Available online at www.sciencedirect.com
ScienceDirect
Biochemical and Biophysical Research Communications 349 (2006) 289–295
BBRC
www.elsevier.com/locate/ybbrc

Structural and genetic characterization of *Shigella boydii* type 17 O antigen and confirmation of two new genes involved in the synthesis of glucolactilic acid

Sof'ya N. Senchenkova [a,1], Lu Feng [b,c,1], Quan Wang [b,c,1], Andrei V. Perepelov [a], Dan Qin [b,c], Sergei D. Shevelev [a], Yi Ren [b,c], Alexander S. Shashkov [a], Yuriy A. Knirel [a], Lei Wang [b,c,*]

[a] *N.D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Moscow 119991, Russian Federation*
[b] *TEDA School of Biological Sciences and Biotechnology, Nankai University, 23 HongDa Street, TEDA, Tianjin 300457, PR China*
[c] *Tianjin Key Laboratory of Microbial Functional Genomics, TEDA College, Nankai University, 23 HongDa Street, TEDA, Tianjin 300457, PR China*

Received 3 August 2006
Available online 15 August 2006

Abstract

Shigella strains are human pathogens and normally identified based on their O antigens. The chemical structure and gene cluster of *Shigella boydii* type 17 O antigen were studied. As judged by sugar and methylation analyses along with NMR spectroscopy data, the O antigen of *S. boydii* type 17 has a linear trisaccharide O unit, which consists of two residues of *N*-acetylgalactosamine (GalNAc) and a 4-*O*-[(*R*)-1-carboxyethyl]-D-glucose (glucolactilic acid). The O antigen gene cluster of *S. boydii* type 17 was sequenced and genes encoding UDP-*N*-acetylglucosamine C4 epimerase for GalNAc synthesis, O unit flippase, O antigen polymerase, and glycosyltransferases were putatively identified based on sequence similarities and the presence of conserved motifs. Two genes, whose functions could not be clearly indicated by homology search, were confirmed to be involved in the synthesis of glucolactilic acid by mutation and structural verification of the O antigens from the mutants. To our knowledge, this is the first time that genes involved in the synthesis of glucolactilic acid have been reported. Two genes specific to *S. boydii* type 17 were also identified.

Keywords: *Shigella boydii* type 17; Lipopolysaccharide; O antigen structure; O antigen gene cluster; 4-*O*-[(*R*)-1-carboxyethyl]-D-glucose

Shigella strains are well-known human pathogens causing diseases such as diarrhea and bacillary dysentery [1]. Four species including *Shigella boydii*, *Shigella dysenteriae*, *Shigella flexneri*, and *Shigella sonnei* have been recognized [2]. Subdivision of *Shigella* strains into different serotypes is solely based on their O antigens as they lack flagellar (H) and capsular (K) antigens. There are 46 *Shigella* serotypes, but only 33 distinct O antigen forms, as the rest are variants arising from phage modification. *S. boydii* type 17, first described as a provisional *Shigella* serotype [3], was added to the *Shigella* schema based on biochemical and serological studies in 1985 [4].

The O antigen (O polysaccharide), which consists of many repeats of an oligosaccharide unit (O unit), is the outer component of lipopolysaccharide (LPS) in the surface of Gram-negative bacteria [5]. It is one of the most variable cell constituentes due to presence of different sugars and sugar linkages. The O antigen variation plays an important role in bacterial evasion of host defense systems [6]. A recent study showed that O antigen was also essential for the virulence of *Yersinia enterocolitica* O:8 [7].

Genes involved in O antigen synthesis are normally clustered between two housekeeping genes *galF* and *gnd* in *Shigella*, *Escherichia coli*, and *Salmonella*, and are commonly

* Corresponding author. Fax: +86 22 66229596.
E-mail address: wanglei@nankai.edu.cn (L. Wang).
[1] These authors contributed equally to this report.

0006-291X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2006.08.040

classified into three main classes: sugar biosynthetic pathway genes, sugar transferase genes, and O antigen processing genes including the flippase (Wzx) and polymerase (Wzy) genes [5]. The different O antigen forms are almost entirely due to genetic variations in their respective O antigen gene clusters.

In this study, we investigated the chemical structure and gene cluster of *S. boydii* type 17 O antigen. The O unit of *S. boydii* type 17 was shown to contain two residues of *N*-acetylgalactosamine (GalNAc) and a 4-*O*-[(*R*)-1-carboxyethyl]-D-glucose (glucolactilic acid). The O antigen gene cluster of *S. boydii* type 17 was found to contain genes involved in the synthesis of GalNAc and glucolactilic acid, genes encoding putative sugar transferases, O unit flippase (Wzx) and O antigen polymerase (Wzy). By generating the mutants and examining the O antigen structures of the mutants, two genes were confirmed, for the first time, to be involved in the synthesis of glucolactilic acid. Genes specific to *S. boydii* type 17 were also identified by screening against representative strains of all 186 *Shigella* and *E. coli* O serotypes.

Materials and methods

Bacterial strains and plasmids. All plasmids used in this study were maintained in *E. coli* K-12 strain DH5α, which was purchased from Beijing Dingguo Biotechnology Development Center (Beijing, China). *S. boydii* type 17 type strain LSPQ 3692 (laboratory stock No. G1214) was provided by the Institute of Medical and Veterinary Science, Adelaide, Australia. Other *Shigella* and *E. coli* type strains used were as previously reported [8].

Cultivation of bacteria and isolation of lipopolysaccharides. Bacteria were grown to late log phase in 8 L of Luria–Bertani broth in a 10-L fermentor (BIOSTAT C-10, B. Braun Biotech International, Germany) under constant aeration at 37 °C. LPSs of *S. boydii* type 17, *wfaW*, and *wfaX* knock-out mutants (824, 280, and 560 mg, respectively) were isolated from dried cells (20.6, 8.7, and 8.2 g, respectively) by the phenol–water method [9] and purified by precipitation of nucleic acids and proteins with aqueous 50% trichloroacetic acid.

Preparation of the O polysaccharides. Delipidation of the LPSs of *S. boydii* type 17, *wfaW*, and *wfaX* knock-out mutants (103, 100, and 91 mg, respectively) was performed with aq 2% HOAc 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, Sweden) in 0.05 M pyridinium acetate buffer (pH 4.5) monitored using a Knauer differential refractometer (Germany). High-molecular-mass polysaccharides were obtained in yields of 50%, 27%, and 25%, respectively, of the LPS weight.

Chemical analyses. The O polysaccharide of *S. boydii* type 17 (1 mg) was hydrolyzed with 2 M CF_3CO_2H (120 °C, 2 h) and sugars were identified by GLC of the alditol acetates on a Hewlett-Packard Model 5880 chromatograph (USA) equipped with an Ultra-2 column (Supelco, USA) using a temperature gradient of 160 to 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 [10]. Methylation of the polysaccharide (1 mg) was performed with CH_3I in dimethyl sulfoxide in the presence of sodium methylsulfinylmethanide [11]. A portion of the methylated polysaccharide was reduced with $LiBH_4$ in aq 70% 2-propanol (20 °C, 2 h). Partially methylated monosaccharides were derived by hydrolysis under the same conditions as in sugar analysis, reduced with $NaDH_4$, acetylated, and analyzed by GLC-MS on a Hewlett-Packard HP 5989A instrument equipped with a 30-m HP-5 ms column (Hewlett-Packard, USA) under the same chromatographic conditions as in GLC.

Isolation of glucolactilic acid. The O polysaccharide of *S. boydii* type 17 (70 mg) was hydrolyzed with CF_3CO_2H at 120 °C for 2 h. After evaporation, products were fractionated on a column (80 × 1.6 cm) of TSK HW-40 (S) (Merck, Germany) in water to give D-Glc4*R*Lac (10 mg). The authentic sample of D-Glc4*R*Lac (3 mg) was isolated in a similar manner from the O polysaccharide of *S. dysenteriae* type 3 (25 mg).

NMR spectroscopy. Samples were deuterium-exchanged by freeze-drying twice from D_2O and then examined as solutions in 99.96% D_2O at 30 °C. NMR spectra were recorded on a Bruker DRX-500 spectrometer (Germany) using internal acetone (δ_H 2.225, δ_C 31.45) as reference. Two-dimensional NMR spectra were obtained using standard Bruker software, and Bruker XWINNMR 2.6 program was used to acquire and process the NMR data. Mixing times of 200 and 300 ms were used in TOCSY and ROESY experiments, respectively.

Construction of a random DNaseI shot gun bank. Chromosomal DNA was prepared as previously described [12]. Primers #1523 and #1524 [13] based on the *galF* and *gnd* genes, respectively, were used to amplify the DNA of *S. boydii* type 17 O antigen gene cluster using the Expand Long Template PCR system from Roche. The PCR products were digested with DNaseI and the resulting DNA fragments were cloned into pGEM-T easy to produce a bank using the method described previously [14].

Sequencing and analysis. The plasmid DNA template for sequencing was prepared using the method of Sambrook [15]. Sequencing was carried out using an ABI 3730 automated DNA sequencer (ABI, Foster City, USA). Sequence data were assembled using the Staden Package [16]. The program Artemis [17] was used to identify open-reading frames (ORFs) and for annotation. The program BlockMaker was used to search conserved motifs. Blast and PSI-BLAST [18] were used for searching databases including GenBank, COG, and Pfam protein motif databases [19,20]. The algorithm described by Eisenberg [21] was used to identify potential transmembrane segments. Sequence alignment and comparison were performed using the program ClustalW. (http://www.ebi.ac.uk/clustalw).

Deletion of wfaW and wfaX from S. boydii type 17. *wfaW* and *wfaX* were replaced, respectively, by a chloramphenicol acetyltransferase (CAT) gene using the RED recombination system of phage lambda [22,23]. The CAT gene was PCR-amplified from plasmid pKK232-8 (Pharmacia) using primer pairs wl-4462/wl-4340 (5′-ATAATTATAAAAGTATTGTCTTTTTTCGGGGGGAATAATGATGGAGAAAAAAATCACTGGAT-3′/ 5′-TTTGCACCTTTAGCTTTCAAATACTTCATTATATAATACATTACGCCCCGCCCTGCCACTCA-3′) and wl-4463/wl-4342 (5′-ATATAATGAAGTATTTGAAAGCTAAAGGTGCAAAATGAGTATGGAGAAAAAAATCACTGGAT-3′/ 5′-CTAGTTTAGGTTCGTTTTTAAGAAACGTATAGAGTTTCCCTTACGCCCCGCCCTGCCACTCA-3′) binding to the 5′ and 3′ ends of *wfaW* and *wfaX*, respectively, with each primer carrying 40bp based on the *S. boydii* type 17 DNA which flanks the target gene. The PCR product was transformed into the *S. boydii* type 17 strain G1214 carrying pKD20, and chloramphenicol-resistant transformants were selected after induction of the RED genes according to the protocol described by Datsenko and Wanner [22]. PCR using primers specific to the CAT gene and the flanking DNA of the target gene was carried out to confirm the replacement. *wfaW* and *wfaX* knock-out mutants were designated as *S. boydii* type 17 strains H1438 and H1435, respectively. To complement the mutants, *wfaW* and *wfaX* were PCR-amplified from the wild-type *S. boydii* type 17 strain G1214 using primer pairs wl-5696/wl-4537 (5′-CATGCCATGGGCATGAGTATTCAAAAGGTAAT-3′/5′-ACTGGGATCCTCATTTTGCACCTTTAGCTT-3′) and wl-5697/wl-4539 (5′-CATGCCATGGGCATGAGTAAGAAGGTTGGTAT-3′/ 5′-ACTGGGATCCCTAGTTTAGGTTCGTTTTTA-3′), respectively. The resulting PCR products were cloned into pTRC99A to make plasmids pLW1201 (containing *wfaW*) and pLW1200 (containing *wfaX*), which were transformed into *S. boydii* type 17 strain to construct the strains H1439 and H1436, respectively. Expression of the cloned *wfaW* and *wfaX* genes was induced by 0.5 mM IPTG.

Membrane preparation, SDS–PAGE, and silver staining for visualization of LPS were carried out as described by Wang and Reeves [24].

PCR specificity assay. Chromosomal DNA was prepared from each of the 186 *E. coli* (including *Shigella*) strains representing all O serotypes.

The quality of DNA was examined by PCR amplification of the *mdh* gene (coding for malate dehydrogenase and presents as a housekeeping gene in *E. coli* and *Shigella*) using primers described previously [14]. A total of 13 DNA pools were made, each containing DNA from 12 to 19 strains [8]. Pools were screened by PCR using primers based on genes specific to *S. boydii* type 17.

Nucleotide sequence Accession No.. The DNA sequence of the *S. boydii* type 17 O antigen gene cluster has been deposited in GenBank under the Accession No. DQ875941.

Results and discussion

Elucidation of the O polysaccharide structure of S. boydii type 17

The O polysaccharide was obtained by mild acid degradation of the LPS isolated from dried cells by the phenol–water procedure. Analyses by GLC of the alditol acetates and (*S*)-2-octyl glycosides revealed D-GalN. Methylation analysis of the O polysaccharide resulted in identification of 3-substituted and 4-substituted GalN residues in the ratio ~1:1. When the methylated polysaccharide was carboxyl-reduced prior to hydrolysis, two additional compounds, major and minor, were detected. Based on the characteristic mass spectrometric fragmentation and published data [25], they were identified as derivatives **1** and **2** from 6-substituted and non-substituted 4-*O*-(1-carboxyethyl)hexose, respectively (compounds **1**, **2** in Fig. 1). Compound **2** was derived evidently from the monosaccharide that occupies the non-reducing end of the polysaccharide and, hence, the biological O unit.

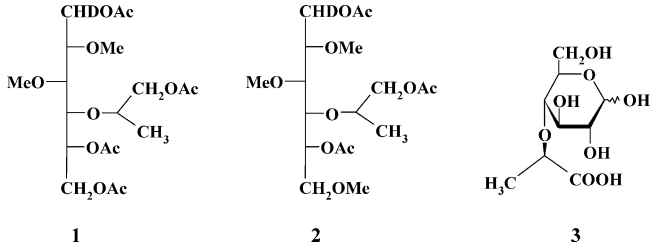

Fig. 1. Structures of compounds derived from the O polysaccharide of *S. boydii* type 17 in methylation analysis (**1** and **2**) and by acid hydrolysis (**3**).

The O polysaccharide was hydrolyzed with CF_3CO_2H and the 4-*O*-(1-carboxyethyl)hexose was isolated by fractionation of products using gel-permeation chromatography on TSK HW-40 (S). Assignment of the ^1H and ^{13}C NMR spectra (Table 1) confirmed that this compound is an ether of a hexose and lactic acid. Relative coupling constants of the neighboring protons ($J_{2,3} \approx J_{3,4} \approx J_{4,5}$ 8–10 Hz) indicated that the hexose is glucose. The NMR spectra were indistinguishable from those of 4-*O*-[(*R*)-1-carboxyethyl]-D-glucose (D-Glc4*R*Lac) (compound **3** in Fig. 1), which was isolated from the O polysaccharide of *S. dysenteriae* type 3 studied earlier [25,26].

The ^{13}C NMR spectrum of the O polysaccharide of *S. boydii* type 17 (Fig. 2, A) contained signals for three anomeric carbons at δ 95.0–105.2, two nitrogen-bearing carbons (C-2 of two GalN) at δ 51.1–52.1, one O-substituted and two non-substituted $HOCH_2$-C groups (C-6 of Glc and two GalN) at δ 69.0 and 62.2–62.6 (data of the attached proton test), one CH_3-C group (C-3 of lactic acid) at δ 19.9, COOH group of lactic acid at δ 179.1, 11 other carbons in the region δ 65.1–78.6 (oxygen-linked sugar ring carbons and C-2 of lactic acid), and two *N*-acetyl groups at δ 23.5, 23.9 (both CH_3), 175.8, 176.2 (both CO). Therefore, the O polysaccharide has a trisaccharide O unit containing two residues of GalNAc ($GalNAc^I$ and $GalNAc^{II}$) and a Glc4*R*Lac.

The ^1H and ^{13}C NMR spectra were assigned using two-dimensional COSY, TOCSY, ROESY, and H-detected ^1H, ^{13}C HSQC experiments, and spin systems for three mono-

Table 1
^1H and ^{13}C NMR data of the O polysaccharides of *S. boydii* type 17 and mutant H1438 (δ, ppm)

Sugar residue		1	2	3	4	5	6a,b	CH_3CO	CH_3CO	1′	2′	3′
S. boydii type 17												
→6)-β-D-Glcp4RLac-(1→	^1H	4.64	3.37	3.56	3.40	3.59	3.83, 4.23				4.27	1.43
	^{13}C	105.2	75.0	76.6	78.6	75.0	69.0			179.1	77.9	19.9
→4)-α-D-Gal*p*NAc^I-(1→	^1H	5.07	4.29	3.87	4.24	3.84	3.77, 3.83	2.04				
	^{13}C	95.0	51.1	69.5	78.6	72.2	62.2	23.5	176.2			
→3)-β-D-Gal*p*NAc^{II}-(1→	^1H	4.53	4.07	3.79	4.10	3.64	3.77, 3.82	2.12				
	^{13}C	102.1	52.1	76.7	65.1	76.3	62.6	23.9	175.8			
S. boydii type 17 mutant H1438												
→6)-β-D-Glc*p*-(1→	^1H	4.65	3.37	3.49	3.40	3.59	3.89, 4.17					
	^{13}C	105.3	75.0	77.2	71.0	76.2	69.7					
→4)-α-D-Gal*p*NAc^I-(1→	^1H	5.08	4.31	3.86	4.26	3.86	3.75, 3.83	2.04				
	^{13}C	95.2	51.5	70.0	78.6	72.2	62.0	23.4	176.0			
→3)-β-D-Gal*p*NAc^{II}-(1→	^1H	4.56	4.09	3.81	4.11	3.65	3.77, 3.82	2.09				
	^{13}C	103.2	52.1	76.7	65.3	76.6	62.4	23.9	175.8			
4-*O*-[(*R*)-1-carboxyethyl]-D-glucose												
α-D-Glc4*R*Lac	^1H	5.19	3.52	3.80	3.37	3.85	3.84, 3.84		1.38		4.43	1.37
	^{13}C	93.3	73.0	74.5	78.7	71.8	62.2			n.d.	78.8	20.0
β-D-Glc4*R*Lac	^1H	4.60	3.24	3.59	3.37	3.47	3.79, 3.86		1.37		4.41	1.38
	^{13}C	97.4	75.8	77.5	78.4	76.3	62.3			n.d.	78.7	20.0

n.d., not determined.

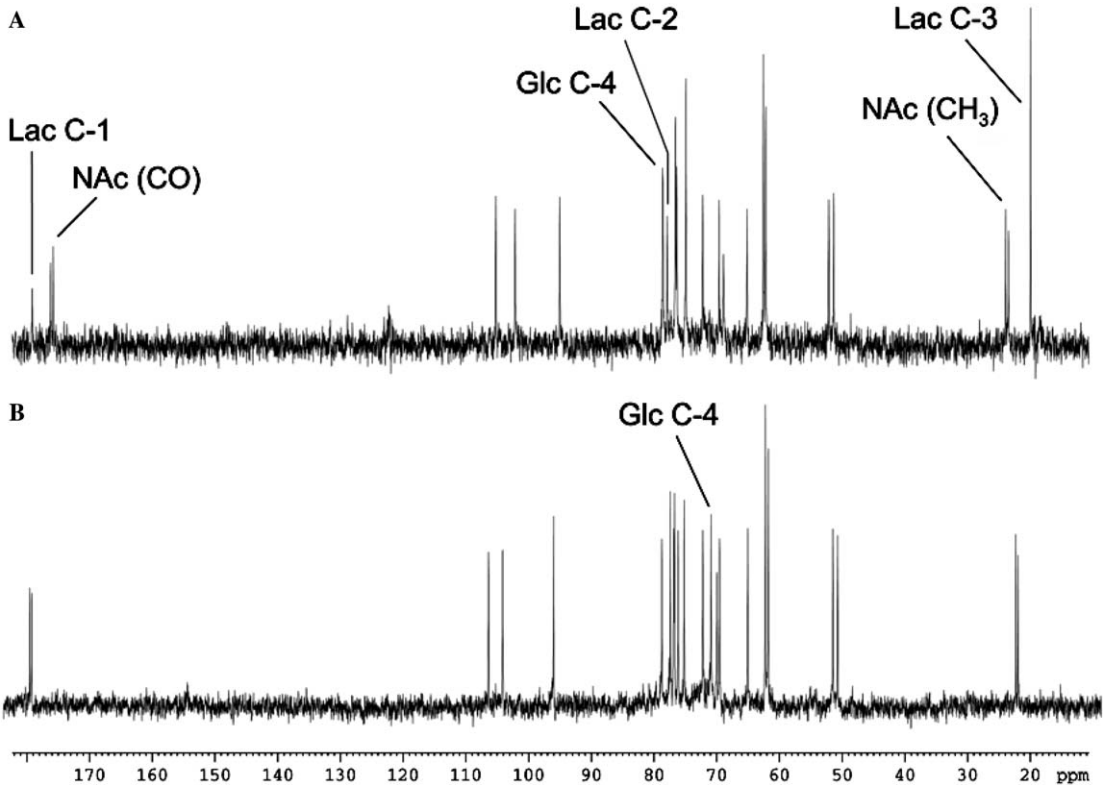

Fig. 2. One hundred and twenty-five megahertz ^{13}C NMR spectra of the O polysaccharides of *S. boydii* type 17 (A) and mutant strain H1438 (B). Lac (*R*)-1-carboxyethyl [(*R*)-lactic acid residue].

saccharide residues and 1-carboxyethyl group were recognized (Table 1). The $J_{1,2}$ coupling constant values of ~3 Hz indicated that GalNAc^{I} is α-linked, whereas GalNAc^{II} and Glc that are characterized by $J_{1,2} \sim 9$ Hz are β-linked. The pyranose form of all sugar residues was inferred by the absence of any ^{13}C NMR signals for non-anomeric sugar ring carbons in the region δ 80–88 [26].

Low-field displacements of signals for C-6 of Glc, C-4 of GalNAc^{I}, and C-3 of GAlNAc^{II} at δ 69.0, 78.6, and 76.7, respectively, as compared with their positions in the corresponding non-substituted monosaccharides [27], confirmed the modes of glycosylation of the GalNAc residues and demonstrated substitution of Glc4*R*Lac at position 6. A ROESY experiment revealed cross-peaks between anomeric protons and protons at the linkage carbons (Glc H-1/GalNAc^{I} H-4, GalNAc^{I} H-1/GalNAc^{II} H-3, and GalNAc^{II} H-1/Glc H-6a, 6b at δ 4.64/4.24, 5.07/3.79 and 4.53/3.83, 4.23, respectively), thus showing the monosaccharide sequence in the O unit.

A cross-peak between H-2 of lactic acid and H-4 of Glc at δ 4.27/3.40 in the ROESY spectrum confirmed the Glc4*R*Lac structure. The *R* configuration of lactic acid was confirmed by correlation between H-3 of lactic acid and H-3 and H-4 of Glc at δ 1.43/3.56 and 1.43/3.40, which is in agreement with molecular modeling data (in case of the *S* configuration, H-3 of lactic acid would be spatially close to H-6 of Glc) (author's unpublished data).

Taking into account these data together, it was concluded that the biological O unit of *S. boydii* type 17 has the structure shown in Fig. 3A.

Sequencing of the region between galF and gnd from S. boydii type 17

A sequence of 10,432 bases including *galF* and *gnd* was obtained. Seven ORFs excluding *galF* and *gnd* were identified, and all transcribes from *galF* to *gnd* (Fig. 4).

A →6)-β-D-Glc*p*-(1→4)-α-D-Gal*p*NAc^{I}-(1→3)-β-D-Gal*p*NAc^{II}-(1→

4

|

*R*Lac

B →6)-β-D-Glc*p*-(1→4)-α-D-Gal*p*NAc^{I}-(1→3)-β-D-Gal*p*NAc^{II}-(1→

Fig. 3. Structures of the O polysaccharides of *S. boydii* type 17 (A), mutant strains H1438 and H1435 (B).

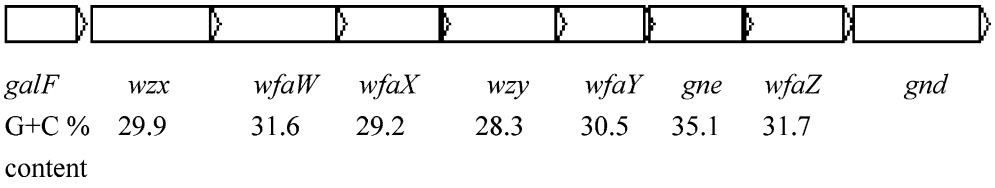

Fig. 4. The O antigen gene cluster organization of *S. boydii* type 17.

Table 2
Characteristics of the ORFs in the *S. boydii* type 17 O antigen gene cluster

Gene	Base positions	% G + C content	Similar protein, strain, (GenBank Accession Number)	% Identity/% similarity (No. of aa overlap)	Putative function
wzx	926–2191	29.9	Wzx, *Yersinia enterocolitica* type O:8 (AAC60766)	21/48 (380)	O unit flippase
wfaW	2191–3555	31.6	Coenzyme F_{420}-reducing hydrogenase, *Methanothermobacter thermautotrophicus* (AAB84847)	28/49 (302)	Hydrogenase
wfaX	3552–4655	29.2	MurB family protein, *Enterococcus faecalis* V583 (AAO82205)	28/46 (330)	Pyruvyl transferase
wzy	4681–5892	28.3	O antigen polymerase, *Salmonella enterica* C1 (AAB49386)	27/46 (357)	O antigen polymerase
wfaY	5909–6850	30.5	WbnA, *Escherichia coli* O113 (AAD50485)	33/50 (276)	Glycosyltransferase
gne	6890–7906	35.1	Gne, *Yersinia enterocolitica* type O:8 (AAC60777)	59/71 (337)	UDP-*N*-acetylglucosamine 4-epimerase
wfaZ	7929–8987	31.7	WbcQ, *Yersinia enterocolitica* type O:3 (CAA87705)	44/63 (337)	Glycosyltransferase

The ORFs were assigned functions based on their sequence similarities to genes in available databases (Table 2).

(i) *Sugar biosynthetic pathway genes*. *orf6* shares 59% identity to *gne* of *Yersinia enteritica* O:8. In *Y. enteritica* O:8, *gne* encodes UDP-GlcNAc C4 epimerase, which catalyzes the conversion of GlcNAc to GalNAc [28]. o*rf6* was assigned the same function in *S. boydii* type 17 and named *gne*. The substrate of Gne, UDP-GlcNAc, is a common sugar in bacteria, and the genes responsible for its synthesis are located elsewhere in the chromosome.

The proteins encoded by *orf2* belongs to the FrhB_FdhB_C family (PF04432, E-value $= 4.3 \times e^{-7}$), which contains the C termini of F_{420} hydrogenase and dehydrogenase beta subunits. The protein encoded by *orf3* belongs to the Polysaccharide pyruvyl transferase family (PF04230 E-value $= 6.3 \times e^{-4}$). *orf2* and *orf3* were confirmed to be involved in the synthesis of glucolactilic acid by mutation tests (see below), and named *wfaW* and *wfaX*, respectively.

(ii) *Sugar transferase genes*. The O unit of *S. boydii* type 17 contains two residues of GalNAc and a residue of glucolactilic acid. GlcNAc or GalNAc is usually the first sugar to be added onto the carrier lipid undecaprenyl phosphate (UndP) in *E. coli* and *Shigella*, and *wecA*, which is located outside the O antigen gene cluster, is responsible for this linkage [29]. It is likely that, one of the GalNAc residues acts as the first sugar in the *S. boydii* type 17 O unit, therefore, only two transferase genes were expected in its O antigen gene cluster.

The protein encoded by *orf5* and *orf7* belong to the families of Glycosyltransferase group 2 (PF00535, E-value $= 9.2 \times e^{-35}$) and Glycosyltransferase group 1 (PF00534, E-value $= 7.7 \times e^{-26}$), respectively. *orf5* shares 33% identity and 50% similarity with WbnA, a proposed glycosyltransferase for the synthesis of *E. coli* O113 O antigen [30]. *orf7* shares 44% identity and 63% similarity with WbcQ, which was proposed to function as either a Gal or GalNAc transferase in *Y. enterocolitica* O:3 [31]. *orf5* and *orf7* were proposed to encode glycosyltransferases and named *wfaY* and *wfaZ*, respectively.

(iii) *O unit processing genes*. The protein encoded by *orf1* was found to have 12 predicted transmembrane segments, which is a feature of Wzx proteins [32]. Orf1 belongs to the Polysaccharide biosynthesis protein family (PF01943, E-value $= 1.0 \times e^{-5}$). It also shares 22 and 21% identity to the putative Wzx proteins of *Clostridium acetobutylicum*

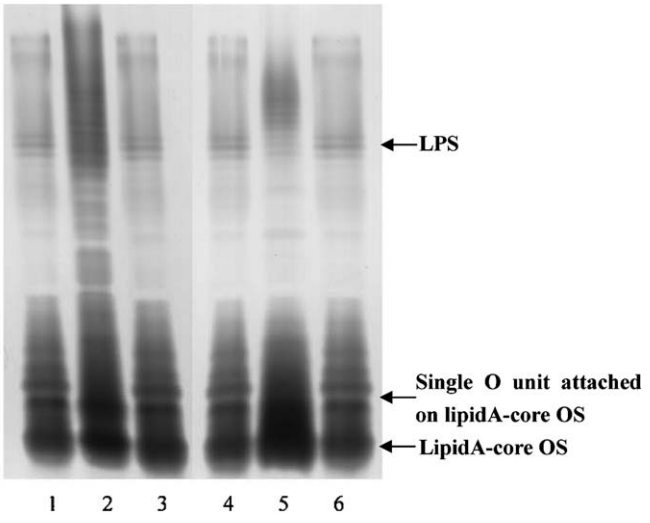

Fig. 5. Deletion and complementation analysis of *S. boydii* type 17 *wfaW* and *wfaX*. Membrane extracts were run on gels and silver stained. Lane 1, *S. boydii* type 17 type strain G1214. Lane 2, H1438 (G1214 missing the *wfaW* gene). Lane 3, H1439 (H1438 with plasmid pLW1201 containing *wfaW*). Lane 4, *S. boydii* type 17 type strain G1214. Lane 5, H1435 (G1214 missing the *wfaX* gene). Lane 6, H1436 (H1435 with plasmid pLW1200 containing *wfaX*).

Table 3
PCR testing of *S. boydii* type 17 specific genes

Gene	Base positions	Forward primer (base positions)	Reverse primer (base positions)	Length of PCR fragment (bp)	Annealing temperature (°C) of PCR
wzx	926–2191	wl-361 (1043–1061)	wl-362 (2018–2035)	993	52[a]
		wl-363 (1137–1154)	wl-364 (1875–1892)	756	55
wzy	4681–5892	wl-365 (4772–4789)	wl-366 (5306-5323)	552	52
		wl-367 (4967–4984)	wl-368 (5562–5579)	613	52

[a] PCR condition used: denaturation at 94 °C for 15 s, annealing at 52 °C (or 55 °C) for 30 s, and extension at 72 °C for 1 min, 30 cycles.

ATCC 824 and *Y. enterocolitica* O:8, respectively. When Orf1 and the two putative Wzx proteins were analyzed using the BlockMaker program, four conserved motifs were revealed (14, 20, 14, and 17 amino acids, respectively). The consensus sequences of those motifs were used to run the PSI-BLAST program to search the GenBank database, and other distantly related Wzx proteins were also retrieved (E-value $= 4 \times e^{-27}$) after three iterations. *orf1* was therefore proposed to be an O unit flippase gene, *wzx*, and named accordingly.

The protein encoded by *orf4* was found to have 12 predicted transmembrane segments with a large periplasmic loop of 20 amino acid residues, which is the typical topological character of Wzy proteins [33]. It also shares 25% and 24% identity to Wzy proteins of *Salmonella enterica* C1 and *S. boydii* type 11, respectively. *orf4* was therefore proposed to be an O antigen polymerase gene, *wzy*, and named accordingly.

wfaW and wfaX genes are involved in the synthesis of glucolactilic acid

Two mutant strains H1438 and H1435, in which *wfaW* and *wfaX* were replaced by a CAT gene, respectively, produced different LPS band patterns from that of the *S. boydii* type 17 wild-type strain as revealed by silver-stained SDS–PAGE (Fig. 5). When the two mutants were complemented by plasmids pLW1201 (containing *wfaW*) and pLW1200 (containing *wfaX*), respectively, the same LPS phenotype as produced by *S. boydii* type 17 wide type strain was restored (Fig. 5). Those data indicated that mutation in either the *wfaW* or *wfaX* gene resulted in structural change of the O antigen.

O polysaccharides from the two mutants were studied by ^{1}H and ^{13}C NMR spectroscopy as described above for the *S. boydii* type 17 O polysaccharide. The ^{13}C NMR spectra (Fig. 2B) were found to be identical to each other and to lack signals for lactic acid in both. Full assignment of the NMR spectra enabled structure elucidation of the two mutant O units, which are both linear and contain two residues of GalNAc and one residue of Glc (Fig. 3B). Glucose was found to have the D configuration, thus further confirming the absolute configuration of Glc4*R*Lac in the parent O polysaccharide. The data obtained showed that mutation in either of the *wfaW* and *wfaX* genes affected the glucolactilic acid synthesis pathway.

Although both *wfaW* and *wfaX* were confirmed to be involved in the biosynthesis of glucolactilic acid, we could not be sure of the exact process of its synthesis by the present experimental data, and the enzymes of the glucolactilic acid biosynthetic pathway need to be characterized biochemically in future study.

Identification of the S. boydii type 17 specific genes

wzx and *wzy* genes usually display a low level of similarity in terms of their primary sequences [34], so they are commonly used as specific genes for identification of individual O antigens. A total of four pairs of primers based on O unit processing genes *wzx* and *wzy* (two pairs for each gene) were designed (Table 3) and used to screen DNA pools containing representatives of the 186 known O antigen forms of *E. coli* and *Shigella* strains. With either primer pair, except the pools containing *S. boydii* type 17, which gave the expected PCR products, no PCR products were detected. Therefore, all of the four primer pairs are specific to *S. boydii* type 17 and can be used to develop a PCR assay for identification and detection of *S. boydii* type 17 strains rapidly.

Acknowledgments

This work was supported by grants of the Russian Foundation for Basic Research (03-04-39020), the President of the Russian Federation for Support of Young Russian Scientists (MK-1597.2005.3), the Foundation for Support of National Science (Russia), the NSFC (the National Natural Science Fundation of China) Key Programs (30530010, 20536040), NSFC General Programs (30370023, 30370339), and Tianjin Municipal Fund for Science and Technology Innovation (05FZZDSH00800).

References

[1] K. Wachsmuth, G.K. Morris, *Shigella*, in: M.P. Doyle (Ed.), Foodborne Bacterial Pathogens, Marcel Dekker Inc., New York, 1989, pp. 447–462.
[2] W.H. Ewing, *Shigella* nomenclature, J. Bacteriol. 57 (1949) 633–638.
[3] W.H. Ewing, R.W. Reavis, B.R. Davis, Provisional *Shigella* serotypes, Can. J. Microbiol. 4 (1958) 89–107.
[4] H.G. Wathen-Grady, B.R. Davis, G.K. Morris, Addition of three new serotypes of *Shigella boydii* to the *Shigella* schema, J. Clin. Microbiol. 21 (1985) 129–132.

[5] P.R. Reeves, L. Wang, Genomic organization of LPS-specific loci, Curr. Top. Microbiol. Immunol. 264 (2002) 109–135.
[6] P.R. Reeves, Role of O-antigen variation in the immune response, Trends Microbiol. 3 (1995) 381–386.
[7] J.A. Bengoechea, H. Najdenski, M. Skurnik, Lipopolysaccharide O antigen status of *Yersinia enterocolitica* O:8 is essential for virulence and absence of O antigen affects the expression of other *Yersinia* virulence factors, Mol. Microbiol. 52 (2004) 451–469.
[8] L. Feng, W. Wang, J. Tao, H. Guo, G. Krause, L. Beutin, L. Wang, Identification of *Escherichia coli* O114 O-antigen gene cluster and development of an O114 serogroup-specific PCR assay, J. Clin. Microbiol. 42 (2004) 3799–3804.
[9] O. Westphal, K. Jann, Bacterial lipopolysaccharides. Extraction with phenol- water and further applications of the procedure, in: R.L. Whistler (Ed.), Methods in Carbohydrate Chemistry, vol. 5, Academic Press, New York and London, 1965, pp. 83–91.
[10] K. Leontein, J. Lonngren, Determination of the absolute configuration of sugars by gas–liquid chromatography of their acetylated 2-octyl glycosides, Methods Carbohydr. Chem. 9 (1993) 87–89.
[11] H.E. Conrad, Methylation of carbohydrates with methyl iodide in dimethyl sulfoxide in the presence of methylsulfinylanion, Methods Carbohydr. Chem. 6 (1972) 361–364.
[12] D.A. Bastin, P.R. Reeves, Sequence and analysis of the O antigen gene (*rfb*) cluster of *Escherichia coli* O111, Gene 164 (1995) 17–23.
[13] L. Wang, W. Qu, P.R. Reeves, Sequence analysis of four *Shigella boydii* O-antigen loci: implication for *Escherichia coli* and *Shigella* relationships, Infect. Immun. 69 (2001) 6923–6930.
[14] L. Wang, P.R. Reeves, Organization of *Escherichia coli* O157 O antigen gene cluster and identification of its specific genes, Infect. Immun. 66 (1998) 3545–3551.
[15] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989.
[16] R. Staden, The Staden sequence analysis package, Mol. Biotech. 5 (1996) 233–241.
[17] K. Rutherford, J. Parkhill, J. Crook, T. Horsnell, P. Rice, M.A. Rajandream, B. Barrell, Artemis: sequence visualisation and annotation, Bioinformatics 16 (2000) 944–945.
[18] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3398–3402.
[19] R.L. Tatusov, D.A. Natale, I.V. Garkavtsev, T.A. Tatusova, U.T. Shankavaram, B.S. Rao, B. Kiryutin, M.Y. Galperin, N.D. Fedorova, E.V. Koonin, The COG database: new developments in phylogenetic classification of proteins from complete genomes, Nucleic Acids Res. 29 (2001) 22–28.
[20] A. Bateman, E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S.R. Eddy, S. Griffiths-Jones, K.L. Howe, M. Marshall, E.L. Sonnhammer, The Pfam protein families database, Nucleic Acids Res. 30 (2002) 276–280.
[21] D. Eisenberg, E. Schwarz, M. Komaromy, R. Wall, Analysis of membrane and surface protein sequences with the hydrophobic moment plot, J. Mol. Biol. 179 (1984) 125–142.
[22] K.A. Datsenko, B.L. Wanner, One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products, Proc. Natl. Acad. Sci. USA 97 (2000) 6640–6645.
[23] D. Yu, H.M. Ellis, E.C. Lee, N.A. Jenkins, N.G. Copeland, D.L. Court, An efficient recombination system for chromosome engineering in *Escherichia coli*, Proc. Natl. Acad. Sci. USA 91 (2000) 5978–5983.
[24] L. Wang, P.R. Reeves, Involvement of the galactosyl-1-phosphate transferase encoded by the *Salmonella enterica rfbP* gene in O antigen subunit processing, J. Bacteriol. 176 (1994) 4348–4356.
[25] N.K. Kochetkov, B.A. Dmitriev, V.L. Lvov, 4-*O*-[(*R*)-1-Carboxyethyl]-D-glucose: a new acidic sugar from *Shigella dysenteriae* type 3 lipopolysaccharide, Carbohydr. Res. 54 (1977) 253–259.
[26] B.A. Dmitriev, V.L. Lvov, N.K. Kochetkov, Complete structure of the repeating unit of the O-specific polysaccharide chain of *Shigella dysenteriae* type 3 lipopolysaccharide, Carbohydr. Res. 56 (1977) 207–209.
[27] K. Bock, C. Pedersen, Carbon-13 nuclear magnetic resonance spectroscopy of monosaccharides, Adv. Carbohydr. Chem. Biochem. 41 (1983) 27–66.
[28] J.A. Bengoechea, E. Pinta, T. Salminen, C. Oertelt, O. Holst, J. Radziejewska-Lebrecht, Z. Piotrowska-Seget, R. Venho, M. Skurnik, Functional characterization of Gne (UDP-*N*-acetylglucosamine-4-epimerase), Wzz (chain length determinant), and Wzy (O-antigen polymerase) of *Yersinia enterocolitica* serotype O:8, J. Bacteriol. 184 (2002) 4277–4287.
[29] D.C. Alexander, M.A. Valvano, 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 (1994) 7079–7084.
[30] A.W. Paton, J.C. Paton, Molecular characterization of the locus encoding biosynthesis of the lipopolysaccharide O antigen of *Escherichia coli* serotype O113, Infect. Immun. 67 (1999) 5930–5937.
[31] M. Skurnik, R. Venho, P. Toivanen, A. Alhendy, A novel locus of *Yersinia enterocolitica* serotype O-3 involved in lipopolysaccharide outer core biosynthesis, Mol. Microbiol. 17 (1995) 575–594.
[32] D. Liu, R. Cole, P.R. Reeves, An O-antigen processing function for Wzx (RfbX): a promising candidate for O-unit flippase, J. Bacteriol. 178 (1996) 2102–2107.
[33] C. Daniels, C. Vindurampulle, R. Morona, Overexpression and topology of the *Shigella flexneri* O-antigen polymerase (Rfc/Wzy), Mol. Microbiol. 28 (1998) 1211–1222.
[34] L. Wang, H. Curd, W. Qu, P.R. Reeves, Sequencing of *Escherichia coli* O111 O antigen gene cluster and identification of O111 specific genes, J. Clin. Microbiol. 36 (1998) 3182–3187.