The O-antigen gene cluster of *Shigella boydii* O11 and functional identification of its *wzy* gene

Jiang Tao, Lu Feng, Hongjie Guo, Yayue Li, Lei Wang *

**TEDA School of Biological Sciences and Biotechnology, Nankai University, 23# Hong Da Street, TEDA, Tianjin 300457, PR China**

**College of Life Sciences, Nankai University, Tianjin 300071, PR China**

**Tianjin State Laboratory of Microbial Functional Genomics, TEDA College, Nankai University, Tianjin 300457, PR China**

Received 24 January 2004; received in revised form 26 February 2004; accepted 8 March 2004

First published online 20 March 2004

**Abstract**

*Shigella* strains are human pathogens and their identification is usually based on their O-antigens. The O-antigen gene cluster of *Shigella boydii* O11 was sequenced. All the expected genes for the synthesis of the O-antigen were identified on the basis of homology and genes for the biosynthesis of dTDP-L-Rhamnose, genes encoding sugar transferases, as well as genes encoding O unit flippase (*wzx*) and O-antigen polymerase (*wzy*). The identity of the putative *wzy* gene was confirmed by showing that a *wzy* deficient mutant strain of *S. boydii* O11 produced a semi-rough LPS phenotype. The predicted *wzx* gene has an opposite transcription direction to that of all of the other genes in the *S. boydii* O11 O-antigen gene cluster. This unusual feature for the *wzx* gene has only previously been reported in *S. boydii* O6. Further comparison revealed an evolutionary relationship between O6 and O11 O-antigen gene clusters. Adjacent-gene PCR showed that *Escherichia coli* O105 and *S. boydii* O11, which share the identical O-antigen, also have the same genes and organization for their respective O-antigen gene clusters. Three genes specific for the *S. boydii* O11 and *E. coli* O105 gene clusters were identified.

© 2004 Federation of European Microbiological Societies. Published by Elsevier B.V. All rights reserved.

**Keywords:** O-Antigen polymerase; Molecular typing; *Shigella boydii* type 11

1. **Introduction**

*Shigella* strains are human pathogens, which cause diseases such as diarrhea and bacillary dysentery. Although four species are recognized including *Shigella boydii*, *Shigella dysenteriae*, *Shigella flexneri* and *Shigella sonnei*, they all can be placed into the same species as *Escherichia coli* based on their sequence similarity [1]. *Shigella* strains are usually identified by their O-antigens, of which 33 forms have been recognized in *Shigella* [2]. Among them, 12 are identical to other known *E. coli* O-antigens and 21 are unique to *Shigella* clones.

The O-antigen, which is part of the lipopolysaccharide (LPS) in the outer membrane of Gram-negative bacteria, consists of many repeats of an oligosaccharide unit (O unit). The O-antigen contributes major antigenic variability to the cell surface. The O unit synthesis is initiated by transfer of a sugar phosphate to a carrier lipid, undecaprenol phosphate (UndP), followed by addition of other sugars sequentially. O units are then “flipped” and polymerized into O-antigen [3]. In *E. coli*, genes for O-antigen synthesis are normally clustered and mapped between two housekeeping genes, *galF* and *gnd*. The O-antigen gene cluster includes genes for synthesis of nucleotide sugar precursors, genes encoding transferases for addition of sugars and genes for O unit processing to form O-antigen. The O-antigen diversity of *E. coli* can be achieved by two means: by obtaining new clusters from other species and by modifying the existing *E. coli* O-antigen gene cluster [2].

Assembly of heteropolymer O-antigen in *Shigella*, *E. coli*, *Salmonella enterica* and other Gram-negative
bacteria normally requires three O unit processing genes, \( wzx \), \( wzy \) and \( wzz \). While the O unit flippase (Wzx) transports the UndPP-linked O unit from the cytoplasmic side of the cytoplasmic membrane to the periplasmic side [4], the O-antigen polymerase (Wzy) links the O units via a glycosidic linkage to form a long chain O-antigen [5]. Wzz is the O-antigen chain length determinant, which controls the chain length of the O-antigen via an interaction with Wzy protein [6]. Mutants with deficient \( wzy \) gene produce LPS consisting of lipidA-core bound to a single O unit [5]. While about 30 \( wzy \) genes have been identified in \( E. coli \) and \( Shigella \) by bioinformatics methods, only those of \( S. flexneri \) type 2A, \( S. dysenteriae \) type 1, \( E. coli \) K-12 (O16), O4, O6, have been confirmed biochemically [5,7–10].

In this study, the O-antigen gene cluster of \( S. boydii \) O11 was sequenced. Genes required for the synthesis of the O-antigen were identified on the basis of homology. A mutant with deficient \( wzy \) gene was examined for the production of LPS. The O-antigen gene cluster of \( E. coli \) O105, which has the same O-antigen form as \( S. boydii \) type 2A, \( S. dysenteriae \) type 1, \( E. coli \) K-12 (O16), O4, O6, have been confirmed biochemically [5,7–10].

2. Materials and methods

2.1. Bacterial strains and plasmids

The bacterial strains, plasmids and primers used in this study are listed in Table 1. \( E. coli \) and \( Shigella \) O serotype type strains were previously described [11].

2.2. Construction of DNaseI shot gun bank

Chromosomal DNA was prepared as previously described [12]. Primers \(^{1523}\) and \(^{1524}\), which were based on the galF and gnd genes, respectively, were used to amplify O-antigen gene cluster of \( S. boydii \) O11 type strain using the Expand Long Template PCR system (Roche). The PCR cycles used were 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 by DNaseI and the resulting fragments were cloned into pGEM-T easy vector to produce a bank using the method described previously [13].

2.3. Sequencing and analysis

Sequencing was carried out using an ABI 3730 automated DNA sequencer. Sequence data were assembled using the Staden Package [14]. The program Artemis [15] was used for gene annotation. The program BlockMaker (http://blocks.fhcrc.org/make_blocks.html) was used for searching conserved motifs. BLAST and PSI-BLAST were used for searching database including GenBank, COG and Pfam (http://www.ncbi.nlm.nih.gov/BLAST/). Potential transmembrane segments were identified using the TMHMM (http://www.cbs.dtu.dk/services/TMHMM/). Sequence alignment and comparison were done using the program ClustalW (http://www.ebi.ac.uk/clustalw/).

2.4. Deletion of \( wzy \) and \( wzx \) genes from \( S. boydii \) O11 type strain

The \( wzy \) and \( wzx \) genes were both replaced by a chloramphenicol acetytransferase (CAT) gene using the RED recombination system of phage lambda [16,17]. The CAT gene was PCR-amplified from plasmid pKK232-8 using primers binding to the 5′ and 3′ ends of the gene, with each primer carrying 40 bp based on the \( S. boydii \) O11 DNA which flanks \( wzy \) (upstream primer WL-884, downstream primer WL-885) and \( wzx \) (upstream primer WL-882, downstream primer WL-883), respectively. The PCR product was transformed into \( S. boydii \) O11 type strain carrying pKD20, and chloramphenicol-resistant transformants were selected after induction of the RED genes according to the protocol described by Datsenko and Wanner [16]. PCR using primers specific to the CAT gene and \( S. boydii \) O11 DNA flanking the \( wzy \) and \( wzx \) genes, respectively, were carried out to confirm the replacement. Membrane preparation, SDS–PAGE and silver staining for visualization of the LPS were carried out as described by Wang and Reeves [18]. To complement the \( wzy \) and \( wzx \) deficient mutants of \( S. boydii \) O11, the \( wzy \) and \( wzx \) genes were PCR-amplified from \( S. boydii \) type strain using primer pairs WL-884 (upstream primer), WL-888 (downstream primer) binding to the ends of \( wzy \) and WL-886 (upstream primer), WL-887 (downstream primer) binding to the ends of \( wzx \), respectively. The resulting PCR products were cloned into pTRC99A to make plasmids pLW1055 and pLW1056 containing \( wzy \) and \( wzx \), respectively.

2.5. Specificity assay by PCR

Chromosomal DNA was isolated as described. The quality of chromosomal DNA from each strain was examined by PCR amplification of the \( mdh \) gene (coding for malate dehydrogenase and present as a housekeeping gene in \( E. coli \) using primers described previously [11]. A total of 28 pools of \( E. coli \) and \( Shigella \) O serotype type strains were made, and each pool contained between 6 and 10 strains [11]. Pools were screened using primers based on specific genes of \( S. boydii \) O11. PCR was car-
ried out in a total volume of 25 μl, of which 10 μl was run on an agarose gel to check for amplified DNA.

2.6. Nucleotide sequence accession number

The DNA sequence of the S. boydii O11 O-antigen gene cluster has been deposited in GenBank under the Accession No. AY529126.

3. Results and discussion

3.1. Sequencing

A sequence of 13,367 bases from galF to gnd was obtained, which contained 11 open reading frames (ORFs) excluding galF and gnd (Fig. 1). All of the ORFs had the same transcriptional direction from galF to gnd except orf11, which transcribes from the opposite direction (Fig. 1).

3.2. O-antigen gene cluster of S. boydii O11

The O unit of S. boydii O11 consists of six sugar residues: a N-acetylglucosamine (GlcNAc), a glucuronic acid (GlcA), a ribofuranose (Ribf) and three rhamnoses (Rha) (Fig. 2). While genes for the synthesis of rhamnose are expected to be present in the O-antigen gene cluster, genes for the synthesis of other sugars are also involved in housekeeping functions and located elsewhere in chromosome [3]. Also expected in the gene cluster are genes encoding O unit flippase, O-antigen polymerase and transferases for UDP-GlcA, UDP-Ribf and three dTDP-Rha. UDP-GlcNAc is used as the first

Table 1

<table>
<thead>
<tr>
<th>Bacterial strains, plasmids or primers</th>
<th>Genotype, comments or oligonucleotide sequences</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G1191 S. boydii O11 type strain</td>
<td></td>
<td>The Institute of Medical and Veterinary Science, Adelaide, Australia</td>
</tr>
<tr>
<td>G1084 E. coli O105 type strain (serotype O105:H8)</td>
<td></td>
<td>The Institute of Medical and Veterinary Science, Adelaide, Australia</td>
</tr>
<tr>
<td>G2001 E. coli K-12 strain DH5α</td>
<td></td>
<td>Beijing Dingguo Biotechnology Development Center, Beijing, PR China</td>
</tr>
<tr>
<td>G1368 Derivative of G1191; Δwzy; Cm′</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>G1369 Derivative of G1191; Δwzx; Cm′</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>H1055 G1368 with plasmid pLW1055</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>H1056 G1369 with plasmid pLW1056</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pGEM-T easy Cloning vector; Amp′</td>
<td></td>
<td>Promega</td>
</tr>
<tr>
<td>pKK232-8 Carrying chloramphenicol resistance (cat) gene</td>
<td></td>
<td>Pharmacia</td>
</tr>
<tr>
<td>pKD20 Encoding the Red recombinase, containing a temperature sensitive replicon</td>
<td></td>
<td>N. Patrick Higgins, Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham</td>
</tr>
<tr>
<td>pTRC99A Cloning vector; Amp′</td>
<td></td>
<td>Pharmacia</td>
</tr>
<tr>
<td>pLW1055 The PCR product of S. boydii O11 wzy gene was cloned into NcoI and BamHI sites of pTRC99A</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>pLW1056 The PCR product of S. boydii O11 wtx gene was cloned into NcoI and BamHI sites of pTRC99A</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td><strong>Primers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#1523 5′-ATT GTG GCT GCA GGG ATC AAA GAA AT-3′</td>
<td></td>
<td>[2]</td>
</tr>
<tr>
<td>#1524 5′-TAG TCG CGT GNG CCT GGA TTA AGT TCG C-3′</td>
<td></td>
<td>[2]</td>
</tr>
<tr>
<td>mdh-5′ 5′-ATG AAA GTC GCA GTC CTC-3′</td>
<td></td>
<td>[25]</td>
</tr>
<tr>
<td>mdh-3′ 5′-GGC GAT ATC TTT CTT CAG CG-3′</td>
<td></td>
<td>[25]</td>
</tr>
<tr>
<td>WL-884 5′-CAG AAT TAT TAA AAT ATA ACA GTA TGA AGG TTT ATT TAT AAT GGA GAA AAA AAT CAC TGG-3′</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>WL-885 5′-TGT TAT CAA AAT CCT AAC TGA TTA AGA TCA ATG GCT TTA AAA AAT ACG CCC CG-5′</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>WL-882 5′-TTC AGG TAT TAA AAA ATC GGA GAA AAA AAT CAC TGG-3′</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>WL-883 5′-CTG ATAT GGG CAA TAG AAG CAG TAA ATA AAT GTT TCC ACC TAA AAA AAT TAC GCC CCG-3′</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>WL-888 5′-TTTACCATGGTTTACGGTTTCTATGG</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>WL-889 5′-TAG GGA TCC TCA TTT ATG TAT CAA AAT CC-3′</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>WL-886 5′-ATGAATTCATGAGCATTTTCAATAATTTAAAATGG</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>WL-887 5′-ATA GGA TCC GTT CTT TGT TGT ATG ATA TAC C-3′</td>
<td></td>
<td>This study</td>
</tr>
</tbody>
</table>
sugar to be added onto UndP, it is transferred by a transferase encoded by \textit{wecA}, which is located outside the O-antigen gene clusters [3]. Functions of each putative gene in \textit{S. boydii} O11 O-antigen gene cluster were predicted on the bases of homology by searching available databases and are summarized in Table 2.

3.2.1. \textit{dTDP-L-rhamnose biosynthesis genes} orfs 1–4 were identified as \textit{rmlB rmlD rmlA rmlC}, respectively, based on their high level identity to other known \textit{rmlB rmlD rmlA rmlC} from other \textit{E. coli} and \textit{Shigella} strains (identity between 89% and 98%). The set of \textit{rml} genes was typical for the synthesis of \textit{dTDP-L-rhamnose} in many O-antigen gene clusters including \textit{S. boydii} O11. \textit{orfs} 1–4 were named accordingly.

3.2.2. \textit{O unit flippase (Wzx) and O-antigen polymerase (Wzy) genes} Wzx and Wzy are both typical membrane proteins [3]. Wzx proteins generally have 12 transmembrane segments. A periplasmic loop of more than 30 amino residues appears to be a characteristic of the Wzy protein [19]. O-antigen gene clusters with deficient \textit{wzy} produce the semi-rough LPS phenotype, in contrast to those with functional Wzy, which produce the smooth phenotype. On this difference, \textit{wzy} genes can be confidently identified [7] based by construction of \textit{Δwzy} mutant and comparison of LPS phenotypes between the mutant and the wild-type.

The only two ORFs encoding predicted membrane proteins are \textit{orf6} and \textit{orf11}. Orf6 has nine predicted transmembrane segments with a large periplasmic loop of 32 amino acid residues, which is a typical topology of Wzy. It also shares 45% and 39% similarity with Wzy of \textit{E. coli} O104 and \textit{S. enterica} strain M40, respectively. When Orf6 and the two forms of Wzy were grouped and analyzed using the BlockMaker program, four conserved motifs (7, 15, 36 and 52 residues, respectively) were revealed. The consensus sequence of these conserved motifs was used to run the program PSI-BLAST to search the Genpept database. Except the two Wzy proteins from strains O104 and M40, other distantly related Wzy proteins were also retrieved (E value = 3 × e⁻⁵⁰) after three iterations. \textit{orf6} was putatively identified as the \textit{wzy} gene. The identity of \textit{orf6} as \textit{wzy} was confirmed by mutation test. The mutant strain G1368, in which \textit{orf6} was replaced by a CAT gene, produced only lipid-A/core part of the LPS and one oligosaccharide unit, while the wild-type \textit{S. boydii} O11 strain produced normal LPS (Fig. 3). The mutation of this semi-rough LPS (SR-LPS) phenotype was trans-complemented by plasmid pLW1055 containing \textit{orf6} to
the smooth phenotype. orf6 was therefore identified as an O-antigen polymerase gene (wzy) and was named accordingly.

Orf11 has 12 predicted transmembrane segments and belongs to the RfbX (Wzx) family (COG2244, E value = 2×e⁻¹¹). Members of this family are membrane proteins involved in the export of O-antigen and teichoic acid. Orf11 also shares 76% and 44% similarity with the Wzx proteins of *S. boydii* O6 and *E. coli* O7, respectively. When Orf11 and the two Wzx proteins were grouped and analyzed using the BlockMaker program, six conserved motifs (16, 27, 39/2, 42 and 55 residues, respectively) were revealed. The consensus sequence of these conserved motifs was used to run the program PSI-BLAST to search the Genpept database. Apart from the two Wzx proteins from *S. boydii* O6 and *E. coli* O7, other distantly related Wzx proteins were also retrieved (E value = 5×e⁻¹⁷) after three iterations. Orf11 was identified as a putative wzx gene. In contrast to most genes of O-antigen gene clusters in *E. coli* and *Shigella*, which are transcribed from galF to gnd, orf11 is transcribed in the direction gnd to galF. This transcription orientation has only been reported in *S. boydii* O6.

To check if this atypical feature of orf11 affects its function, orf11 was replaced with a CAT gene to make the mutant strain G1369. While the wild-type strain produced normal LPS, the mutant produced no O-antigen (Fig. 3), and the mutation was trans-complemented by plasmid pLW1056 containing orf11. This fact indicated that orf11 was functional. orf11 was identified as an O unit flippase gene (wzx) and named accordingly.
3.2.3. Putative transferase genes

Orf9 belongs to the glycosyltransferase 2 family (pfam00535, \(E\) value = \(1 \times 10^{-10}\)) whose members transfer a range of sugar substrates. Orf9 also shares 48% identity with RfbG of \(S.\) flexneri 2a, which is a putative dTDP-rhamnosyltransferase of the O-antigen gene cluster of \(S.\) flexneri type 2a. Based on their high level similarity, Orf9 and RfbG are expected to transfer rhamnose by the identical linkage: L-Rha-\(\alpha\) (1→3)-L-Rha in \(S.\) boydii O11 and \(S.\) flexneri type 2a. Therefore, orf9 is proposed to encode a putative dTDP-rhamnosyltransferase for the formation of L-Rha-\(\alpha\) (1→3)-L-Rha in \(S.\) boydii O11 and \(S.\) flexneri type 2a.

Orf10 belongs to a family of diverse phosphoribosyltransferases (pfam00156, \(E\) value = \(5 \times 10^{-4}\)). Orf10 also shares 43% identity with GptA-1 of \(P.\) furiosus DSM3638, which was suggested to be a putative phosphoribosyltransferase. Thus, we proposed orf10 is a putative ribosyltransferase for transferring Ribf in \(S.\) boydii O11 and named it \(wbsZ\).

Orf7 belongs to the glycosyltransferase 2 family (pfam00535, \(E\) value = \(2 \times 10^{-4}\)) and shares 26% identity with WbdE of \(S.\) boydii O4, which was suggested to be a putative glycosyltransferase. Orf8 shares 38% identity with a putative glycosyltransferase of \(B.\) cereus ATCC 14579. Orf5 shares 31% identity with a putative glycosyltransferase of exopolysaccharide operon in \(S.\) thermophilus. orf7, orf8 and orf5 are proposed to be putative transferase genes and named \(wbsW, wbsX, wbsV\) respectively.

The O unit of \(S.\) boydii O11 also contains four O-acetyl groups branched out from the main chain. Genes encoding O-acetyl transferases may or may not locate in O-antigen gene clusters [3]. All of the genes in the O-antigen gene cluster of \(S.\) boydii O11 have been assigned functions and none of them share any similarity with acetyltransferase genes. This suggests that acetyltransferase genes required are located outside the O-antigen gene cluster in \(S.\) boydii O11.

3.3. The \(E.\) coli O105 and \(S.\) boydii O11 O-antigen gene clusters have the same genes and organization

It is known that \(E.\) coli O105 and \(S.\) boydii O11 have the same O-antigen form [20]. PCR was carried out for \(E.\) coli O105 and \(S.\) boydii O11 type strains using primers based on the sequence of \(S.\) boydii O11 gene clusters for each adjacent gene (including the flanking galF and gnd genes). The respective PCR products from each primer pair were same in size in \(S.\) boydii O11 and \(E.\) coli O105 strains. Sequencing was also performed for selected PCR products from \(E.\) coli O105, and all have the same sequence as their corresponding parts in \(S.\) boydii O11 (data not shown). These results indicate that the O-antigen gene cluster of \(E.\) coli O105 has the same genes and organization as of \(S.\) boydii O11. The putative \(wzx\) gene of \(E.\) coli O105 also has an opposite orientation. The same genes and organization in the O-antigen gene clusters are also present in other strains which share the same O-antigen forms including \(S.\) boydii O4 and \(E.\) coli O53 [2], \(S.\) boydii O5 and \(E.\) coli O79 [2], and \(S.\) dysenteriae O7 and \(E.\) coli O121 [21]. All these studies indicate that the O-antigen forms in those strains are solely dependent on their O-antigen gene clusters. The presence of identical O-antigen gene clusters in \(S.\) boydii O11 and \(E.\) coli O105 could have been formed in three possible processes. The first one is they were derived from a common ancestral gene cluster in the ancestor of \(S.\) boydii O11 and \(E.\) coli O105. This suggestion was supported by the data in \(E.\) coli and \(S.\) enterica. The second one is they were separately transferred from a common source into \(S.\) flexneri and \(E.\) coli after the \(S.\) flexneri clones arose in \(E.\) coli. The third one is they were recently transferred from one to the other. In other cases where the O-antigen is the same between a \(S.\) flexneri clone and \(E.\) coli, the two forms are nearly identical and show features typical of the \(E.\) coli O-antigen gene cluster [2,21]. However, the O-antigen gene clusters of \(S.\) boydii O11 and \(E.\) coli O105 show one atypical feature, reverse orientation of \(wzx\), suggesting that it may have been recently assembled in one of the two strains then spread to the other.

3.4. Identification of genes specific for \(S.\) boydii O11 and \(E.\) coli O105

Sugar transferase genes are commonly specific to individual O-antigens [3]. Primer pairs were designed based on predicted transferase genes \(wbsW, wbsY\) and \(wbsZ\) (Table 3), and used to screen the DNA pools containing representatives of the 186 known O-antigen forms of \(E.\) coli and \(S.\) flexneri strains. Except the pools containing \(S.\) boydii O11 and \(E.\) coli O105, which gave the expected PCR products, no PCR products were detected with other pools. Therefore, all of the three genes are specific for \(S.\) boydii O11 and \(E.\) coli O105 strains. While \(S.\) boydii O11 is human pathogen, \(E.\) coli O105 is also a pathogenic verotoxin-producing \(E.\) coli (VTEC) strain, which has been isolated from livestock, food and clinical specimens in recent years [22,23]. The genes specific to those two strains may be used for their rapid identification in PCR-based methods.

3.5. Comparison of \(S.\) boydii O11 and \(S.\) boydii O6 gene clusters

\(E.\) coli O-antigen gene clusters usually have their genes transcribed in the same direction from \(galF\) to \(gnd\). However, in the case for \(S.\) boydii O11 O-antigen gene cluster, \(wzx\) is transcribed from the opposite direction. Coincidently, the only reported gene of O-antigen gene clusters that has an opposite transcription direction is
orf* of *S. boydii* O6 [2]. When the O-antigen gene clusters of *S. boydii* O11 and O6 were compared, an evolutionary relationship was revealed between the two clusters. We found the region containing two genes upstream of *gnd* is very similar between the two clusters (Fig. 1). Apart from having an opposite transcription direction, both forms of the orf* gene are located at the 3′ end of their respective gene clusters and they also share identical size and 67% DNA identity. The gene upstream of *wzx* is a transferase gene designated as *wbsZ* in *S. boydii* O11, and a non-functional orf* (named orf*) interrupted by an IS in *S. boydii* O6. The two parts of orf* separated by the IS in *S. boydii* O6 share 67% and 67% identity to corresponding parts of *wbsZ* in *S. boydii* O11. The whole orf* excluding the IS shares identical size and 67% identity to *wbsZ*. This suggests that orf* and *wbsZ* are closely related. The fact that neither half of orf* has any indels or stop codons indicates that the IS was inserted recently [2]. Interruption of the orf* indicates that the orf* is not required for the biosynthesis of *S. boydii* O6 O-antigen. We also found that 209 bases of the non-encoding region downstream *galF* in *S. boydii* O11 share 98% identity to the corresponding part in *S. boydii* O6, while the rest of the non-encoding region is significantly different between the two clusters. This further suggests that *S. boydii* O11 and *S. boydii* O6 are related.

*Shigella* strains were thought to have arisen from *E. coli* [1]. For those *Shigella* strains whose O-antigen form are also found in *E. coli*, their O-antigen gene clusters may come from the same ancestors as their respective *E. coli* counterparts. *Shigella* strains with unique O-antigen forms are likely to be evolved more recently by modifying the existing gene clusters [2]. While the O-antigen form of *S. boydii* O11 is identical to that of *E. coli* O105, the O-antigen form of *S. boydii* O6 is unique. Therefore, the O-antigen forms of *S. boydii* O11 and *E. coli* O105 is considered to be older than that of *S. boydii* O6. In addition, presence of the IS also indicates that the O-antigen gene cluster of *S. boydii* O6 was assembled more recently. We propose that the O-antigen gene cluster of *S. boydii* O6 was derived from *S. boydii* O11. The mechanism might involve the replacement of the region downstream of the 209 bp non-encoding sequence at the *galF* end and upstream of *wbsZ* at the *gnd* end of *S. boydii* O11 O-antigen gene cluster with the region now present in *S. boydii* O6. orf*, which appeared to be evolved from *wbsZ*, a predicted transferase gene for Ribf in *S. boydii* O11, is not required for the synthesis of the *S. boydii* O6 O-antigen and it was later interrupted by insertion of an IS. The fact that sequence identity at the amino acid level (57%) is lower than that of the DNA (67%) between the orf* gene forms of *S. boydii* O6 and *S. boydii* O11, together with a relatively high value for *Ka/Ks* (0.18), suggest an occurrence of adaptive changes after divergence of orf* in *S. boydii* O6 and *S. boydii* O11.

### 4. Conclusions

The sequence of the O-antigen gene cluster of the *S. boydii* O11 was determined. The O-antigen gene cluster was shown to contain 11 putative genes, all of which were assigned functions on the basis of homology. The O-antigen gene cluster of *S. boydii* O11 contains four *rml* genes, five putative transferase genes, and O unit processing genes *wzy* and *wzx*. These genes account for all functions expected in the synthesis of *S. boydii* O11 O-antigen. The *wzy* gene is functionally identified. *E. coli* O105 and *S. boydii* O11, which share the same O-antigen form, also have the same genes and organization in their O-antigen gene clusters. The genes specific for *S. boydii* O11 and *E. coli* O105 were identified, which can be potentially used for the fast identification of the two strains. An atypical feature of *S. boydii* O11 O-antigen gene cluster is the opposite transcription direction of the orf* gene. Further comparison between *S. boydii* O11 and *S. boydii* O6 O-antigen gene clusters revealed an evolutionary relationship.

### Acknowledgements

We thank all the individuals and institutes for kindly supplying strains and plasmids listed in Table 1. This...
work was supported by the 863 Program (2002AA2Z2051), the Chinese National Science Fund for Distinguished Young Scholars (30125001), NSFC General Program (30270029) and funding from the Science and Technology Committee of Tianjin City (013181711) to L.W.

References


