *wbbG* gene with the *wbbP* gene, and it had the LPS structure and antigenic properties of *S. dysenteriae* type 1, indicating that the *S. dysenteriae* type 1 O antigen evolved from that of *E. coli* O148. It was also confirmed that *wbbG* of *E. coli* O148 is a glucosyltransferase gene, and two serotype-specific genes of *E. coli* O148 and *S. dysenteriae* type 1 were identified.

Received 10 August 2006

Revised 11 September 2006

Accepted 14 September 2006

## INTRODUCTION

*Escherichia coli* serotypes are normally classified by a combination of their O and H (and sometimes K) antigens. For shigellae, only the O antigen classification system is used, as they lack H and K antigens. Based on multilocus enzyme electrophoresis (MLEE) and analysis of house-keeping gene sequences, it was shown that *Shigella* species in

reality are *E. coli* serotypes. At least 166 O antigen types have been recognized in *E. coli* and 33 in *Shigella*. Thirteen of them were common to both based on cross-reactions summarized by Ewing (1986) and chemical structural data (Parolis *et al.*, 1997), which also indicates the close relationship between *Shigella* and *E. coli*. It has been proposed that *Shigella* originated from *E. coli* in the Neolithic about 10 000 years ago when agriculture had been developed (Ochman *et al.*, 1983; Pupo *et al.*, 2000). The O antigen forms unique to *Shigella* were proposed to have been obtained since the *Shigella* mode of pathogenicity arose in *E. coli* (Wang *et al.*, 2001b).

The O antigen, as part of the LPS in the outer membrane of Gram-negative bacteria, is a major target of both the immune system and bacteriophages, and plays an important

†These authors contributed equally to this work.

**Abbreviations:** CAT, chloramphenicol acetyltransferase; HMBC, heteronuclear multiple band correlation; HSQC, heteronuclear single quantum correlation; MLEE, multilocus enzyme electrophoresis; ROESY, rotating Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy.

The GenBank/EMBL/DDBJ accession number for the sequence reported in this paper is DQ167407.

role in the bacterium–host interplay. It is one of the most variable cell constituents and also plays an important role in virulence. Genes for O antigen synthesis are normally clustered between *galF* and *gnd* in the chromosome of *E. coli*.

Of *Shigella* serotypes, *S. dysenteriae* type 1 was the first described, and it attracts special attention for frequency of epidemics, the severity of symptoms, high attack rate, high case-fatality rate, and various complications (Bennish *et al.*, 1990). It was responsible for large dysentery epidemics in Guatemala and other parts of Central America (Mendizabal-Morris *et al.*, 1971), Bangladesh (Ronsmans *et al.*, 1988), Zaire (Group, 1995), Kenya (Iijima *et al.*, 1995), and recently West Africa (Guerin *et al.*, 2003) and India (Pazhani *et al.*, 2004). The O antigen of *S. dysenteriae* type 1 is essential for virulence, and there is indirect evidence that antibodies against the O antigen are protective (Passwell *et al.*, 2001; Robbins *et al.*, 1992; Taylor *et al.*, 1993). The O antigen consists of tetrasaccharide repeating units of the following structure:  $\rightarrow 3$ - $\alpha$ -L-Rhap-(1 $\rightarrow$ 3)- $\alpha$ -L-Rhap-(1 $\rightarrow$ 2)- $\alpha$ -D-Galp-(1 $\rightarrow$ 3)- $\alpha$ -D-GlcpNAc-(1 $\rightarrow$ ) (Dmitriev *et al.*, 1976). The enzymes for the biosynthesis of O antigen are encoded by an O antigen gene cluster and the *wbbP* gene (previously known as *rfpB*), located on the chromosome near the *his* locus and a 9 kb plasmid pHW400, respectively (Gohmann *et al.*, 1994). The *wbbP* gene was shown to encode a membrane-bound galactosyltransferase, adding the second sugar of the structure shown above. A number of *E. coli* and *Shigella* O antigen gene clusters have been cloned and characterized (Reeves & Wang, 2002), with *Shigella sonnei* the only other example with genes essential for O antigen synthesis being on a plasmid (Shepherd *et al.*, 2000). The presence of this essential gene on a plasmid indicated that the galactose component of the *S. dysenteriae* type 1 O antigen was gained recently.

*E. coli* O148:H28 is one of the most common causes of diarrhoea in children in developing countries as well as in travellers to these areas (Black, 1986). In this study, the *E. coli* O148 O antigen repeating unit was shown to differ from that of *S. dysenteriae* type 1 only by the presence of a glucose residue in place of the galactose residue. The genes and organization of the O antigen gene cluster were shown to be the same in both bacteria, except that *wbbG* is interrupted by a deletion in *S. dysenteriae* type 1, which together with the presence in *S. dysenteriae* type 1 of *wbbP* accounts for the difference in structure. After deleting *wbbG* in *E. coli* O148 and cloning the *S. dysenteriae* type 1 *wbbP* gene into the O antigen deficient mutant strain, the construct strain was shown to have the same O antigen structure as *S. dysenteriae* type 1. Therefore, *wbbG* was shown to be responsible for transferring glucose to *N*-acetylglucosamine (GlcNAc) in *E. coli* O148 and it also appears that *S. dysenteriae* type 1 gained its current O antigen gene cluster from *E. coli* O148. Genes specific to *E. coli* O148 and *S. dysenteriae* type 1 were identified by screening strains representing 186 *E. coli* serotypes including *Shigella*.

## METHODS

**Bacterial strains and growth conditions.** All strains were grown on liquid or solid Luria–Bertani medium. The *E. coli* O148:H28 type strain E 519-66 (Ewing, 1986) and *S. dysenteriae* type 1 type strain LSPQ3472 (Pupo *et al.*, 2000) were from the Institute of Medical and Veterinary Science, Adelaide, Australia. All other *Shigella* and *E. coli* strains used were previously described (Feng *et al.*, 2004). *E. coli* DH5 $\alpha$  was used as a host strain for plasmids throughout the study. When required, antibiotics were included at the following concentration: ampicillin 100  $\mu$ g ml<sup>-1</sup> and chloramphenicol 10  $\mu$ g ml<sup>-1</sup>.

**Construction of a random DNase I shotgun bank.** Chromosomal DNA was prepared as previously described (Bastin & Reeves, 1995). Primers wl-1098 (5'-ATTGGTAGCTGTAAGCCAAGGGCGGTAGCGT-3') and wl-1524 (5'-TAGTCGCGCTGNGCCTG-GATTAAGTTCGC-3'), based on the JUMPStart and *gnd* gene, respectively, were used to amplify the *E. coli* O148 O antigen gene cluster by using the Expand Long Template PCR system (Roche). PCR amplifications were in a final volume of 50  $\mu$ l containing 1  $\mu$ l purified DNA, 10 mM dNTP mix and 50 pmol each primer. The PCR conditions were as follows: 1 cycle 94 °C for 2 min, 30 cycles 94 °C for 15 s, 60 °C for 30 s, 68 °C for 15 min, final extension 68 °C for 5 min. To limit any PCR errors, five individual PCR products were pooled. The PCR products were digested with DNase I, and the resulting DNA fragments were cloned into pGEM-T Easy to produce a bank by using the method described previously (Wang & Reeves, 1998).

**Sequencing and analysis.** The plasmid DNA template used for sequencing was prepared by the method of Sambrook *et al.* (1989). Sequencing was carried out with an ABI PRISM 3730 automated DNA sequencer using ABI BigDye terminator chemistry. The Staden package (Staden, 1996) and the Artemis program (Rutherford *et al.*, 2000) were used for sequence assembly and gene annotation, respectively. The BLOCKMAKER program (Henikoff *et al.*, 1995) was used for searching conserved motifs. BLAST and PSI-BLAST (Altschul *et al.*, 1997) were used for searching databases, including GenBank and the Pfam protein motif database (Bateman *et al.*, 2002), for possible gene functions. The algorithm of Eisenberg *et al.* (1984) was used to identify potential transmembrane segments. Sequence alignment and comparison were done with the CLUSTALW program (Thompson *et al.*, 1994).

**Construction of an *E. coli* O148 *wbbG* knockout mutant.** The *wbbG* gene was replaced by a chloramphenicol acetyltransferase (CAT) gene by using the Red recombination system of phage lambda (Datsenko & Wanner, 2000; Yu *et al.*, 2000). The CAT gene was PCR amplified from plasmid pKK232-8 (Pharmacia) by using primers binding to the 5' and 3' ends of the gene, and each primer carried 40 bp based on the *E. coli* O148 DNA which flanks *wbbG*: wl-1333 (5'-ATAATACATATCCAAGTGTGCCGAAAA-TGTCTGGGGTCTTCAGGAGCTAAGGAAGCTA-3') and wl-1334 (5'-TTCAACCATGATATCTTTGGTGAAATGCTTTAAGACATATAATTACGCCCGCCCTGCCA-3'). The *wbbG*::*cat* PCR product was gel-purified using the UNIQ-10 Gel Extraction Kit (Sangon, Shanghai, China) and transformed into an *E. coli* O148 strain carrying pKD20. Chloramphenicol-resistant transformants were selected after induction of the Red genes by the protocol described by Datsenko & Wanner (2000). Gene replacement was verified by PCR and sequencing using the primer pairs wl-1337 (5'-CGTATGCGTTCGTATGGC-3')/wl-1335 (5'-TCTTTACGATGCGATTGG-3') and wl-1336 (5'-GGATAGTGTTCACCCTTGT-3')/wl-1338 (5'-AGACGGTATAACCGAC-3'), which are located in *cat* and *wbbG* flanking genes. One positive colony was retained as *E. coli* O148 strain H1229.

**Cloning of the *S. dysenteriae* type 1 *wbbP* gene into *E. coli* O148 derivative H1229.** The *wbbP* gene and its own promoter were PCR-amplified from the *S. dysenteriae* type 1 plasmid using primer pair wl-4109 (5'-ACTGGAATTCACATAATTTTTACT-TATCAA-3') and wl-2152 (5'-AGTGGGATCCGAATGAATCAGAGCCGC-3') bearing *EcoRI* and *BamHI* restriction sites (not present elsewhere in the amplification product). The PCR product was digested with *EcoRI* and *BamHI* and ligated into the vector pUC19 digested using the same two enzymes. The plasmid pLW1059 was transformed into *E. coli* O148 strain H1229 and a positive colony was retained as *E. coli* O148 strain H1230.

**Electroporation.** All electroporations were carried out in a Bio-Rad GenePulser apparatus using 0.2 cm gap cuvettes. Preparation of *E. coli* electrocompetent cells and electroporation were performed using standard protocols (Ausubel *et al.*, 1999). The *E. coli* cultures electroporated with the *wbbG::cat* PCR product were incubated at 37 °C for 3 h before plating on selective plates.

**Serological tests.** The production of *S. dysenteriae* type 1 O antigen was determined by slide agglutination with *S. dysenteriae* type 1 type-specific antiserum. A fresh overnight bacterial culture in liquid Luria-Bertani medium was used as the antigen for slide agglutination tests.

**Bacterial cultivation and extraction of the O antigen.** LPS of *E. coli* O148 and *E. coli* O148 strain H1230 (1047 and 737 mg, respectively) was extracted from dried cells (12.6 and 8.9 g, respectively) by the phenol/water method (Wang *et al.*, 2001a) and purified by precipitation of nucleic acids and proteins with aqueous 50% trichloroacetic acid as described (Wang *et al.*, 1998). Delipidation of LPS of *E. coli* O148 and mutant strain H1230 (170 and 160 mg, respectively) was performed with aqueous 2% acetic acid (6 ml) at 100 °C until lipid A precipitation. The precipitate was removed by centrifugation (13 000 g, 20 min), and the supernatant was fractionated on a column (56 × 2.6 cm) of Sephadex G-50 (S) (Amersham Biosciences) in 0.05 M pyridinium acetate buffer pH 4.5 monitored using a differential refractometer (Knauer). High-molecular-mass polysaccharides were obtained in yields of 40% and 35%, respectively, of the LPS weight.

**Chemical analyses of the O antigen.** The polysaccharides were hydrolysed with 2 M CF<sub>3</sub>CO<sub>2</sub>H (120 °C, 2 h) and sugars were identified by GLC of the alditol acetates on a Hewlett Packard 5890 chromatograph equipped with an Ultra-2 column (Supelco) using a temperature gradient of 160 °C to 290 °C at 3 °C min<sup>-1</sup>. The absolute configurations of the monosaccharides were determined by GLC of the acetylated (-)-2-octyl glycosides as described by Leontein & Longgren (1993).

**NMR spectroscopy analyses of the O antigen.** Samples were deuterium-exchanged by freeze-drying twice from D<sub>2</sub>O and then examined as solutions in 99.96% D<sub>2</sub>O at 27 °C. NMR spectra were recorded on a Bruker DRX-500 spectrometer using internal acetone ( $\delta_{\text{H}}$  2.225,  $\delta_{\text{C}}$  31.45) as reference. Two-dimensional NMR spectra were obtained using standard Bruker software, and the Bruker XWINMR 2.6 program was used to acquire and process the NMR data. Mixing times of 200 and 100 ms were used in total correlation spectroscopy (TOCSY) and rotating Overhauser effect spectroscopy (ROESY) experiments, respectively.

**PCR specificity assay.** Chromosomal DNA was prepared from 186 *E. coli* strains, including *Shigella* strains of different O antigen serotypes. The quality of DNA was examined by PCR amplification of the *mdh* gene (encoding malate dehydrogenase and present as a housekeeping gene in *E. coli*) by using primers as described previously (Pupo *et al.*, 1997). A total of 14 pools of DNA were made, each containing DNA from 15–20 *E. coli* or *Shigella* O serotypes as described previously (Feng *et al.*, 2004). The pools were screened by

PCR by using primers based on *E. coli* O148 and *S. dysenteriae* type 1 specific genes *wzx* [wl-2412 (5'-TTCCCGTTTATAGTTC-3')/wl-2413 (5'-TTTGCCACAACATAGTCA-3'); wl-2414 (5'-CAACAA-CAATCGCTAAAC-3')/wl-2415 (5'-TGAAATAATGGCAGGTAG-3')] and *wzy* [wl-2416 (5'-GTCTTAATAGCCAATGCG-3')/wl-2417 (5'-GATAAAGCGAACAGTCCA-3'); wl-2418 (5'-GGAGTGGGAAGGG-TTGTTT-3')/wl-2419 (5'-GGGAGCGTGACATTCTTT-3')]. Each PCR was carried out in a 25  $\mu$ l (total volume) mixture, and 15  $\mu$ l of the mixture was loaded on an agarose gel to check for amplified DNA.

**Other methods.** Membrane preparation, SDS-PAGE, and silver staining for visualizing the LPS were carried out as described by Wang & Reeves (1994).

## RESULTS AND DISCUSSION

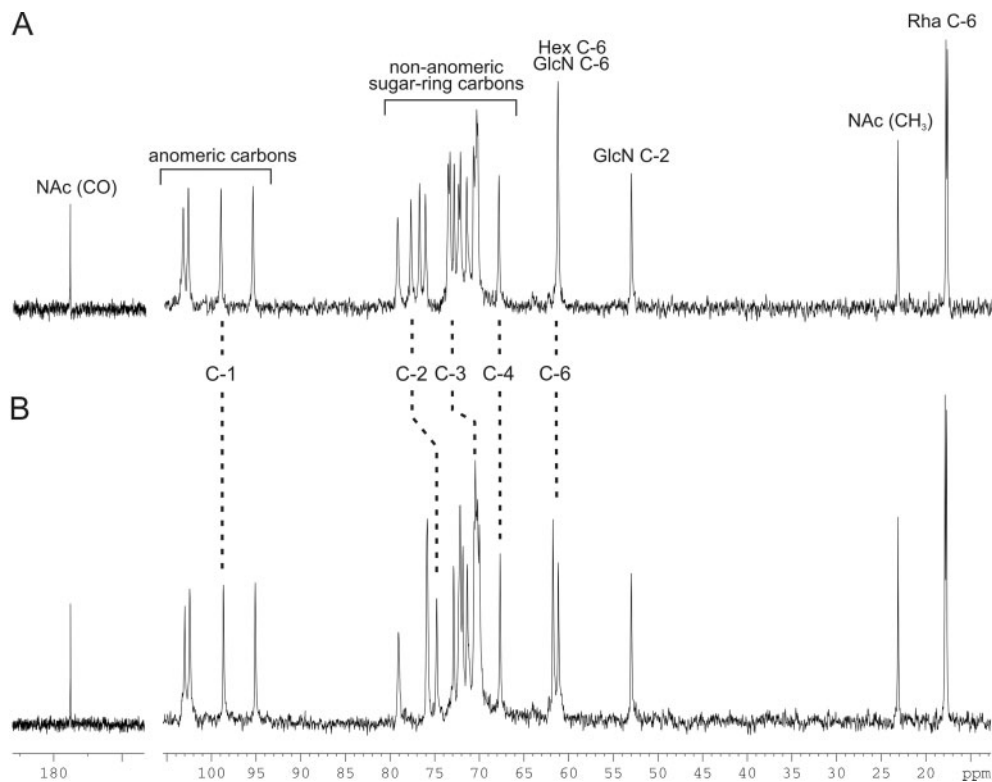
### Structure of the *E. coli* O148 O antigen

The O antigen was obtained by mild acid degradation of the corresponding LPS isolated from dried cells by the phenol/water procedure (Wang *et al.*, 2001a). Sugar analysis after full acid hydrolysis of the polysaccharide revealed rhamnose (Rha), glucose and glucosamine (GlcN) in the ratios 1.2:1:0.2. GLC of the acetylated glycosides with (S)-2-octanol showed that glucosamine and glucose have the D configuration while rhamnose has the L configuration.

The <sup>13</sup>C NMR spectrum of the polysaccharide (Fig. 1, top) contained signals for four anomeric carbons at  $\delta$  95.6–103.4, one nitrogen-bearing carbon (C-2 of GlcN) at  $\delta$  53.4, two non-substituted HOCH<sub>2</sub>-C groups at  $\delta$  61.7 (double signal), two CH<sub>3</sub>-C groups (C-6 of Rha) at  $\delta$  18.1 and 18.3, one *N*-acetyl group at  $\delta$  23.6 (CH<sub>3</sub>) and 175.5 (CO), and 15 other carbons in the region  $\delta$  68.0–80.0. Accordingly, the <sup>1</sup>H NMR spectrum of the polysaccharide contained signals for four anomeric protons at  $\delta$  5.02–5.53, one *N*-acetyl group at  $\delta$  2.07 and two methyl groups of rhamnose at  $\delta$  1.29 and 1.33, as well as signals of the other sugar protons. Therefore, the polysaccharide is regular and has a repeating unit containing one glucose, one glucosamine and two rhamnose residues.

The signals in the <sup>1</sup>H and <sup>13</sup>C NMR spectra of the polysaccharide were assigned using two-dimensional correlation spectroscopy, TOCSY, H-detected <sup>1</sup>H,<sup>13</sup>C heteronuclear single quantum correlation (HSQC) and heteronuclear multiple band correlation (HMBC) experiments (Tables 1 and 2). Based on the coupling constant values estimated from the two-dimensional spectra, the spin systems were assigned to two rhamnose residues (Rha<sup>I</sup> and Rha<sup>II</sup>) and one residue each of GlcNAc and Glc. The spin system for GlcNAc was recognized by a correlation between the nitrogen-bearing carbon (C-2) and H-2 at  $\delta$  53.4/4.12 in the <sup>1</sup>H,<sup>13</sup>C HSQC spectrum.

The *J*<sub>1,2</sub> coupling constant values of ~3 Hz indicated that Glc and GlcNA are  $\alpha$ -linked, and the position of the signals for C-5 at  $\delta$  70.8 indicated that Rha<sup>I</sup> and Rha<sup>II</sup> are  $\alpha$ -linked too (compare published data: Jansson *et al.*, 1989). The pyranose form of all monosaccharide residues was demonstrated by the absence from the <sup>13</sup>C NMR spectrum of any



**Fig. 1.** 125-MHz  $^{13}\text{C}$  NMR spectra of the O antigens of *E. coli* O148 (A) and construct strain H1230 (B). Hex stands for glucose in *E. coli* O148 or galactose in H1230. The most significant chemical shift differences between the spectra are observed for C-2 and C-3 of Hex.

signals for non-anomeric sugar carbons at a lower field than  $\delta$  80 (Bock & Pedersen, 1983).

Relatively low-field positions of the signals for C-3 of Rha<sup>I</sup> and Rha<sup>II</sup>, C-2 of Glc and C-3 of GlcNAc at  $\delta$  76.4, 79.5, 78.0 and 77.1, respectively, as compared with their positions in the corresponding non-substituted monosaccharides

(Lipkind *et al.*, 1988), demonstrated the modes of glycosylation of the monosaccharides in the repeating unit.

The sequence of the sugar residues was determined by a  $^1\text{H}$ ,  $^1\text{H}$  ROESY experiment, which revealed strong inter-residue Rha<sup>I</sup> H-1/Rha<sup>II</sup> H-3; Rha<sup>II</sup> H-1/Glc H-2; Glc H-1/GlcNAc H-3 and GlcNAc H-1/Rha<sup>I</sup> H-3 cross-peaks at  $\delta$

**Table 1.**  $^1\text{H}$  NMR data ( $\delta$ , p.p.m.) of the O antigens

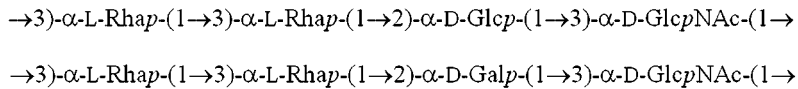
The chemical shift for NAc is  $\delta$  2.07 (Me).

Sugar residue	H-1	H-2	H-3	H-4	H-5	H-6
<b><i>E. coli</i> O148</b>						
$\rightarrow 3$ )- $\alpha$ -L-Rhap <sup>I</sup> -(1 $\rightarrow$ )	5.09	4.22	3.92	3.55	3.87	1.33
$\rightarrow 3$ )- $\alpha$ -L-Rhap <sup>II</sup> -(1 $\rightarrow$ )	5.08	4.15	3.86	3.55	3.87	1.29
$\rightarrow 2$ )- $\alpha$ -D-Glcp-(1 $\rightarrow$ )	5.53	3.64	3.75	3.48	3.63	3.81
$\rightarrow 3$ )- $\alpha$ -D-GlcpNAc-(1 $\rightarrow$ )	5.02	4.12	4.04	3.78	4.04	3.81
<b>Construct strain H1230</b>						
$\rightarrow 3$ )- $\alpha$ -L-Rhap <sup>I</sup> -(1 $\rightarrow$ )	5.10	4.22	3.93	3.55	3.90	1.33
$\rightarrow 3$ )- $\alpha$ -L-Rhap <sup>II</sup> -(1 $\rightarrow$ )	5.05	4.16	3.87	3.56	3.90	1.29
$\rightarrow 2$ )- $\alpha$ -D-Galp-(1 $\rightarrow$ )	5.59	3.93	3.88	4.01	3.92	3.77
$\rightarrow 3$ )- $\alpha$ -D-GlcpNAc-(1 $\rightarrow$ )	5.03	4.13	4.08	3.78	4.20	3.82

**Table 2.**  $^{13}\text{C}$  NMR data ( $\delta$ , p.p.m.) of the O antigens

The chemical shifts for NAc are  $\delta$  23.6 (Me) and 175.5 (CO).

Sugar residue	C-1	C-2	C-3	C-4	C-5	C-6
<b><i>E. coli</i> O148</b>						
$\rightarrow 3$ )- $\alpha$ -L-Rhap <sup>I</sup> -(1 $\rightarrow$ )	103.4	68.2	76.4	71.8	70.8	18.3
$\rightarrow 3$ )- $\alpha$ -L-Rhap <sup>II</sup> -(1 $\rightarrow$ )	102.8	71.1	79.5	72.7	70.8	18.1
$\rightarrow 2$ )- $\alpha$ -D-Glcp-(1 $\rightarrow$ )	99.2	78.0	73.8	70.6	73.7	61.7
$\rightarrow 3$ )- $\alpha$ -D-GlcpNAc-(1 $\rightarrow$ )	95.6	53.4	77.1	72.5	73.2	61.7
<b>Construct strain H1230</b>						
$\rightarrow 3$ )- $\alpha$ -L-Rhap <sup>I</sup> -(1 $\rightarrow$ )	103.4	68.1	76.4	71.8	70.8	18.3
$\rightarrow 3$ )- $\alpha$ -L-Rhap <sup>II</sup> -(1 $\rightarrow$ )	102.8	71.0	79.5	72.7	70.6	18.1
$\rightarrow 2$ )- $\alpha$ -D-Galp-(1 $\rightarrow$ )	99.1	75.2	70.5	70.9	72.3	62.2
$\rightarrow 3$ )- $\alpha$ -D-GlcpNAc-(1 $\rightarrow$ )	95.6	53.4	76.3	72.6	73.3	61.6



**Fig. 2.** Structures of the O antigens of *E. coli* O148 (top) and of *S. dysenteriae* type 1 and construct strain H1230 (bottom).

5.09/3.86; 5.08/3.64; 5.53/4.04 and 5.02/3.92, respectively. This pattern was confirmed by the  $^1\text{H}$ ,  $^{13}\text{C}$  HMBC experiment (data not shown), and it was concluded that the O antigen of *E. coli* O148 has the structure shown in Fig. 2.

The same experimental approaches were used to determine the O antigen structure of *E. coli* O148 derivative H1230 (Tables 1 and 2), which was shown to differ from the parent form only by replacement of  $\rightarrow 2)\text{-}\alpha\text{-D-Glcp-(1}\rightarrow$  by  $\rightarrow 2)\text{-}\alpha\text{-D-Galp-(1}\rightarrow$  (Fig. 2). Thus, *E. coli* O148 derivative H1230 has the same O antigen structure as *S. dysenteriae* type 1.

### The *E. coli* O148 and *S. dysenteriae* type 1 O antigen gene clusters are closely related

A sequence of 10241 bases, including *gnd* (positions 8902–10241), was obtained from *E. coli* O148 strain E 519-66. Nine ORFs excluding *gnd* were identified, all transcribed from JUMPstart to *gnd* (Fig. 3). Functions of each ORF in the *E. coli* O148 O antigen gene cluster were predicted on the basis of homology by searching available databases and are summarized in Table 3.

ORFs 1–4 were identified as *rmlB*, *rmlD*, *rmlA* and *rmlC*, respectively, based on their high level of identity (81–98%) to known *rml* genes from other *Shigella* and *E. coli* strains. The set of *rml* genes is responsible for the synthesis of dTDP-L-rhamnose in many O antigen gene clusters including *E. coli* O148. ORFs 1–4 were named accordingly.

ORFs 5–8 were identified as *wzx*, *wzy*, *wbbR* and *wbbQ* respectively, based on the high level of identity (97–99%) of the genes to those of *S. dysenteriae* type 1. In *S. dysenteriae* type 1, *wzx*, *wzy*, *wbbR* and *wbbQ* are proposed to encode O unit flippase, O antigen polymerase, rhamnosyltransferase II

and rhamnosyltransferase I, respectively (Klena & Schnaitman, 1993; Sturm *et al.*, 1986). They are presumed to have the same functions in *E. coli* O148.

ORF 9 was confirmed to be the glucosyltransferase gene (see below) and named as *wbbG*.

The O antigen gene clusters of *E. coli* O148 and *S. dysenteriae* type 1 have the same genes in the same order with DNA identity ranging from 89.8% to 99.5% (Fig. 3). However, *wbbG* in *S. dysenteriae* type 1 has a deletion of 22 bases flanked by 8 bp direct repeat sequences (Fig. 4), which generated a frame-shift and stop codon. ORF 9 was not assigned a function in *S. dysenteriae* type 1 (Klena & Schnaitman, 1993).

### The first sugar in the O antigen repeat unit of *E. coli* O148 and *S. dysenteriae* type 1

The *E. coli* O148 gene cluster does not include a gene for an initial sugar phosphate transferase. However, the structure includes GlcNAc, and we assume that, as shown for several such O antigens, GlcNAc is the initial sugar, with GlcNAc-*P* transferred to undecaprenol phosphate by *WecA*, encoded by the *wecA* gene of the enterobacterial common antigen gene cluster. We note that early workers on the genetics of the related *S. dysenteriae* type 1 suggested that the first sugar of the repeat unit was galactose (Sturm *et al.*, 1986), based essentially on the claim that the presence of the *wbbP* gene conferred addition of galactose to the *E. coli* LPS core. However, given the current understanding of the role of *wecA* in *E. coli* and *Shigella* O antigen synthesis, we consider that conclusion most improbable, and have shown GlcNAc as first sugar of the repeat unit in Fig. 2.

	D1									
	<i>rmlB</i>	<i>rmlD</i>	<i>rmlA</i>	<i>rmlC</i>	<i>wzx</i>	<i>wzy</i>	<i>orf7</i>	<i>orf8</i>	<i>orf9/orf10</i>	<i>gnd</i>
G+C (%)	43.7	47.1	43.2	35.5	32.2	31.0	30.0	30.3	31.3 32.1	51.1
DNA identity (%)	94.6	89.8	94.2	97.1	98.5	99.0	98.5	99.5	98.4 97.8	96.9
Protein identity (%)	97.0	95.7	98.3	98.4	97.7	98.4	98.6	99.0	97.8 96.9	99.9
Protein similarity (%)	98.1	97.7	98.6	100	98.5	98.7	98.9	99.7	98.5 98.8	100
G+C (%)	42.5	46.9	43.1	35.7	31.7	31.1	29.7	30.0	31.0	51.2
	O148									
	<i>rmlB</i>	<i>rmlD</i>	<i>rmlA</i>	<i>rmlC</i>	<i>wzx</i>	<i>wzy</i>	<i>wbbR</i>	<i>wbbQ</i>	<i>wbbG</i>	<i>gnd</i>

**Fig. 3.** Comparison of the O antigen gene clusters of *E. coli* O148 and *S. dysenteriae* type 1. All genes are transcribed in the direction from JUMPstart to *gnd*.

**Table 3.** Genes of the *E. coli* O148 O antigen gene cluster with selected properties of the genes and encoded proteins

Gene	G+C content (mol%)	No. of residues	Conserved domain(s)	Putative function	Related protein	
					Protein (accession no.)	%Identical aa/%Similar aa (no. of aa)
<i>rmlB</i>	42.5	361	NAD-dependent epimerase/dehydratase family PF01370 <i>E</i> value = $2.0 \times 10^{-212}$	dTDP-glucose 4,6-dehydratase	RmlB <i>Shigella boydii</i> (AAL27347)	97/98 (361)
<i>rmlD</i>	46.9	299	RmlD substrate-binding domain PF04321 <i>E</i> value = $8.3 \times 10^{-76}$	dTDP-6-deoxy-L-mannose dehydrogenase	RmlD <i>Shigella boydii</i> (AAL27348)	94/97 (299)
<i>rmlA</i>	43.1	292	Nucleotidyl transferase PF00483 <i>E</i> value = $3.6 \times 10^{-118}$	Glucose-1-phosphate thymidyltransferase	RmlA <i>Shigella boydii</i> (AAL27349)	97/98 (289)
<i>rmlC</i>	35.7	185	dTDP-4-dehydrorhamnose 3,5-epimerase PF00908 <i>E</i> value = $3.9 \times 10^{-117}$	dTDP-4-dehydrorhamnose 3,5-epimerase	RmlC <i>Shigella boydii</i> (AAL27325)	81/88 (179)
<i>wzx</i>	31.8	396	Polysaccharide biosynthesis protein PF01943 <i>E</i> value = $1.2 \times 10^{-83}$	O antigen transporter	RfbX <i>Shigella dysenteriae</i> type 1 (AAA16934)	97/98 (396)
<i>wzy</i>	31.1	380	None	O antigen polymerase	Rfc <i>Shigella dysenteriae</i> type 1 (AAA16935)	98/98 (380)
<i>wbbR</i>	30.0	282	Glycosyltransferase PF00535 <i>E</i> value = $2.3 \times 10e^{-2}$	Rhamnosyltransferase	RfbR <i>Shigella dysenteriae</i> type 1 (AAA16936)	98/98 (282)
<i>wbbQ</i>	30.0	303	Glycosyltransferase PF00535 <i>E</i> value = $3.4 \times 10^{-18}$	Rhamnosyltransferase	RfbQ <i>Shigella dysenteriae</i> type 1 (AAA16937)	99/99 (303)
<i>wbbG</i>	31.0	363	Glycosyltransferases group 1 PF00534 <i>E</i> value = $1.7 \times 10^{-31}$	Galactosyltransferase	WbsH <i>Escherichia coli</i> (AAO37690)	33/50 (380)

#### Construction of a mimic of the *S. dysenteriae* type 1 O antigen from *E. coli* O148 and identification of the glycosyltransferase gene

A *wbbG* knockout mutant of *E. coli* O148 was constructed by replacing it with a CAT gene. This mutant, named H1229, produced no O antigen, whereas the wild-type strain E 519-66 produced normal LPS (Fig. 5). The *wbbP* gene and its

own promoter were cloned from *S. dysenteriae* type 1 into pUC19, and the resultant plasmid pLW1059 was electroporated into H1229 to give strain H1230.

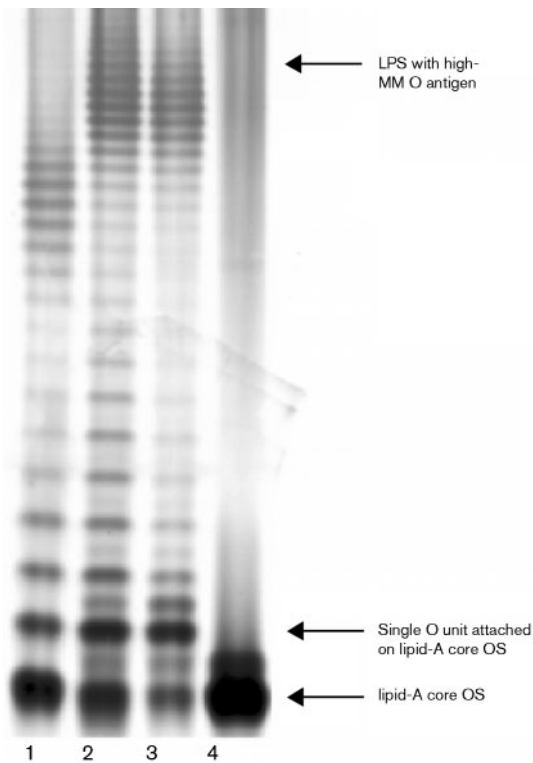
Strain H1230 produced normal LPS with polymeric O antigen (Fig. 5), showing that the plasmid pLW1059 could complement the mutation in strain H1229. Slide agglutination tests showed that strain H1230 reacted with *S.*

```

E. coli O148      TAAGGAAAATTTTITACTATTGCTCTGGA AAAATTTTTCGACTC
S. dysenteriae 1 TAAGGAAAATTTT-----GCGACTC

```

**Fig. 4.** Sequence comparison of part of the *E. coli* O148 *wbbG* gene (top) and the corresponding sequence in *S. dysenteriae* type 1 (bottom). The sequences in the frame are direct repeats in *E. coli* O148 thought to have been involved in generating the deletion in *S. dysenteriae* type 1.



**Fig. 5.** Deletion and complementation analysis of *E. coli* O148 *wbbG*. Membrane extracts were run on SDS-PAGE gels and silver stained. Lanes: 1, LSPQ3472 (*S. dysenteriae* type 1 type strain); 2, E 519-66 (*E. coli* O148 type strain); 3, H1230 (H1229 with plasmid pLW1059); 4, H1229 (*E. coli* O148 lacking the *wbbG* gene). MM, molecular mass.

*dysenteriae* type 1 specific antiserum while the *E. coli* O148 type strain E 519-66 did not. Using the same chemical and NMR spectroscopy (Fig. 1, bottom; Tables 1 and 2) approaches as described above, the O antigen structure of strain H1230 was shown to differ from the parent form only by replacement of the  $\rightarrow 2$ - $\alpha$ -D-Glcp-(1 $\rightarrow$ ) unit by a  $\rightarrow 2$ - $\alpha$ -D-Galp-(1 $\rightarrow$ ) unit and, hence, to be same as that of *S. dysenteriae* type 1 (Fig. 2). It can be seen in Fig. 5 that the replacement of glucose by galactose in *E. coli* O148 has not affected the spacing in the LPS 'ladder' of molecules with increasing number of repeat units. The high-molecular-mass LPS has a modal distribution similar to that of the parent *E. coli* O148 strain, with maximum staining at about 19 repeat units, whereas the *S. dysenteriae* type 1 strain has maximum staining at about 13 repeat units.

These findings confirmed that *wbbG* is the glucosyltransferase gene responsible for transferring UDP-glucose to GlcNAc via an  $\alpha$ -1,3-linkage in the O antigen of *E. coli* O148. The protein encoded by *wbbG* belongs to the glycosyltransferase group 1 (PF00534, *E* value =  $1.7 \times 10^{-31}$ ). It also has 33 and 31 % identity (50 and 52 % similarity) to WbsH of *E. coli* O128 and WbgM of *E. coli* O55, both of which are putative glycosyltransferases (Shao *et al.*, 2003; Wang *et al.*, 2002).

The fact that the *E. coli* O148 Rha<sup>II</sup>-(1 $\rightarrow$ 2)- $\alpha$ -D-Glcp link is replaced by a Rha<sup>II</sup>-(1 $\rightarrow$ 2)- $\alpha$ -D-Galp in derivative strain H1230 shows that the *E. coli* O148 Rha<sup>II</sup> transferase is relaxed for glucose or galactose as acceptor sugar, and thus no change in this transferase gene would have been required for the change in O antigen structure.

### ***S. dysenteriae* type 1 gained its O antigen genes from *E. coli* O148**

The above data suggest that the *S. dysenteriae* type 1 O antigen gene cluster is derived from the *E. coli* O148 gene cluster by loss of the proposed glucosyltransferase gene *wbbG* and gain of a plasmid-borne galactosyltransferase. The *S. dysenteriae* type 1 O antigen is one of those not otherwise known to occur in *E. coli* but is shown here to be derived from the *E. coli* O148 O antigen.

Although the exact origin of the galactosyltransferase gene is unknown, it may well have been transferred from another species in the human enteric tract on the plasmid, which presumably arose by transfer of a galactosyltransferase gene from a typical polysaccharide gene cluster to a plasmid. It is most probable that the plasmid was gained before the inactivation of the glucosyltransferase gene *wbbG*, as with neither *wbbG* or *wbbP*, the O unit could not be completed and the LPS would lack O antigen.

MLEE and *mdh* sequence data show that *E. coli* O148:H28 and *S. dysenteriae* type 1 are not very closely related (Pupo *et al.*, 1997). However, the *E. coli* O148 O antigen is found in association with several H antigens and further work would be needed to determine if *S. dysenteriae* type 1 was derived from a different *E. coli* O148 serotype.

### **Identification of *E. coli* O148 and *S. dysenteriae* type 1 specific genes**

PCR assays based on specific genes have been developed for many pathogenic *E. coli* strains such as *E. coli* O157 and O111 (Wang *et al.*, 1998; Wang & Reeves, 1998). Two pairs of primers designed for each of *wzx* and *wzy* from *E. coli* O148 were shown to be specific to *E. coli* O148 and *S. dysenteriae* type 1 by screening against 186 O serotypes of *E. coli* and *Shigella* strains by PCR. It would also be possible to distinguish the two closely related O antigens by use of primers for the plasmid-encoded *wbbP* gene found only in *S. dysenteriae* type 1.

### **Concluding remarks**

In summary, the *E. coli* O148 O antigen structure and gene cluster were elucidated. These data showed that *E. coli* O148 has a similar O antigen and gene cluster to *S. dysenteriae* type 1, and a mimic of *S. dysenteriae* type 1 O antigen was constructed from *E. coli* O148. Based on these results, it is concluded that *S. dysenteriae* type 1 gained its O antigen from *E. coli* O148. In addition, a glucosyltransferase gene of *E. coli* O148 was identified and two genes were confirmed to be serotype-specific to *E. coli* O148 and *S. dysenteriae* type 1.

## ACKNOWLEDGEMENTS

This work was supported by the NSFC (National Natural Science Foundation of China) Key Programs (30530010, 20536040), NSFC General Programs (30370023, 30370339), Tianjin Municipal Fund for Science and Technology Innovation (05FZZDSH00800), the Russian Foundation for Basic Research (05-04-48992) and the Council on Grants of the President of the Russian Federation (MK-1597.2005.3).

## REFERENCES

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3398–3402.
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. E. (1999). *Current Protocols in Molecular Biology*. New York: Wiley.
- Bastin, D. A. & Reeves, P. R. (1995). Sequence and analysis of the O antigen gene (*rfb*) cluster of *Escherichia coli* O111. *Gene* **164**, 17–23.
- Bateman, A., Birney, E., Cerruti, L., Durbin, R., Ewlinger, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M. & Sonnhammer, E. L. (2002). The Pfam protein families database. *Nucleic Acids Res* **30**, 276–280.
- Bennish, M. L., Harris, J. R., Wojtyniak, B. J. & Struelens, M. (1990). Death in shigellosis: incidence and risk factors in hospitalized patients. *J Infect Dis* **160**, 500–506.
- Black, R. E. (1986). The epidemiology of cholera and enterotoxigenic *Escherichia coli* diarrheal disease. In *Development of Vaccines and Drugs Against Diarrhea*, pp. 23–32. Edited by J. Holmgren, A. Lindberg & R. Mollby. Lund, Sweden: Studentlitteratur.
- Bock, K. & Pedersen, C. (1983). Carbon-13 nuclear magnetic resonance spectroscopy of monosaccharides. *Adv Carbohydr Chem Biochem* **41**, 27–66.
- Datsenko, K. A. & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640–6645.
- Dmitriev, B. A., Knirel, Y. A., Kochetkov, N. K. & Hofman, I. L. (1976). Somatic antigens of *Shigella*. Structural investigation on the O-specific polysaccharide chain of *Shigella dysenteriae* type 1 lipopolysaccharide. *Eur J Biochem* **66**, 559–566.
- Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. (1984). Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J Mol Biol* **179**, 125–142.
- Ewing, W. H. (1986). *Edwards and Ewing's Identification of the Enterobacteriaceae*, 4th edn. Amsterdam: Elsevier Science Publishers.
- Feng, L., Wang, W., Tao, J., Guo, H., Krause, G., Beutin, L. & Wang, L. (2004). Identification of *Escherichia coli* O114 O-antigen gene cluster and development of an O114 serogroup-specific PCR assay. *J Clin Microbiol* **42**, 3799–3804.
- Gohmann, S., Manning, P. A., Alpert, C. A., Walker, M. J. & Timmis, K. N. (1994). Lipopolysaccharide O-antigen biosynthesis in *Shigella dysenteriae* serotype 1: analysis of the plasmid-carried *rfp* determinant. *Microb Pathog* **16**, 53–64.
- Group, G. E. (1995). Public health impact of Rwandan refugee crisis: what happened in Goma, Zaire, in July 1994? *Lancet* **345**, 339–344.
- Guerin, P. J., Brasher, C., Baron, E., Mic, D., Grimont, F., Rayan, M., Aavitsland, P. & Legros, D. (2003). *Shigella dysenteriae* serotype 1 in West Africa: intervention strategy for an outbreak in Sierra Leone. *Lancet* **362**, 705–706.
- Henikoff, S., Henikoff, J. G., Alford, W. J. & Pietrokovski, S. (1995). Automated construction and graphical presentation of protein blocks from unaligned sequences. *Gene* **163**, GC17–26.
- Iijima, Y., Oundo, J. O., Taga, K., Saidi, S. M. & Honda, T. (1995). Simultaneous outbreak due to *Vibrio cholerae* and *Shigella dysenteriae* in Kenya. *Lancet* **345**, 69–70.
- Jansson, P. E., Kenne, L. & Widmalm, G. (1989). Computer-assisted structural analysis of polysaccharides with an extended version of CASPER using <sup>1</sup>H- and <sup>13</sup>C-n.m.r. data. *Carbohydr Res* **188**, 169–191.
- Klena, J. D. & Schnaitman, C. A. (1993). Function of the *rfb* gene cluster and the *rfe* gene in the synthesis of O antigen by *Shigella dysenteriae* 1. *Mol Microbiol* **9**, 393–402.
- Leontein, K. & Lonngren, J. (1993). Determination of the absolute configuration of sugars by gas-liquid chromatography of their acetylated 2-octyl glycosides. *Methods Carbohydr Chem* **9**, 87–89.
- Lipkind, G. M., Shashkov, A. S., Knirel, Y. A., Vinogradov, E. V. & Kochetkov, N. K. (1988). A computer-assisted structural analysis of regular polysaccharides on the basis of <sup>13</sup>C-n.m.r. data. *Carbohydr Res* **175**, 59–75.
- Mendizabal-Morris, C. A., Mata, L. J., Gangarosa, E. J. & Guzman, G. (1971). Epidemic Shiga-bacillus dysentery in Central America. Derivation of the epidemic and its progression in Guatemala, 1968–69. *Am J Trop Med Hyg* **20**, 927–933.
- Ochman, H., Whittam, T. S., Caugant, D. A. & Selander, R. K. (1983). Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*. *J Gen Microbiol* **129**, 2715–2726.
- Parolis, H., Parolis, L. A. & Olivieri, G. (1997). Structural studies on the *Shigella*-like *Escherichia coli* O121 O-specific polysaccharide. *Carbohydr Res* **303**, 319–325.
- Passwell, J. H., Harlev, E., Ashkenazi, S., Chu, C., Miron, D., Ramon, R., Farzan, N., Shiloach, J., Bryla, D. A. & other authors (2001). Safety and immunogenicity of improved *Shigella* O-specific polysaccharide-protein conjugate vaccines in adults in Israel. *Infect Immun* **69**, 1351–1357.
- Pazhani, G. P., Sarkar, B., Ramamurthy, T., Bhattacharya, S. K., Takeda, Y. & Niyogi, S. K. (2004). Clonal multidrug-resistant *Shigella dysenteriae* type 1 strains associated with epidemic and sporadic dysenteries in Eastern India. *Antimicrob Agents Chemother* **48**, 681–684.
- Pupo, G. M., Karaolis, D. K. R., Lan, R. & Reeves, P. R. (1997). Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and *mdh* sequence studies. *Infect Immun* **65**, 2685–2692.
- Pupo, G. M., Lan, R. & Reeves, P. R. (2000). Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. *Proc Natl Acad Sci U S A* **97**, 10567–10572.
- Reeves, P. R. & Wang, L. (2002). Genomic organization of LPS-specific loci. *Curr Top Microbiol Immunol* **264**, 109–135.
- Robbins, J. B., Chu, C. & Schneerson, R. (1992). Hypothesis for vaccine development: protective immunity to enteric diseases caused by nontyphoidal salmonellae and shigellae may be conferred by serum IgG antibodies to the O-specific polysaccharide of their lipopolysaccharides. *Clin Infect Dis* **15**, 346–361.
- Ronsmans, C., Bennish, M. L. & Wierzba, T. (1988). Diagnosis and management of dysentery by community health workers. *Lancet* **2**, 552–555.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. (2000). Artemis: sequence visualisation and annotation. *Bioinformatics* **16**, 944–945.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.



- Shao, J., Li, M., Jia, Q., Lu, Y. & Wang, P. G. (2003). Sequence of *Escherichia coli* O128 antigen biosynthesis cluster and functional identification of an alpha-1,2-fucosyltransferase. *FEBS Lett* **553**, 99–103.
- Shepherd, J. G., Wang, L. & Reeves, P. R. (2000). Comparison of O-antigen gene clusters of *Escherichia coli* (*Shigella*) sonnei and *Plesiomonas shigelloides* O17: sonnei gained its current plasmid-borne O-antigen genes from *P. shigelloides* in a recent event. *Infect Immun* **68**, 6056–6061.
- Staden, R. (1996). The Staden sequence analysis package. *Mol Biotechnol* **5**, 233–241.
- Sturm, S., Jann, B., Jann, K., Fortnagel, P. & Timmis, K. N. (1986). Genetic and biochemical analysis of *Shigella dysenteriae* 1 O antigen polysaccharide biosynthesis in *Escherichia coli* K-12: structure and function of the *rfb* gene cluster. *Microb Pathog* **1**, 307–324.
- Taylor, D. N., Trofa, A. C., Sadoff, J., Chu, C., Bryla, D., Shiloach, J., Cohen, D., Ashkenazi, S., Lerman, Y. & other authors (1993). Synthesis, characterization, and clinical evaluation of conjugate vaccines composed of the O-specific polysaccharides of *Shigella dysenteriae* type 1, *Shigella flexneri* type 2a, and *Shigella sonnei* (*Plesiomonas shigelloides*) bound to bacterial toxoids. *Infect Immun* **61**, 3678–3687.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.
- Wang, L. & Reeves, P. R. (1994). Involvement of the galactosyl-1-phosphate transferase encoded by the *Salmonella enterica rfbP* gene in O antigen subunit processing. *J Bacteriol* **176**, 4348–4356.
- Wang, L. & Reeves, P. R. (1998). Organization of *Escherichia coli* O157 O antigen gene cluster and identification of its specific genes. *Infect Immun* **66**, 3545–3551.
- Wang, L., Curd, H., Qu, W. & Reeves, P. R. (1998). Sequencing of *Escherichia coli* O111 O antigen gene cluster and identification of O111 specific genes. *J Clin Microbiol* **36**, 3182–3187.
- Wang, L., Briggs, C. E., Rothmund, D., Fratamico, P., Luchansky, J. B. & Reeves, P. R. (2001a). Sequence of the *E. coli* O104 antigen gene cluster and identification of O104 specific genes. *Gene* **270**, 231–236.
- Wang, L., Qu, W. & Reeves, P. R. (2001b). Sequence analysis of four *Shigella boydii* O-antigen loci: implication for *Escherichia coli* and *Shigella* relationships. *Infect Immun* **69**, 6923–6930.
- Wang, L., Huskic, S., Cisterne, A., Rothmund, D. & Reeves, P. R. (2002). The O-antigen gene cluster of *Escherichia coli* O55:H7 and identification of a new UDP-GlcNAc C4 epimerase gene. *J Bacteriol* **184**, 2620–2625.
- Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., Copeland, N. G. & Court, D. L. (2000). An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci U S A* **91**, 5978–5983.

---

Edited by: N. J. High