Microbiology (2010), **156**, 1642–1649 **DOI** 10.1099/mic.0.037325-0

Structure of the O-antigen of *Salmonella* O66 and the genetic basis for similarity and differences between the closely related O-antigens of *Escherichia coli* O166 and *Salmonella* O66

Bin Liu,[1,2]† Andrei V. Perepelov,[3]† Dan Li,[1,2]† Sof'ya N. Senchenkova,[3] Yanfang Han,[1,2] Alexander S. Shashkov,[3] Lu Feng,[1,2,4] Yuriy A. Knirel[3] and Lei Wang[1,2,4,5]

Correspondence
Lei Wang
wanglei@nankai.edu.cn

[1]TEDA School of Biological Sciences and Biotechnology, Nankai University, 23 Hongda Street, TEDA, Tianjin 300457, PR China

[2]Tianjin Key Laboratory of Microbial Functional Genomics, Tianjin 300457, PR China

[3]N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 119991 Moscow, Russian Federation

[4]Tianjin Research Center for Functional Genomics and Biochip, Tianjin 300457, PR China

[5]The Key Laboratory of Molecular Microbiology and Technology, Ministry of Education, PR China

O-antigen is a component of the outer membrane of Gram-negative bacteria and is one of the most variable cell surface constituents, leading to major antigenic variability. The O-antigen forms the basis for bacterial serotyping. In this study, the O-antigen structure of *Salmonella* O66 was established, which differs from the known O-antigen structure of *Escherichia coli* O166 only in one linkage (most likely the linkage between the O-units) and *O*-acetylation. The O-antigen gene clusters of *Salmonella* O66 and *E. coli* O166 were found to have similar organizations, the only exception being that in *Salmonella* O66, the *wzy* gene is replaced by a non-coding region. The function of the *wzy* gene in *E. coli* O166 was confirmed by the construction and analysis of deletion and *trans*-complementation mutants. It is proposed that a functional *wzy* gene located outside the O-antigen gene cluster is involved in *Salmonella* O66 O-antigen biosynthesis, as has been reported previously in *Salmonella* serogroups A, B and D1. The sequence identity for the corresponding genes between the O-antigen gene clusters of *Salmonella* O66 and *E. coli* O166 ranges from 64 to 70 %, indicating that they may originate from a common ancestor. It is likely that after the species divergence, *Salmonella* O66 got its specific O-antigen form by inactivation of the *wzy* gene located in the O-antigen gene cluster and acquisition of two new genes (a *wzy* gene and a prophage gene for *O*-acetyl modification) both residing outside the O-antigen gene cluster.

Received 17 December 2009
Revised 19 February 2010
Accepted 23 February 2010

INTRODUCTION

Lipopolysaccharide (LPS), a key component of the outer membrane of Gram-negative bacteria, consists of three distinct regions: lipid A, an oligosaccharide core and an O-specific polysaccharide (O-antigen). The O-antigen comprises a number of oligosaccharide repeats (O-units), which usually contain between two and eight sugar residues. The O-antigen exhibits extensive variation in the types of sugar present, their arrangement within the O-unit and the linkages within and between O-units, making LPS one of the most variable cell constituents. The O-antigen appears to be a major target of the immune system and bacteriophages. The variability of the O-antigen provides the major basis for serotyping schemes for many Gram-negative bacteria, such as *Escherichia coli* and *Salmonella*.

O-antigen diversity is thought to be important in allowing the various clones to present variations in surface structures that may offer a selective advantage in their specific niche (Reeves, 1992). O-antigen is also essential for the full function of bacteria and is related to bacterial

†These authors contributed equally to this work.

Abbreviations: GalNAc, *N*-acetylgalactosamine; GlcNAc, *N*-acetyl-D-glucosamine; LPS, lipopolysaccharide; SR-type, semi-rough type.

The GenBank/EMBL/DDBJ accession number for the sequences of the *E. coli* O166 and *Salmonella* O66 O-antigen gene clusters are GU299794 and GU299799, respectively.

virulence. It has been shown that the virulence of *Shigella flexneri* is reduced if the O-antigen is changed (Gemski *et al.*, 1972).

The majority of O-antigen biosynthetic genes reside within a gene cluster that maps between the *galF* and *gnd* genes in *E. coli*, *Salmonella* and *Shigella* genomes. Genetic variation in the O-antigen gene cluster contributes to a diverse array of O-antigen forms. O-antigen biosynthetic genes fall into three main classes: (i) nucleotide sugar precursor synthesis genes, (ii) sugar transfer genes associated with O-unit formation and (iii) O-unit processing genes associated with the conversion of the O-unit to O-antigen.

The clonal species *E. coli* includes both commensal and pathogenic strains, which are usually identified by a combination of their O- and H- (and sometimes K-) antigens. To date, 174 O-serogroups have been described for *E. coli* (Liu *et al.*, 2008). *Salmonella* is recognized as a major pathogen of both animals and humans, and in many countries, it is the leading cause of outbreaks of food-borne infections. Forty-six O-serogroups are described for *Salmonella* in the Kauffmann–White serotyping scheme, in which members of each serogroup have the same basic O-antigen structure and each serogroup is divided into serotypes based on H-antigens.

Though *E. coli* and *Salmonella* are known to be evolutionarily closely related, previous analysis indicates that they have only a few O-antigen structures in common (Samuel *et al.*, 2004). Combined structural and genetic studies will improve our understanding of the relationship between the O-antigens of *E. coli* and *Salmonella*. In this study, the O-antigen structure of *Salmonella* O66 was determined and compared with the related structure of *E. coli* O166 reported previously (Ali *et al.*, 2007). The O-antigen gene clusters of *Salmonella* O66 and *E. coli* O166 were sequenced and analysed. A *wzy* mutant of *E. coli* O166 was constructed to characterize the gene function. The evolutionary relationship between the O-antigens of *E. coli* O166 and *Salmonella* O66 is discussed.

METHODS

Cultivation of bacteria and isolation of LPS. *Salmonella* O66, strain G1601, was grown to late-exponential phase in 8 l LB using a 10 l fermenter (Biostat C-10, B. Braun Biotech International) under constant aeration at 37 °C and pH 7.0, and bacterial cells were washed and dried as described by Robbins & Uchida (1962). Dried cells (5 g) were extracted with a phenol/water mixture as described by Westphal & Jann (1965). After dialysis of combined phenol and water layers, contaminations were precipitated by adding 50 % aqueous trichloroacetic acid at 4 °C; the supernatant was dialysed against distilled water and lyophilized to give LPS (0.5 g).

Preparation and *O*-deacetylation of the O-polysaccharide. Delipidation of LPS (100 mg) was performed with aqueous 2 % acetic acid (6 ml) at 100 °C for 2 h. The precipitate was removed by centrifugation (13 000 ***g***, 20 min) and the supernatant was fractionated on a Sephadex G-50 Superfine column (56 × 2.6 cm; Amersham Biosciences) in 0.05 M pyridinium acetate buffer pH 5.5 with monitoring, using a differential refractometer (Knauer). The O-polysaccharide was obtained in a yield ~30 % of the LPS mass.

The O-polysaccharide (30 mg) was treated with aqueous 12.5 % ammonia at 37 °C for 16 h. Ammonia was removed with a stream of air and the remaining solution was desalted on a TSK HW-40 (S) column (90 × 2.5 cm; Merck) in water and freeze-dried to give an *O*-deacetylated polysaccharide (25 mg).

Chemical analyses. The O-polysaccharide was hydrolysed with 2 M trifluoroacetic acid (120 °C, 2 h) and the monosaccharides were identified as the alditol acetates by GLC on a Hewlett Packard 5890 chromatograph equipped with an Ultra-2 column (Agilent) using a temperature gradient of 160–290 °C at 3 °C min^{-1}. The absolute configuration of the monosaccharides was determined by GLC of the acetylated (*S*)-2-octyl glycosides as described by Leontein & Lönngren (1993).

NMR spectroscopy. Samples were deuterium-exchanged by freeze-drying twice from 99.9 % D_2O and then examined as solutions in 99.95 % D_2O at 27 °C. NMR spectra were recorded on an Avance 600 spectrometer (Bruker) using internal sodium 3-(trimethylsilyl)propionate-2,2,3,3-d_4 (δ_H 0.00) and acetone (δ_C 31.45) as references. Two-dimensional NMR spectra were obtained using standard Bruker software, and the Bruker TOPSPIN program was used to acquire and process the NMR data. Mixing times of 200 and 100 ms were used in TOCSY and ROESY experiments, respectively.

Construction of a DNase I shotgun bank, DNA sequencing and data analysis. Chromosomal DNA from *E. coli* and *Salmonella* was prepared as described previously (Bastin & Reeves, 1995). Primers WL_1098 (5′-ATTGGTAGCTGTAAGCCAAGGGCGGTAGCGT-3′) and WL_2211 (5′-CACTGCCATACCGACGACGCCGATCTGTTGC-TTGG-3′) (Wang & Reeves, 1998), based on sequences from the JUMPStart site and *gnd* genes, respectively, were used to amplify *E. coli* O166 and *Salmonella* O66 O-antigen gene clusters using the expand long template PCR system (Roche). The PCR cycle used was as follows: denaturation at 94 °C for 10 s, annealing at 60 °C for 30 s and extension at 68 °C for 15 min. The PCR products were digested with DNase I and the resulting DNA fragments were cloned into pGEM-T Easy (Promega) to produce a bank using the method described previously (Wang & Reeves, 1998). DNA sequencing was carried out by the Tianjin Biochip Corporation, using an ABI 3730 automated DNA sequencer. Sequence data were assembled and analysed as described previously (Liu *et al.*, 2006).

Deletion and complementation of the *wzy* gene. The *wzy* gene from *E. coli* O166 was replaced by a chloramphenicol acetyltransferase (*cat*) gene using the RED recombination system of phage lambda (Datsenko & Wanner, 2000; Yu *et al.*, 2000). The *cat* gene was amplified by PCR from plasmid pKK232-8 (Pharmacia) using primers that bind to the 5′ and 3′ ends of the gene, and each primer carried 40 bp of *E. coli* O166 DNA flanking the *wzy* gene (upstream primer WL-30957, 5′-TGCAAATAAACATATACTTTAAATATTAT-AAGAGATTAGTCAGGAGCTAAGGAAGCTAA-3′; downstream primer WL-30958, 5′-GTCTCCGACGAAAAAACAAATATTTTTCAC-AATAACCCTCAAAAAAATTACGCCCCGC-3′). The PCR product was transformed into *E. coli* O166 type strain G1216 carrying plasmid pKD20, and chloramphenicol-resistant transformants were selected after induction of the RED genes according to the protocol described by Datsenko & Wanner (2000). To complement the *wzy*-deficient mutants, the *wzy* gene of *E. coli* O166 was amplified using primers WL-30961 (upstream primer, 5′-GGAATTCGATGTACCATATAGC-AATAGCGT-3′) and WL-30962 (downstream primer, 5′-CGGGA-TCCTTAAATACTAAAAAATGTTTTATAC-3′), and the resulting PCR products were cloned into pUC18 to create plasmid pLW1539. Membrane preparation, SDS-PAGE and silver staining were carried

out for visualization of the LPS, as described previously (Wang & Reeves, 1994).

RESULTS AND DISCUSSION

Structure elucidation of the O-polysaccharide of *Salmonella* O66

LPS was obtained from dried bacterial cells of *Salmonella* O66 by using the Westphal procedure (Westphal & Jann, 1965) and degraded with mild acid to yield an O-polysaccharide isolated by gel-permeation chromatography on Sephadex G-50. Sugar analysis by GLC of the alditol acetates derived after full acid hydrolysis of the polysaccharide revealed glucose and galactose in the ratio 1 : 2.2, as well as *N*-acetylgalactosamine (GalNAc). GLC of the acetylated glycosides with (*S*)-2-octanol showed that all monosaccharides have the D configuration.

The ^{13}C NMR spectrum of the O-polysaccharide (Fig. 1a) contained signals of different intensities, most likely owing to nonstoichiometric *O*-acetylation, as there was a signal for CH_3 of an *O*-acetyl group at δ 21.9. The ^{1}H-NMR spectrum of the O-polysaccharide showed signals for one *O*-acetyl group at δ 2.13 and two *N*-acetyl groups at δ 2.04 and 2.06. The NMR spectra of the *O*-deacetylated polysaccharides were typical of a regular polymer. The ^{13}C NMR spectrum (Fig. 1b) showed signals for five anomeric carbons at δ 95.1–106.5, two nitrogen-bearing carbons (C-2 of two GalNAc residues) at δ 50.1 and 52.5, five CH_2O groups (C-6 of monosaccharides) at 61.2–62.7 (4 carbons) and 66.5 (shown by attached-proton test), other sugar carbons in the region δ 65.0–79.9 and two *N*-acetyl groups at δ 23.6, 23.9 (both CH_3), 176.0 and 176.1 (both CO). The ^{1}H NMR spectrum of the *O*-deacetylated polysaccharide contained, among other things, signals for five anomeric protons at δ 4.45–5.16 and two *N*-acetyl groups at δ 2.03 and 2.05.

The ^{1}H and ^{13}C NMR spectra of the *O*-deacetylated polysaccharide were assigned (Table 1) using 2D correlation spectroscopy, including ^{1}H,^{1}H COSY, TOCSY and ^{1}H,^{13}C HSQC experiments, and five sugar spin systems were revealed. Based on coupling constant values, the spin systems were assigned to two residues each of galactose (A and B) and GalNAc (C and D) and one residue of glucose (E). $J_{1,2}$ coupling constant values of ~3 Hz indicated that units A, B and C are α-linked, whereas significantly larger $J_{1,2}$ values of 7–8 Hz showed that units D and E are β-linked.

The signals for C-2 of unit A, C-6 of unit B, C-3 and C-4 of unit C and C-3 of unit D were shifted significantly downfield to δ 79.9, 66.5, 77.7, 76.9 and 76.3, respectively, compared with their positions in the corresponding non-substituted monosaccharides (Lipkind *et al.*, 1988). These displacements were due to glycosylation and thus defined the glycosylation pattern in the O-unit. The sequence of the sugar residues was determined by a 2D ROESY experiment, which showed inter-residue cross-peaks between the following anomeric protons and protons at the linkage carbons: A H-1/B H-6a,6b; B H-1/C H-4; C H-1/D H-3; D H-1/A H-2; and E H-1/C H-3. The sequence

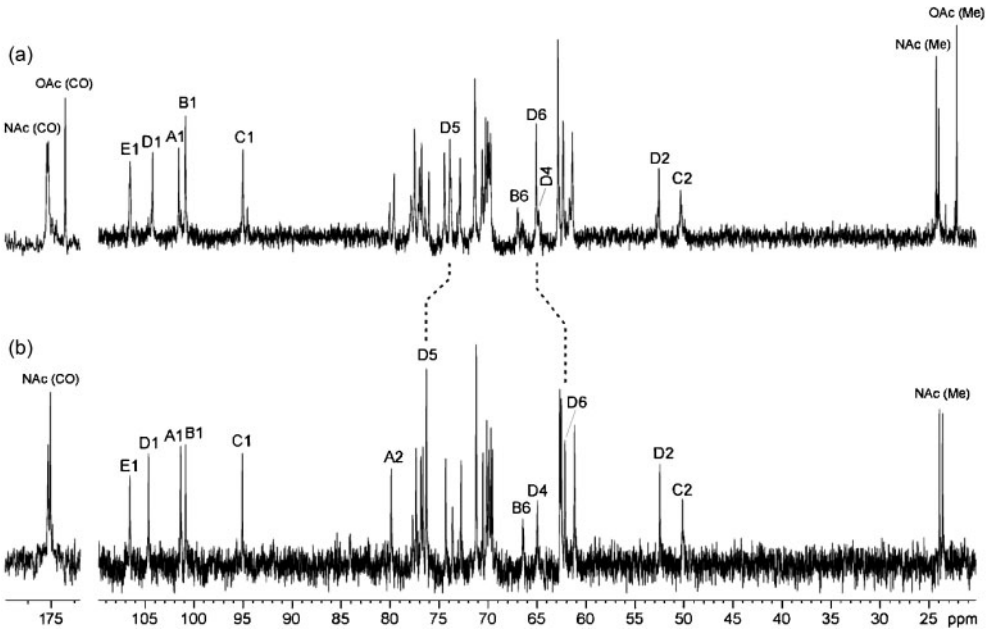

Fig. 1. ^{13}C NMR spectra (125 MHz) of the O-polysaccharide (a) and *O*-deacetylated polysaccharide (b) from *Salmonella* O66. Numbers refer to the carbons in sugar residues denoted by letters, as shown in Table 1.

Table 1. ^{1}H and ^{13}C NMR chemical shifts of the polysaccharides from *Salmonella* O66 (δ, ppm)

Sugar residue	Nucleus	Polysaccharide					
		1	2	3	4	5	6 (6a, 6b)
O*-deacetylated polysaccharide							
→2)-α-D-Gal*p*-(1→	^{1}H	5.16	3.84	3.92	3.96	3.94	3.73; 3.73
A	^{13}C	101.4	79.9	69.6	71.2	72.8	62.7
→6)-α-D-Gal*p*-(1→	^{1}H	5.00	3.83	4.06	4.28	4.58	3.75; 3.75
B	^{13}C	100.9	70.1	70.5	69.7	69.9	66.5
→3,4)-α-D-Gal*p*NAc-(1→	^{1}H	5.10	4.49	3.88	4.27	3.91	3.85; 3.85
C	^{13}C	95.1	50.1	77.7	76.9	73.6	61.2
→3)-β-D-Gal*p*NAc-(1→	^{1}H	4.66	4.09	3.81	4.13	3.65	3.74; 3.88
D	^{13}C	104.7	52.5	76.3	65.0	76.3	62.1
β-D-Glc*p*-(1→	^{1}H	4.45	3.30	3.51	3.42	3.42	3.76; 3.91
E	^{13}C	106.5	74.3	76.7	71.2	77.3	62.5
O-polysaccharide†							
→2)-α-D-Gal*p*-(1→	^{1}H	5.11	3.84	3.93	3.97	3.96	3.75; 3.75
A	^{13}C	101.7	79.6	69.7	71.3	72.8	62.8
→6)-α-D-Gal*p*-(1→	^{1}H	5.02	3.85	4.07	4.26	4.58	3.71; 3.79
B	^{13}C	101.0	70.2	70.6	69.9	70.0	66.9
→3,4)-α-D-Gal*p*NAc-(1→	^{1}H	5.11	4.51	3.89	4.29	3.92	3.86; 3.86
C	^{13}C	95.1	50.2	77.9	77.0	73.7	61.3
→3)-β-D-Gal*p*NAc6Ac-(1→	^{1}H	4.69	4.11	3.83	4.17	3.87	4.29; 4.40
D	^{13}C	104.4	52.4	76.0	64.7	73.9	65.0
β-D-Glc*p*-(1→	^{1}H	4.45	3.29	3.51	3.42	3.42	3.75; 3.89
E	^{13}C	106.7	74.4	76.8	71.3	77.5	62.3

*Signals for *N*-acetyl groups are at δ_H 2.03 and 2.05; δ_C 23.6 and 23.9 (Me); 176.0 and 176.1 (CO).
†Signals for an *O*-acetyl group are at δ_H 2.13, δ_C 21.9 (Me) and 175.5 (CO); for *N*-acetyl groups at δ_H 2.04 and 2.06, δ_C 23.7 and 24.0 (Me), 176.2 and 176.3 (CO).

was confirmed independently by a $^{1}H,^{13}C$ HMBC experiment, which showed correlations between anomeric protons and linkage carbons and vice versa (data not shown).

The position of the *O*-acetyl group in the O-polysaccharide was determined by a $^{1}H,^{13}C$ HSQC experiment, which showed a significant downfield displacement of ~80 % of the D H-6a,6b/C-6 cross-peaks from δ 3.74, 3.88/62.1 in the *O*-deacetylated polysaccharide to δ 4.29, 4.40/65.0 in the initial O-polysaccharide. This shift was evidently due to a deshielding effect of the *O*-acetyl group (α-effect of *O*-acetylation) and indicated *O*-acetylation of ~80 % unit D at position 6. This conclusion was confirmed by a downfield shift by 2.9 p.p.m. and an upfield shift by 2.4 p.p.m. of the signals for C-6 and C-5 of unit D (Fig. 1) caused by α- and β-effects of *O*-acetylation, respectively (Jansson *et al.*, 1987).

Therefore, the O-polysaccharide of *Salmonella* O66 has structure 1, as shown in Fig. 2. A closely related O-polysaccharide of *E. coli* O166, whose structure (2, Fig. 2) has been elucidated previously (Ali *et al.*, 2007), differs only in one linkage (at position 2 of Gal A in *Salmonella* O66 vs position 3 in *E. coli* O166), which is, most likely, the linkage between the O-units, and by the lack of *O*-acetylation.

Characterization of the O-antigen gene cluster of *E. coli* O166 and confirmation of the function of the *wzy* gene

A sequence of 10 140 bases was obtained between the JUMPStart site and the *gnd* gene from *E. coli* O166 type strain G1216. Seven ORFs were identified, all of which have the same transcriptional direction from the JUMPStart site to the *gnd* gene. All of the genes were assigned functions based on their similarities to genes from available databases (Table 2).

Genes for the synthesis of the nucleotide precursors of common sugars, including Glc and Gal, are located outside the O-antigen gene cluster (Samuel & Reeves, 2003). Therefore, only the biosynthetic gene for GalNAc was expected to be located in the O-antigen gene cluster of *E. coli* O166. Orf7 shares 60 % identity with Gne (a UDP-*N*-acetyl-D-glucosamine-4-eipmerase) of *E. coli* O107, which converts UDP-*N*-acetyl-D-glucosamine (-GlcNAc) to UDP-GalNAc (Wang *et al.*, 2009). Therefore, *orf7* was identified as a UDP-GlcNAc-4-epimerase gene and designated *gne* accordingly.

Both Wzx and Wzy are highly hydrophobic membrane proteins. Orf1 and Orf3 are the only two proteins with predicted transmembrane segments. Orf1 was predicted to

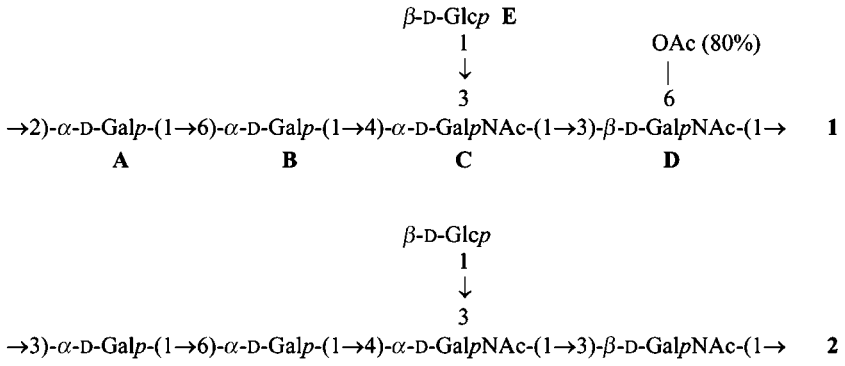

Fig. 2. Structure of the O-polysaccharides of *Salmonella* O66 (structure 1) and *E. coli* O166 (structure 2).

have 11 well-proportioned transmembrane segments, and shares 47 % similarity with the putative O-antigen flippase of *E. coli* strain S88. Orf3 was found to have 10 predicted transmembrane segments with a large periplasmic loop of 46 aa, a typical topological characteristic of Wzy proteins (Daniels *et al.*, 1998). Orf3 shares 48 % similarity with the Wzy protein of *S. boydii* type 11. Therefore, *orf1* and *orf3* were proposed to be the O-antigen flippase gene (*wzx*) and O-antigen polymerase gene (*wzy*), respectively, and were named accordingly.

WecA-catalysed transfer of GlcNAc-1-phosphate or GalNAc-1-phosphate to an undecaprenol phosphate carrier triggers O-unit synthesis in *E. coli*, and the *wecA* gene is located outside the O-antigen gene cluster (Alexander & Valvano, 1994). Therefore, four additional glycosyltransferase genes were expected in the O-antigen gene cluster of *E. coli* O166. Orf2, Orf4, Orf5 and Orf6 share 68, 61, 66 and 62 % similarity to WbcL, WbcM, WbcN and WbcQ, respectively, and all four proteins are glycosyltransferases involved in the synthesis of the LPS outer core of *Yersinia enterocolitica* serotype O : 3 (Skurnik *et al.*, 1999). Given these observations, *orf2*, *orf4*, *orf5* and *orf6* were proposed to be glycosyltransferase genes, and were named *weiA*, *weiB*, *weiC* and *weiD*, respectively.

To confirm the function of *wzy* in *E. coli* O166, gene deletion followed by *trans*-complementation was carried out. Deletion of the *wzy* gene in *E. coli* resulted in a semi-rough (SR-type) LPS phenotype with only one O-unit attached to the core-lipid A moiety (Daniels *et al.*, 1998). The mutant strain H1985 was constructed from the *E. coli* O166 type strain by replacing the *wzy* gene with the *cat* gene. A comparison of the wild-type and mutant LPS by SDS-PAGE (Fig. 3) revealed that the mutant strain produced no S-type LPS, but instead produced SR-type LPS with only one O-unit attached to the core. This SR-LPS phenotype was *trans*-complemented by introduction

Table 2. Characteristics of the ORFs in the *E. coli* O166 O-antigen gene cluster

Orf no.	Gene name	Position of gene	G+C content (%)	Conserved domain(s)	Similar protein(s) or strain(s) (GenBank accession no.)	Identity/similarity (%) (no. of aa overlap)	Putative function of protein
1	*wzx*	1175...2428	27.9	None	Putative O-antigen flippase, *E. coli* strain S88 (CAN87668)	23/47 (382)	O-antigen flippase
2	*weiA*	2421...3230	30.0	Glycosyltransferases group 2 (PF00535) *E* value=$2.7e^{-31}$	WbcL, *Yersinia enterocolitica* serotype O : 3 (CAA87700)	88/94 (289)	Glycosyltransferase
3	*wzy*	3319...4392	27.6	None	Wzy, *Shigella boydii* type 11 (AAS98031)	23/48 (192)	O-antigen polymerase
4	*weiB*	4403...5485	30.7	Glycosyltransferases group 1 (PF00534) *E* value=$5.2e^{-43}$	WbcM, *Y. enterocolitica* serotype O : 3 (CAA87701)	70/82 (419)	Glycosyltransferase
5	*weiC*	5496...6587	33.6	Glycosyltransferases group 1 (PF00534) *E* value=$2e^{-42}$	WbcN, *Y. enterocolitica* serotype O : 3 (CAA87702)	62/75 (370)	Glycosyltransferase
6	*weiD*	6574...7650	33.1	Glycosyltransferases group 1 (PF00534) *E* value=$6.6e^{-25}$	WbcQ, *Y. enterocolitica* serotype O : 3 (CAA87705)		Glycosyltransferase
7	*gne*	7699...8712	34.2	NAD-dependent epimerase/ dehydratase family (PF01370) *E* value=$6.1e^{-149}$	Gne, *E. coli* O107 (ACH97137)	60/78 (336)	UDP-GlcNAc-4-epimerase

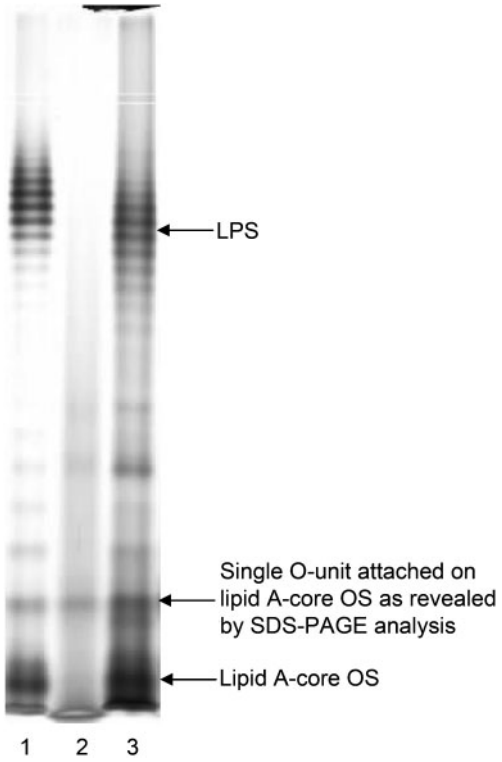

Fig. 3. SDS-PAGE analysis of deletion and *trans*-complemented *wzy* gene mutants of *E. coli* O166. Membrane extracts were run on SDS-PAGE gels and visualized by silver staining. Lane 1, strain G1216 (*E. coli* O166 type strain); lane 2, strain H1985 (G1216 lacking the *wzy* gene); lane 3, strain H1986 (H1985 containing plasmid pLW1539 carrying the *E. coli* O166 *wzy* gene).

of plasmid pLW1539 containing the *wzy* gene. SDS-PAGE showed that the ability to produce an S-type LPS with a long-chain O-polysaccharide was restored in the *trans*-complemented strain (Fig. 3). These data confirmed the function of the *wzy* gene in *E. coli* O166, as well as the identity of the *E. coli* O166 O-antigen gene cluster.

However, it is noted that the LPS modality for the *trans*-complemented strain slightly changed. The number of O-units attached to the core-lipid A has a characteristic modal distribution which is regulated by the chain length determinant gene *wzz* located between *gnd* and *his* (Franco *et al.*, 1998). The interaction and the balance between Wzy and Wzz is suggested to be important in determining the distribution of the repeat unit length (Bastin *et al.*, 1993; Daniels *et al.*, 1998). It is likely that the deletion and *trans*-complementation of the *wzy* gene in *E. coli* O166 affected the ratio of Wzy and Wzz, resulting in a decreased modal chain length.

Genetic relationship between the O-antigen gene clusters of *E. coli* O166 and *Salmonella* O66

To elucidate the genetic basis for the structural similarities between the O-antigens of *E. coli* O166 and *Salmonella* O66, the O-antigen gene cluster of *Salmonella* O66 was also sequenced and compared with that of *E. coli* O166. These two genomic regions contained the majority of the same genes and shared a DNA identity of between 64 and 70 % (Fig. 4). All genes in the O-antigen gene cluster of *Salmonella* O66 were assigned the same functions as the corresponding genes in *E. coli* O166. The only exception was that there was an 874 bp non-coding region between the *weiA* and *weiB* genes in *Salmonella* O66, while the *wzy* gene was present in this location in *E. coli* O166 (Fig. 4). Therefore, no O-antigen polymerase gene is present in the O-antigen gene cluster of *Salmonella* O66.

In *Salmonella* serogroups A, B and D1, the *wzy* genes responsible for the linkage between O-units are not located in the O-antigen gene cluster; instead they are mapped far from the O-antigen gene cluster in the *Salmonella* genome (Wang *et al.*, 2002). Since the *Salmonella* O66 type strain produces normal LPS, it is highly likely that this strain has a functional *wzy* gene for the $\beta 1 \rightarrow 2$ linkage outside the O-antigen gene cluster. The functional *wzy* gene of *Salmonella* O66 may be located and identified by whole-genome sequencing, which will be the subject of future study. It is possible that the ancestor of the O-antigen gene cluster of *Salmonella* O66 had an intact *wzy* gene for the $\beta 1 \rightarrow 3$ linkage located between the *weiA* and *weiB* genes, which became redundant when the bacteria acquired a new *wzy*

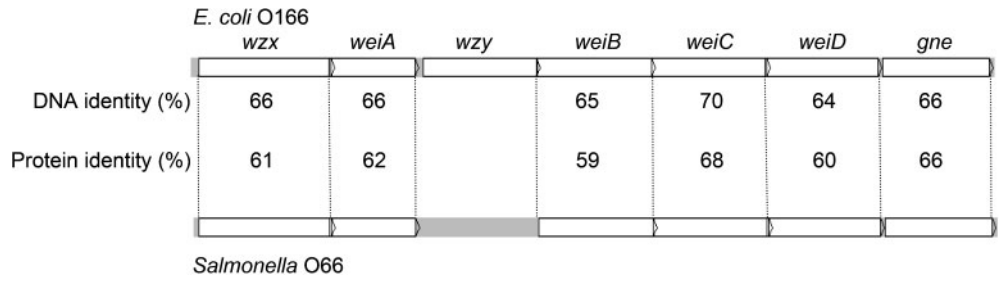

Fig. 4. Comparison of the O-antigen gene clusters of *E. coli* O166 and *Salmonella* O66. The non-coding region is indicated by a shaded area.

gene. The substantial degradation of the *wzy* gene sequence between the *weiA* and *weiB* genes in *Salmonella* O66 suggests that the introduction of the new *wzy* gene into the *Salmonella* O66 genome occurred a long time ago.

The genetic similarity between the O-antigen gene clusters of *E. coli* O166 and *Salmonella* O66 is consistent with the structural similarity between their O-antigens. Three cases are known in which the O-antigen structures are identical in *E. coli* and *Salmonella*: *E. coli* O55 and *Salmonella* O50, *E. coli* O157 and *Salmonella* O30, and *E. coli* O111 and *Salmonella* O35 (Samuel *et al.*, 2004). *Salmonella* O : 6,14 has been recently reported to share an O-unit backbone with the *E. coli* O77 group (Wang *et al.*, 2007). *E. coli* and *Salmonella* appear to have diverged about 140 million years ago, and 93 % of *E. coli* and *Salmonella* housekeeping genes exhibit levels of identity between 76.3 and 100 % (Sharp, 1991). A comparison between the O-antigen gene clusters of *E. coli* and *Salmonella* strains displaying identical O-antigen structures indicates that paired gene clusters share an obvious similarity, which is close to the lower end of the range for housekeeping genes (Samuel *et al.*, 2004; Wang *et al.*, 2007). These data imply that each pair of *E. coli*–*Salmonella*-related gene clusters possibly originates from a common ancestor which then diverged at a higher rate than the housekeeping genes under consistent selection pressure (Samuel *et al.*, 2004; Wang *et al.*, 2007). The similarity level between the O-antigen gene clusters of *E. coli* O166 and *Salmonella* O66 is similar to that previously found for other pairs of *E. coli*–*Salmonella*-related gene clusters. Therefore, we propose that the *E. coli* O166 and *Salmonella* O66 O-antigen gene clusters also have a common ancestor. After the species divergence, *Salmonella* O66 acquired a new *wzy* gene and a prophage gene for *O*-acetyl modification, both of which were located outside the O-antigen gene cluster, and the original *wzy* gene in the O-antigen gene cluster was lost over time. The resulting O-antigen structure, which is specific for *Salmonella* O66, may have aided adaptation of this organism to its environmental niche.

ACKNOWLEDGEMENTS

This work was supported by the Russian Foundation for Basic Research (Projects 08-04-01205 and 08-04-92225-NNSF), the Tianjin Municipal Special Fund for Science and Technology Innovation Grant 05FZZDSH00800, the National Natural Science Foundation of China (NSFC) Key Programs Grants 30530010 and 20536040, the Chinese National Science Fund for Distinguished Young Scholars (30788001), NSFC General Program Grant 30670038, 30870078, 30771175 and 30900041, the National 863 Program of China Grants 2006AA020703 and 2006AA06Z409, the National 973 Program of China Grant 2009CB522603, and National Key Programs for Infectious Diseases of China 2008ZX10004-002, 2008ZX10004-009 and 2009ZX10004-108.

REFERENCES

Alexander, D. C. & Valvano, M. A. (1994). Role of the *rfe* gene in the biosynthesis of the *Escherichia coli* O7-specific lipopolysaccharide and other O-specific polysaccharides containing *N*-acetylglucosamine. *J Bacteriol* **176**, 7079–7084.

Ali, T., Weintraub, A. & Widmalm, G. (2007). Structural determination of the O-antigenic polysaccharide from *Escherichia coli* O166. *Carbohydr Res* **342**, 274–278.

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.

Bastin, D. A., Brown, P. K., Haase, A., Stevenson, G. & Reeves, P. R. (1993). Repeat unit polysaccharides of bacteria: a model for polymerisation resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length. *Mol Microbiol* **7**, 725–734.

Daniels, C., Vindurampulle, C. & Morona, R. (1998). Overexpression and topology of the *Shigella flexneri* O-antigen polymerase (Rfc/Wzy). *Mol Microbiol* **28**, 1211–1222.

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.

Franco, A. V., Liu, D. & Reeves, P. R. (1998). The Wzz (Cld) protein in *Escherichia coli*: amino acid sequence variation determines O antigen chain length specificity. *J Bacteriol* **180**, 2670–2675.

Gemski, P., Jr, Sheahan, D. G., Washington, O. & Formal, S. B. (1972). Virulence of *Shigella flexneri* hybrids expressing *Escherichia coli* somatic antigens. *Infect Immun* **6**, 104–111.

Jansson, P.-E., Kenne, L. & Schweda, E. (1987). Nuclear magnetic resonance and conformational studies on monoacetylated methyl D-gluco- and D-galacto-pyranosides. *J Chem Soc Perkin 1* **2**, 377–383.

Leontein, K. & Lönngren, 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 ^{13}C-n.m.r. data. *Carbohydr Res* **175**, 59–75.

Liu, B., Senchenkova, S. N., Feng, L., Perepelov, A. V., Xu, T., Shevelev, S. D., Zhu, Y., Shashkov, A. S., Zou, M. & other authors (2006). Structural and molecular characterization of *Shigella boydii* type 16 O antigen. *Gene* **380**, 46–53.

Liu, B., Knirel, Y. A., Feng, L., Perepelov, A. V., Senchenkova, S. N., Wang, Q., Reeves, P. R. & Wang, L. (2008). Structure and genetics of *Shigella* O antigens. *FEMS Microbiol Rev* **32**, 627–653.

Reeves, P. R. (1992). Variation in O-antigens, niche-specific selection and bacterial populations. *FEMS Microbiol Lett* **79**, 509–516.

Robbins, P. W. & Uchida, T. (1962). Studies on the chemical basis of the phage conversion of O-antigens in the E-group *Salmonella*. *Biochemistry* **1**, 323–335.

Samuel, G. & Reeves, P. (2003). Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr Res* **338**, 2503–2519.

Samuel, G., Hogbin, J. P., Wang, L. & Reeves, P. R. (2004). Relationships of the *Escherichia coli* O157, O111, and O55 O-antigen gene clusters with those of *Salmonella enterica* and *Citrobacter freundii*, which express identical O antigens. *J Bacteriol* **186**, 6536–6543.

Sharp, P. M. (1991). Determinants of DNA sequence divergence between *Escherichia coli* and *Salmonella typhimurium*: codon usage, map position, and concerted evolution. *J Mol Evol* **33**, 23–33.

Skurnik, M., Venho, R., Bengoechea, J. A. & Moriyon, I. (1999). The lipopolysaccharide outer core of *Yersinia enterocolitica* serotype O:3 is required for virulence and plays a role in outer membrane integrity. *Mol Microbiol* **31**, 1443–1462.

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., Andrianopoulos, K., Liu, D., Popoff, M. Y. & Reeves, P. R. (2002). Extensive variation in the O-antigen gene cluster within one *Salmonella enterica* serogroup reveals an unexpected complex history. *J Bacteriol* **184**, 1669–1677.

Wang, W., Perepelov, A. V., Feng, L., Shevelev, S. D., Wang, Q., Senchenkova, S. N., Han, W., Li, Y., Shashkov, A. S. & other authors (2007). A group of *Escherichia coli* and *Salmonella enterica* O antigens sharing a common backbone structure. *Microbiology* **153**, 2159–2167.

Wang, Q., Perepelov, A. V., Feng, L., Knirel, Y. A., Li, Y. & Wang, L. (2009). Genetic and structural analyses of *Escherichia coli* O107 and O117 O-antigens. *FEMS Immunol Med Microbiol* **55**, 47–54.

Westphal, O. & Jann, K. (1965). Bacterial lipopolysaccharides. Extraction with phenol–water and further applications of the procedure. In *Methods in Carbohydrate Chemistry 5*, pp. 83–91. Edited by R. L. Whistler. New York, London: Academic Press.

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* **97**, 5978–5983.

Edited by: D. L. Gally