ELSEVIER

FEMS Microbiology Letters 248 (2005) 153–161

FEMS MICROBIOLOGY Letters

www.fems-microbiology.org

Characterization of the *Escherichia coli* O59 and O155 O-antigen gene clusters: The atypical *wzx* genes are evolutionary related

Hongjie Guo [a,b,1,2], Qingke Kong [a,1], Jiansong Cheng [a], Lei Wang [a,b], Lu Feng [a,b,*]

[a] *TEDA School of Biological Sciences and Biotechnology, Nankai University, Tianjin 300457, PR China*
[b] *Tianjin Research Center for Functional Genomics and Biochips, TEDA College, Nankai University, Tianjin 300457, PR China*

Received 25 February 2005; received in revised form 29 April 2005; accepted 17 May 2005

First published online 16 June 2005

Edited by R.Y.C. Lo

Abstract

O-antigens are highly polymorphic. The genes specifically involved in O-antigen synthesis are generally grouped together on the chromosome as a gene cluster. In *Escherichia coli*, the O-antigen gene clusters are characteristically located between the housekeeping genes *galF* and *gnd*. In this study, the O-antigen gene clusters of *E. coli* O59 and *E. coli* O155 were sequenced. The former was found to contain genes for GDP-mannose synthesis, glycosyltransferase genes and the O-antigen polymerase gene (*wzy*), while the latter contained only glycosyltransferase genes and *wzy*. O unit flippase genes (*wzx*) were found immediately downstream of the *gnd* gene, in the region between the *gnd* and *hisI* genes in these two strains. This atypical location of *wzx* has not been reported before, and furthermore these two genes complemented in trans despite the fact that different O-antigen structures are present in *E. coli* O59 and O155. A putative acetyltransferase gene was found downstream of *wzx* in both strains. Comparison of the region between *gnd* and *hisI* revealed that the *wzx* and acetyltransferase genes are closely related between *E. coli* O59 and O155, indicating that the two gene clusters arose recently from a common ancestor. This work provides further evidence for the O-antigen gene cluster having formed gradually, and selection pressure will eventually bring O-antigen genes into a single cluster. Genes specific for *E. coli* O59 and O155, respectively, were also identified.

Keywords: O-antigen flippase; Specific genes; *Escherichia coli* serotypes O59 and O155

1. Introduction

The O-antigen, which contains many repeats of an oligosaccharide unit (O unit), is part of the lipopolysaccharide (LPS) present on the surface of gram-negative bacteria. It is one of the most variable components of cells and it varies by the types of sugars present, the arrangement of the sugars and the linkages between the sugars within an O unit, as well as the linkage between the O units [1,4]. Based on the O-antigen variation, a bacterial species is divided into different O serotypes. In *Escherichia coli* (including *Shigella*), 186 O serotypes have been recognized [2]. The O-antigen is subject to intense selection by the host immune system, which may be the major factor for the maintenance of many different O-antigen forms [1,4].

There are three main classes of genes involved in O-antigen synthesis: (1) genes for synthesis of nucleotide

* Corresponding author. Tel.: +86 22 66229592; fax: +86 22 66229596.
E-mail address: fenglu63@nankai.edu.cn (L. Feng).
[1] These authors contributed equally to this work.
[2] Present address: Department of Biochemistry, Ohio State University, 484 W 12th Avenue, Columbus, OH 43210, USA.

0378-1097/$22.00
doi:10.1016/j.femsle.2005.05.036

sugar precursors; (2) genes encoding glycosyltransferases for sequential transfer of sugars from their respective nucleotide precursors to the carrier lipid, undecaprenol-phosphate (Und-P) to form O units; and (3) O unit processing genes encoding O unit flippase (*wzx*) and O-antigen polymerase (*wzy*) [1]. The role of *wzx* is to translocate or flip the O units formed at the cytoplasmic face of the membrane to the periplasmic face. The O units are then polymerised to form a long chain O-antigen at the periplasmic face of the membrane by Wzy. Both *wzx* and *wzy* are specific to individual O-antigens [3].

The genes specifically involved in O-antigen synthesis are generally grouped together on the chromosome as a gene cluster. In *E. coli* and *Salmonella enterica*, the O-antigen gene clusters lie between the *galF* and *gnd* genes [4]. The O-antigen genes are found generally very close to each other, often overlapping in their reading frames, and thought to be transcribed as a unit [5]. There are few cases of one or more O-antigen genes being located outside the O-antigen gene clusters. For example, in *E. coli* O55, two of the colitose synthetic pathway genes, *col1* and *col2*, were found outside the O-antigen gene cluster, downstream of the *gnd* gene, while the rest of the pathway genes were found within the O-antigen gene cluster between *galF* and *gnd* [6]. The location of *col1* and *col2* is perhaps at intermediate stage, and selection pressure will relocate the two genes into one operon with other O-antigen genes.

The structural variation of O-antigens is almost entirely dependent on the variation of the O-antigen gene clusters. Typically, O-antigen gene clusters have a GC content lower than the average level of the genome, indicating the gene clusters have been arisen from other species by lateral transfer [7,8]. Inter- and intraspecies lateral transfer of O-antigen genes appear to play important roles for expanding O-antigen polymorphism. Both transfer of O-antigen genes for assembly of new O-antigen gene clusters and transfer of entire O-antigen gene clusters between clones of a species by homologous recombination were observed [7,9,10]. Interspecies transfer of the entire O-antigen gene cluster from *Plesiomonas shigelloides* to *E. coli* was also reported [11,12].

In this study, the region between *galF* and *his* operon of *E. coli* O59 and O155 chromosome were sequenced, respectively, *wzy* genes were identified functionally and other genes were identified on the basis of homology. It was shown that the O59 and O155 antigen gene clusters are located between *galF* and *gnd*. However, the O unit flippase gene (*wzx*) and an acetyltransferase gene were found outside the main O-antigen gene clusters, locating between the *gnd* gene and the *his* operon, in both strains. Both *wzx* genes were proven to be functional and able to trans-complement with each other, despite the fact that this gene is usually highly specific to a particular O unit. This indicates that the *E. coli* O59 and O155 gene clusters were recently generated from a common ancestor, and they are at a stage to have the *wzx* genes adapted specifically to the respective O units and the *wzx* and acetyltansferase gene moved to the main cluster. Finally, by screening against all 186 *Shigella* and *E. coli* O serotypes, genes specific for each of *E. coli* O59 and O155 were identified.

2. Materials and methods

2.1. Bacterial strains and plasmids

The bacterial strains and plasmids used in this study are listed in Table 1. Primers used in this study are listed in Table 2.

2.2. Construction of DNaseI shot gun bank

Chromosomal DNAs were prepared as previously described [13]. Long PCR was carried out using the Expand Long Template PCR system from Roche. Primers #1523 (5′-ATTGTGGCTGCAGGGATCAAAGAAAT-3′) and #1524 (5′-TAGTCGCGTGNGCCTGGATTAAGTTCGC-3′) [14] were used to amplify the region between *galF* and *gnd*, and the PCR cycles 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. Primers WL-359 and WL-360 were used to amplify the region between the *gnd* gene and the *his* operon, and the PCR cycles were as follows: denaturation at 94 °C for 10 s, annealing at 58 °C for 30 s and extension at 68 °C for 10 min. To limit any possible PCR errors, 10 individual PCR products were pooled. The PCR products were subjected to DNaseI digestion and cloned into pGEM-T Easy to construct banks using the method described previously [15].

2.3. Sequencing and analysis

Sequencing was carried out using an ABI 3730 automated DNA sequencer. Sequence data were assembled using the Staden Package [16]. The program Artemis was used for gene annotation [17]. The program BLOCKMAKER was used for searching conserved motifs [18]. BLAST and PSI-BLAST [19] were used for searching databases including the GenBank database and the Pfam protein motif database [20]. Sequence alignment and comparison were performed using the program Clustal W [21].

2.4. Deletion of wzx, wzy and wclD genes from E. coli O59 and O155

The *wzx*, *wzy* and *wclD* genes were each replaced by the chloramphenicol acetyltransferase (CAT) gene using the RED recombination system of phage lambda [22,23]. The CAT gene was PCR-amplified from plasmid

Table 1
Bacterial strains and plasmids used in this work

Bacterial strains and plasmids	Genotype or comments	Reference or source
Strains		
G1070	*E. coli* O59 type stain (serotype O59:H19)	The Institute of Medical and Veterinary Science, Adelaide, Australia
G1106	*E. coli* O155 type stain (serotype O155:H9)	The Institute of Medical and Veterinary Science, Adelaide, Australia
G2001	*E. coli* K-12 strain DH5α	Beijing Dingguo Biotechnology Development Center, Beijing, PR China
H1411	Derivative of G1106; Δ*wzy*; Cm^r	This study
H1413	Derivative of G1106; Δ*wzx*; Cm^r	This study
H1412	Derivative of G1070; Δ*wzy*; Cm^r	This study
H1414	Derivative of G1070; Δ*wzx*; Cm^r	This study
H1056	Derivative of G1070; Δ*wclD*; Cm^r	This study
H1011	H1414 with plasmid pLW-1011	This study
H1012	H1414 with plasmid pLW-1012	This study
H1013	H1413 with plasmid pLW-1011	This study
H1014	H1413 with plasmid pLW-1012	This study
H1123	H1411 with plasmid pLW-1213	This study
H1124	H1056 with plasmid pLW-1214	This study
H1125	H1412 with plasmid pLW-1211	This study
H1126	H1056 with plasmid pLW-1212	This study
Plasmids		
pGEM-T easy	Cloning vector; Amp^r	Promega
pKK232-8	Carring chloramphenicol resisitance(cat) gene	Pharmacia
pKD20	Encoding the Red recombinase, containing a temperature sensitive replicon	N. Patrick Higgins, Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham
pTRC99A	Cloning vector; Amp^r	Pharmacia
pLW-1011	The PCR product of *E. coli* O59 *wzx* gene was cloned into *Nco*I and *Bam*HI sites of pTRC99A	This study
pLW-1012	The PCR product of *E. coli* O155 *wzx* gene was cloned into *Nco*I and *Bam*HI sites of pTRC99A	This study
pLW-1211	The PCR product of *E. coli* O59 *wzy* gene was cloned into *Nco*I and *Bam*HI sites of pTRC99A	This study
pLW-1212	The PCR product of *E. coli* O59 *wclD* gene was cloned into *Nco*I and *Eco*RI sites of pTRC99A	This study
pLW-1213	The PCR product of *E. coli* O155 *wzy* gene was cloned into *Nco*I and *Bam*HI sites of pTRC99A	This study
pLW-1214	The PCR product of *E. coli* O155 *wclD* gene was cloned into *Nco*I and *Bam*HI sites of pTRC99A	This study

pKK232-8 using primers binding to the 5′ and 3′ ends of the gene, respectively, and each primer also carrying about 40bp which flanks the target gene. Primer pairs WL −995/996 and WL −991/992 were used for replacement of *wzx*; WL −997/998 and WL −993/994 for *wzy* in *E.Coli* O59 and 0155, respectively; WL −983/984 for *wclD* in *E. coli* O59. The PCR products were transformed into *E. coli* O59 and O155, respectively, 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. Membrane preparation, SDS–PAGE and silver staining for visualization of the LPS were carried out as described previously [24].

To complement the deficient mutants, *wzx*, *wzy* and *wclD* genes from *E. coli* O59 and O155 were PCR amplified and cloned into the *Nco*I and *Bam*HI or *Nco*I and *Eco*RI sites of pTRC99A to make plasmids pLW1011, pLW1012, pLW1211, pLW1212, pLW1213, and pLW1214 (see Table 1). Expression of the cloned genes was induced by 2.5 mM IPTG (isopropylthiogalactopyranoside).

2.5. Specificity assay by PCR

Chromosomal DNA was isolated from 186 *E. coli* (including *Shigella*) strains representing all the different O serotypes. 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 house-keeping gene in *E. coli*) using primers described previously [25]. A total of 28 DNA pools were made, and each pool contained between 6 and 10 strains [2]. Pools were screened using primers based on genes

Table 2
Primers used in this work[a]

Primers	Oligonucleotide sequences	Description	Reference
WL-359 (F)	5′-CCTGGTTGATGTGAT(TC)CTGGA(TC)GAAGC(GAT)GC(TAG)AA-3′	From *gnd*, for amplification between *gnd* and *hisI*	This study
WL-360(R)	5′-CTGGCGAATCC(CT)ATCGG(CAG)CC(AG)ACCTGCCACAAA-3′	From *hisI*, for amplification between *gnd* and *hisI*	This study
WL-991(F)	5′-TTATTATTTCTTAGTTCTAAGAGTCTACGTGAGATATTAGATGGAGAAAAAAATCACTGG-3′	*wzx* (O155)flanking DNA, for *wzx* (O155) replacement	This study
WL-992(R)	5′-ACCCCCTTCTAAATAGCATGATAAATAATTGTCTAAATCCTAAAAAAATTACGCCCCG-3′	*wzx* (O155)flanking DNA, for *wzx* (O155) replacement	This study
WL-993(F)	5′-ATAAAGCAGCGTTTTCTGTAAAAGGAAGAGAGTATTGAAAATGGAGAAAAAAATCACTGG-3′	*wzy* (O155)flanking DNA, for *wzy* (O155) replacement	This study
WL-994(R)	5′-AAAAATCTGCCGTTACGATATAATAAATACATCTCGCTAAAAAAATTACGCCCCG-3′	*wzy* (O155)flanking DNA, for *wzy* (O155) replacement	This study
WL-995(F)	5′-TTTGAATATCTTCTAAGGTCAAAATGTAATGAGATAATAGATGGAGAAAAAAATCACTGG-3′	*wzx* (O59)flanking DNA, for *wzx* (O59) replacement	This study
WL-996(R)	5′-CCATTTTTTTTTGTGTTTTTACTGATCATTTTGGCACTCATAAAAAAATTACGCCCCG-3′	*wzx* (O59)flanking DNA, for *wzx* (O59) replacement	This study
WL-997(F)	5′-AATAAATCTTTGCACCTTTATAATTATTACATCGTCATCGAGGAGCTAAGGAAGCTAAAATGG-3′	*wzy* (O59)flanking DNA, for *wzy* (O59) replacement	This study
WL-998(R)	5′-ATTTTACAAAATATTAAAACAAGACAAATACCGTATAGTGTAAAAAAATTACGCCCCG-3′	*wzy* (O59)flanking DNA, for *wzy* (O59) replacement	This study
WL-983(F)	5′-GGAATATTAAGAGACTTTTCTTTCATTTAAATGTAAGATGGGAGCTAAGGAAGCTAAAATG-3′	*wclD* (O59)flanking DNA, for *wclD* (O59) replacement	This study
WL-984(R)	5′-TCATCGGTGAGCCAACAGCAATCTTATCCTCAGGAACGTCTAAAAAAATTACGCCCCG-3′	*wclD* (O59)flanking DNA, for *wclD* (O59) replacement	This study
WL-971(F)	5′-CGAGATCTCCATGGCTAACTATATCAAAAATG-3′	From *wzx* (O155), for amplification of *wzx* (O155)	This study
WL-972(R)	5′-GAGATGGATCCCCTACCCCCTTCTAAATA-3′		
WL-973(F)	5′-TAATCCATGGCAAATTATATTAAAAATGC-3′	From *wzx* (O59), for amplification of *wzx* (O59)	This study
WL-974(R)	5′-TTTGGATCCTTTACTGATCATTTTGGCAC-3′		
WL-3947(F)	5′-TAATCCATGGAGAAAAGTGAATTTCAAG-3′	From *wzy* (O59), for amplification of *wzy* (O59)	This study
WL-3948(R)	5′-TTCGGATCCAATACCTATTTTCATTATG-3′		
WL-3949(F)	5′-TCATCCATGGTAAGATGGCGTGTGTTTCT-3′	From *wclD* (O59), for amplification of *wclD* (O59)	This study
WL-3950(R)	5′-GCCGAATTCTATCAATAGAGAACATTAGC-3′		
WL-3951(F)	5′-TCATCCATGGGAGGCTTTATTACAAG-3′	From *wzy* (O155), for amplification of *wzy* (O155)	This study
WL-3952(R)	5′-TTCGGATCCTTCTCCTTATAAGAAATC-3′		
WL-3953(F)	5′-TGATCCATGGTTGACAGTTATGAAAG-3′	From *wclD* (O155), for amplification of *wclD* (O155)	This study
WL-3954(R)	5′-TTCGGATCCACTGAGATAATTAGTAG-3′		

[a] Primers with (F) are upstream primers, and (R) are downstream primers.

specific to *E. coli* O59 or O155, respectively. PCR was carried 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 region between *galF* and *hisI* from *E. coli* O59 and O155 O-antigen gene cluster have been deposited in GenBank under the accession numbers AY654590 and AY657020, respectively.

3. Results and discussion

3.1. Sequencing the regions between galF and hisI from E. coli O59 and O155

Sequences of 16,573 and 12,755 bases between *galF* and *hisI* were obtained from *E. coli* O59 and O155, respectively. The former contained fourteen open reading frames (ORFs) including *galF* and *hisI*, and the latter contained nine. All of the ORFs in the two sequences are transcribed from *galF* to *hisI* with the exception of *gla*, which transcribed in the opposite direction (Fig. 1). ORFs were assigned functions based on knowledge of previously characterized O-antigen genes and homology comparisons using available databases (Tables 3 and 4).

3.2. The region between galF and gnd is the O-antigen gene cluster

The region between *galF* and *gnd* from *E. coli* O59 was found to contain two genes for the synthesis of GDP-mannose (*manC* and *manB*), three glycosyltransferase genes (*wclA*, *wclB* and *wclC*), and the O-antigen polymerase gene (*wzy*). The same region from *E. coli*

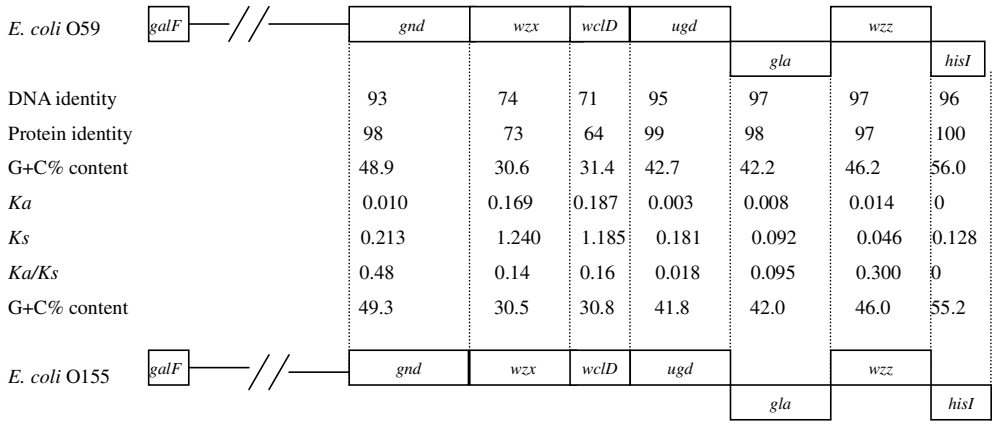

Fig. 1. Comparison of sequences between *gnd* and *hisI* of *E. coli* O59 and O155. Boxes below the line indicated transcription direction is reversed.

O155 was found to contain three glycosyltransferase genes (*wclE*, *wclF*, and *wclG*) and *wzy*. All of the genes are typical of those involved in the synthesis of O-antigen, and the region between *galF* and *gnd* is the normal location for *E. coli* O-antigen gene cluster.

The putative function of *wzy* was further confirmed by comparison of LPS phenotype between the wild type and the mutant in which *wzy* was replaced by a CAT gene. While H1412, the *wzy* deficient mutant strain of *E. coli* O59, produced only lipid-A/core part of the LPS and one oligosaccharide unit, the wild type *E. coli* O59 and complement strain (H1125) produced normal LPS (Fig. 2(b)). The same was found for the mutant and wild type *E. coli* O155 (Fig. 2(a)). This confirms the designation of *wzy*, and the regions between *galF* and *gnd* in *E. coli* O59 and O155 are O-antigen gene clusters. However, the O unit flippase gene is absent in this region in both strains.

The structures of *E. coli* O59 and O155 have not been elucidated, and the sequencing data suggest that they are substantively different. Both *manB* and *manC* are exclusively used for GDP-D-mannose synthesis

Table 3
Genes in the region between *galF* and *hisI* of *E. coli* O59

Orf No.	Gene name	No. of a.a.	%G + C content	Similar proteins (accession number)	%Identitical/%Similar (residues)	Putative function
orf1	*galF*	254	50.0	UTP-glucose-1-phosphate uridylyltransferase, *Escherichia coli* (AB008676)	100/100(254)	UTP-glucose-1-phosphate uridylyltransferase
orf2	*wclA*	403	32.1	Glycosyltransferase, *Agrobacterium tumefaciens str.* C58 (AAL44363)	26/50(411)	Glycosyltransferase
orf3	*wzy*	389	28.6	O-antigen polymerase, *Escherichia coli* O6(AJ426423)	27/45(447)	O-antigen polymerase
orf4	*wclB*	358	34.3	Mannosyltransferase, *Brucella melitensis* (AAC98618)	38/58(376)	Mannosyltransferase
orf5	*wclC*	347	31.2	Glycosyltransferase, *Escherichia coli* O6 (CAD19797)	42/60(401)	Glycosyltransferase
orf6	*manC*	465	37.1	Mannose-1-phosphate guanyltransferase, *Escherichia coli* O157:H7 (BAB36277)	66/80(478)	Mannose-1-phosphate guanyltransferase
orf7	*manB*	476	37.2	Phosphomannomutase, *Shigella boydii* O5 (AAL27334)	92/97(471)	Phosphomannomutase
orf8	*gnd*	468	48.9	6-Phosphogluconate dehydrogenase, *Escherichia coli* K12 (AAA24496)	98/98(468)	6-Phosphogluconate dehydrogenase
orf9	*wzx*	427	30.6	Wzx, *Yersinia enterocolitica* type O8 (AAC60766)	25/52(429)	O-antigen filippase
orf10	*wclD*	152	31.4	Serine acetyltransferase, *Clostridium tetani* E88 (AAO34989)	41/65(186)	Acetyltransferase
orf11	*ugd*	388	42.7	UDP-Glc-6-dehydrogenase Ugd, *Escherichia coli* O113 (AAD50493)	98/98(388)	UDP-Glc-6-dehydrogenase
orf12	*gla*	334	42.2	Gla, *Escherichia coli* O113 (AAD50494)	99/99(334)	GlcA epimerase
orf13	*wzz*	344	46.2	O-antigen chain length determinant Wzz, *Escherichia coli* O55 (AAL67565)	92/96(345)	O-antigen chain length determinant
orf14	*hisI*	115	56.0	Phosphoribosyl-AMP cyclohydrolase *Escherichia coli* O157 (BAA77739)	99/100(203)	Phosphoribosyl-AMP cyclohydrolase

Table 4
Genes in the region between *galF* and *hisI* of *E. coli* O155

Orf number	Gene name	No. of a.a.	%G + C content	Similar proteins (accession number)	%Identitical/%Similar (residues)	Putative function of protein
orf1	*galF*	254	50.0	Glucose-1-phosphate uridylyltransferase, *Escherichia coli* (AAD50483)	100/100(254)	Glucose-1-phosphate uridylyltransferase
orf2	*wclE*	339	30.0	Glycosyltranserase, *Erwinia amylovora* (CAA54883)	28/52(266)	Glycosyltranserase
orf3	*wzy*	318	27.7	O-antigen polymerase, *Shigella boydii* O6(AAL27339)	24/45(305)	O-antigen polymerase
orf4	*wclF*	350	26.6	Glucosyltransferase, *Methanosarcina mazei* Goe1 (AAM30345)	24/52(306)	Glycosyltransferase
orf5	*wclG*	272	32.1	WbwC, *Escherichia coli* O104 (AAK64375)	58/76(268)	Glycosyltransferase
orf6	*gnd*	469	49.3	6-Phosphogluconate dehydrogenase, *Escherichia coli* K12(AAA24496)	98/99(468)	6-Phosphogluconate dehydrogenase
orf7	*wzx*	429	30.5	Wzx, *Yersinia enterocolitica* (type 0:8)(AAC60766)	28/54(410)	O-antigen flippase
orf8	*wclD*	156	30.8	Serine acetyltransferase, *Clostridium tetani* E88 (AAO34989)	41/65(96)	Acetyltransferase
orf9	*ugd*	389	41.8	UDP-Glc-6-dehydrogenase, Ugd *Escherichia coli* O113(AAD50493)	98/98(388)	Dehydrogenase
orf10	*gla*	335	42.0	Gla *Escherichia coli* O113(AAD50494)	99/99(334)	Nucleotide sugar epimerase
orf11	*wzz*	345	46.6	Chain length determinant protein, *Escherichia coli* CFT073 (AAN81009)	92/96(326)	Chain length determinant protein
orf12	*hisI*	116	55.2	Phosphoribosyl-AMP cyclohydrolase *Escherichia coli* O157(BAA77739)	100/100(115)	Cyclohydrolase

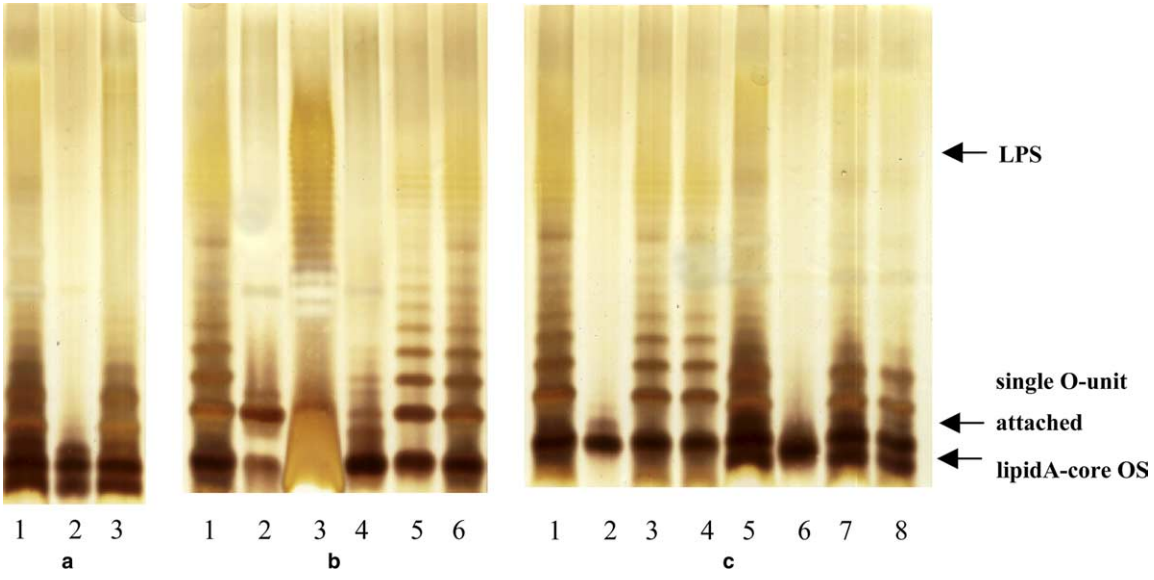

Fig. 2. Deletion and complementation analysis of *E. coli* O59 *and* O155 *wzy*, *wclD* and *wzx* genes. Membrane extracts were run on SDS–PAGE gels and stained by silver staining. a: 1, G1106 (*E. coli* O155 type strain); 2, H1411 (G1106 missing the *wzy* gene); 3, H1123 (H1411 with plasmid pLW-1213 containing *wzy* from 0155); b: 1, G1070 (*E. coli* O59 type strain); 2, H1412 (G1070 missing the *wzy* gene); 3, H1125 (H1412 with plasmid pLW-1211 containing *wzy* from O59); 4, H1056 (G1070 missing the *wclD* gene); 5, H1126 (H1056 with plasmid pLW-1212 containing *wclD* from O59); 6, H1124 (H1056 with plasmid pl,W-1214 containing *wclD* from O155); c: 1, G1070 (*E. coli* O59 type strain); 2, H1414 (G1070 missing the *wzx* gene); 3, H1011 (H1414 with plasmid pLW1011 containing *wzx* from O59); 4, H1012 (H1414 with plasmid pLW1012 containing *wzx* from O155); 5, G1106 (*E. coli* O155 type strain); 6, H1413 (G1106 missing the *wzx* gene); 7, H1013 (H1413 with plasmid pLW1011 containing *wzx* from O59); 8, H1014 (H1413 with plasmid pLW1012 containing *wzx* from O155).

[26], and the presence of *manB* and *manC* in the *E. coli* O59 gene cluster suggests that the O59 antigen contains mannose. The absence of any sugar pathway genes in the O155 gene cluster suggests that this antigen contains common sugars, such as Glc, Gal, GlcNAc or GalA, which are also involved in other cell functions and are synthesized by house-keeping genes [1,5].

3.3. wzx and acetyltransferase genes are located downstream of the main gene cluster in E. coli O59 and O155

The regions between *gnd* and *hisI* in both *E. coli* O59 and O155 were found to contain five ORFs. *orf11*, *orf12* and *orf13* of O59 and *orf9*, *orf10* and *orf11* of O155 showed high level similarity to *ugd*, *gla* and *wzz*, respectively (Tables 3 and 4). *ugd* and *gla* are house-keeping genes, encoding proteins involved in UDP-GlcA and UGP-GalA biosynthesis, respectively [5]. *wzz* encodes the O-antigen chain length determinant protein. *ugd*, *gla* and *wzz* are usually located in this region in other *E. coli* serotypes [27]. *orf11*, *orf12* and *orf13* of O59 and *orf9*, *orf10* and *orf11* of O155 were named *ugd*, *gla*, and *wzz*, respectively.

orf9 of *E. coli* O59 was predicted to encode protein with 12 transmembrane segments, relating to the protein family Pfam062047 and COG2244, which both include Wzx proteins. When Orf9 and the putative Wzx proteins of *Nostoc punctiforme* (GenBank entry ZP_00108476) and *Yersinia enterocolitica* (GenBank entry AAC-60766) were analyzed using the BlockMaker program, six conserved motifs were revealed (of length 37, 9, 28, 49, 33and 41 amino acids, respectively). The consensus sequences of those motifs were used to run the program PSI-BLAST to search the Genpept database. Except for the Wzx proteins of *N. punctiforme* and *Y. enterocolitica*, other distantly related Wzx proteins were retrieved after three iterations (E value = 4×10^{-9}). This indicates that *orf9* is a putative *wzx* gene.

orf7 of O155 was also predicted to encode protein with 12 transmembrane segments. Orf7 showed 73% identity to Orf9 of O59, and was also identified as putative Wzx using programs BlockMaker and PSI-BLAST as mentioned above.

Orf10 of O59 and Orf8 of O155 share 64% identity. Both show 41% identity or 65% similarity to CysE, a serine acetyltransferase of *Clostridium tetani* E88 [28], When searching against the Pfam and COG databases, Orf10 and Orf8 were found belonging to the COG1045 acetyltransferase family (E value = 2×10^{-12}). We proposed that *orf10* of O59 and *orf8* of O155 were putative acetyltransferase genes, and named both genes *wclD*.

To confirm that *wzx* and *wclD* genes are involved in O-antigen synthesis, we replaced *wzx* of both O59 and 0155 and *wclD* gene of O59 with CAT genes to make the mutant strains H1414, H1413, and H1056, respectively. While the wild-type and complement strains produced the normal LPS, the mutant strains produced only the lipid-A/core part of the LPS (Figs. 2(a)–(c)). We also showed that H1056 complemented with *wcld* gene of 0155 produced normal LPS (Fig. 2b), indicating the *wcld* genes of both O59 and 0155 have the same function. This indicates that *wzx* and *wclD* genes are required for the synthesis of O59 and O155 antigens.

3.4. wzx genes of E. coli O59 and O155 are evolutionary related

Wzx flips the undecaprenol-phosphate (Und-P) linked O units across the bacterial inner membrane. As the Und-P is also needed for the synthesis of cell wall peptidoglycan, teichoic acids and other carbohydrate polymers, accumulation of Und-P-linked O units on the cytoplasmic face of inner membrane results in cell death [3]. Thus, it is no surprise that the *wzx* gene is located within the main cluster in the 70 or so known cases (http://www.microbio.usyd.edu.au/BPGD/default.htm). This is the first time that the *wzx* gene was found outside of the main gene cluster.

The region between *gnd* and *hisI* in O59 and O155 contains the same genes in the same order, and the DNA identity between each pair of the genes is also high (Fig. 1). In comparison to *gnd*, which has a GC content of about 0.5, *wzx* and *wclD*, adjacent to *gnd* gene, have a lower GC content of about 0.30, and other genes in this region have a GC content of about 0.43. This suggests that the first two genes (*wzx* and *wclD*) were incorporated into their current position recently from the same ancestor or evolved one from the other.

Plasmids pLW1011 and pLW1012 expressing respective Wzx protein of O59 and O155 restored the O155 and O59 antigen synthesis in *wzx* mutants to levels that were comparable to those obtained with the parental strains (Fig. 2(c)). The fact that the two Wzx proteins are functionally interchangeable further confirms that the two genes were evolved from the same origin. We then calculated the synonymous substitution rate (*Ks*) and nonsynonymous substitution rate (*Ka*) between the O59 and O155 genes using program *K*-estimator (Fig. 1) [29]. Comparison between the two *wzx* genes gave a *Ks* of 1.2397, a *Ka* of 0.1685, and a *Ka*/*Ks* ratio of 0.14, the two *wclD* had a *Ks* of 1.1854, a *Ka* of 0.1873, and a *Ka*/*Ks* ratio of 0.16. Comparison of 67 house-keeping genes present in both *E. coli* K-12 and *S. enterica* serovar gave average values for *Ks* and *Ka* of 0.94 and 0.039, respectively, and an average *Ka*/*Ks* ration of 0.04 [30]. The much higher *Ka*/*Ks* ratio of *wzx* and *wclD* of the two strains suggests an occurrence of adaptive change after divergence for more efficiently flipping of different O units or acetylating on different O units. At this stage, it is not clear if the *wzx* and *wclD* genes of the two strains evolved from a common ancestor or one from the other, but in no doubt that they are both at an intermediate stage and the two genes will eventually be brought into the main gene between *galF* and *gnd*.

3.5. Identification of genes specific for E. coli O59 and O155

Glycosyltransferase and O unit processing genes (*wzx* and *wzy*) are usually specific to individual O-antigens [15]. Four primer pairs based on *wclA*, *wclB*, *wclC* and *wzy* from *E. coli* O59, and two pairs based on *wzy* and *wzx* from *E. coli* O155 were designed, and used to screen the DNA pools contained representatives of the 186 known O-antigen forms of *E. coli* and *Shigella* strains. Except for the pools containing *E. coli* O59 *and* O155, which gave the expected PCR products with respective primer pairs, none of the other pools produced the expected PCR products. Therefore, all of the genes tested were specific to *E. coli* O59 and O155, respectively. PCR based assays using the primers based on the O-antigen specific genes may be developed for the rapid detection and identification of *E. coli* O59 and O155 strains as well as their DNA serotyping.

3.6. Conclusions

The O-antigen gene clusters of *E. coli* O59 *and* O155 were found between *galF* and *hisI*, which contained twelve and ten genes, respectively. The designation of *wzy*, *wzx* and *wlcD* genes in *E. coli* O59 and O155 were further confirmed by mutation analysis. The *wzx* genes of the two strains were found downstream of *gnd* and complementary in trans. An acetyltransferase gene, *wclD*, is adjacent to *wzx* in both strains. Further comparison indicates that *wzx* and *wclD* genes of the two strains were evolved from the same origin. We suggest that *wzx* and *wclD* genes of the two strains site at the present position as an intermediate stage, they are expected to relocate into the main O-antigen gene clusters under selection pressure. The genes specific for *E. coli* O59 and O155 were also identified, which can be potentially used for rapid identification of the two strains by PCR.

Acknowledgements

This work was supported by the Chinese National Science Fund for Distinguished Young Scholars (30125001), NSFC General Program (30270029, 30370339, 30370023), the 863 Program (2002AA2Z2051) and funding from the Science and Technology Committee of Tianjin City (013181711) to LW.

References

[1] Reeves, P.R. and Wang, L. (2002) Genomic organization of LPS-specific loci. Current Topics in Microbiology and Immunology 264, 109–135.
[2] Feng, L., Senchenkova, S.N., Yang, J., Shashkov, A.S., Tao, J., Guo, H., Zhao, G., Knirel, Y.A., Reeves, P.R. and Wang, L. (2004) Structural and genetic characterization of the *Shigella boydii* type 13 O antigen. Journal of Bacteriology 182, 383–392.
[3] Liu, D., Cole, R. and Reeves, P.R. (1996) An O-antigen processing function for Wzx (RfbX): a promising candidate for O-unit flippase. Journal of Bacteriology 178, 2102–2107.
[4] Reeves, P.R. (1994) Biosynthesis and assembly of lipopolysaccharide In: Bacterial Cell Wall (A., Neuberger and L.L.M., van Deenen, Eds.), Vol. 27, pp. 281–314. Elsevier Science Publishers, Amsterdam, The Netherlands.
[5] Samuel, G. and Reeves, P.R. (2003) Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. Carbohydrate Research 338, 2503–2519.
[6] Wang, L., Huskic, S., Cisterne, A., Rothemund, D. and Reeves, P.R. (2002) The O-antigen gene cluster of Escherichia coli O55:H7 and identification of a new UDP-GlcNAc C4 epimerase gene. Journal of Bacteriology 184, 2620–2625.
[7] Lan, R. and Reeves, P.R. (1996) Gene transfer is a major factor in bacterial evolution. Molecular Biology and Evolution 13, 47–55.
[8] Aoyama, K.M., Haase, A.M. and Reeves, P.R. (1994) Evidence for effect of random genetic drift on G + C content after Lateral transfer of fucose pathway genes to *Escherichia coli* K-12. Molecular Biology and Evolution 11, 829–838.
[9] Wang, L., Andrianopoulos, K., Liu, D., Popoff, M.Y. and Reeves, P.R. (2002) Extensive variation in the O-antigen gene cluster within one Salmonella enterica serogroup reveals an unexpected complex history. Journal of Bacteriology 184, 1669–1677.
[10] Lan, R. and Reeves, P.R. (2000) Intraspecies variation in bacterial genomes: the need for a species genome concept. Trends in Microbiology 8, 396–401.
[11] Shepherd, J.G., Wang, L. and Reeves, P.R. (2000) Comparison of O-antigen gene clusters of *E.scheria coli* (*Shigella*) Sonnei and *Plesiomonas shigelloides* O17: Sonnei gained its current plasmid-borne O-antigen genes from *P. shigelloides* in a recent event. Infection and Immunity 68, 6056–6061.
[12] Lai, V., Wang, L. and Reeves, P.R. (1998) *Escherichia coli* clone Sonnei (*Shigella sonnei*) had a chromosomal O-antigen gene cluster prior to gaining its current plasmid-borne O-antigen genes. Journal of Bacteriology 180, 2983–2986.
[13] Feng, L., Tao, J., Guo, H., Xu, J., Li, Y., Rezwan, F., Reeves, P.R. and Wang, L. (2004) Structure of the *Shigella dysenteriae* 7 O antigen gene cluster and identification of its antigen specific genes. Microbial Pathogenesis 36, 109–115.
[14] Wang, L., Qu, W. and Reeves, P.R. (2001) Sequence analysis of four *Shigella boydii* O-antigen loci: implication for *Escherichia coli* and *Shigella* relationships. Infection and Immunity 69, 6923–6930.
[15] Wang, L. and Reeves, P.R. (1998) Organization of *Escherichia coli* O157 O antigen gene cluster and identification of its specific genes. Infection and Immunity 66, 3545–3551.
[16] Staden, R. (1996) The Staden sequence analysis package. Molecular Biotechnology 5, 233–241.
[17] Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A. and Barrell, B. (2000) Artemis: sequence visualisation and annotation. Bioinformatics 16, 944–945.
[18] Henikoff, S., Henikoff, J.G., Alford, W.J. and Pietrokovski, S. (1995) Automated construction and graphical presentation of protein blocks from unaligned sequences. Gene 163, GC17-26.
[19] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25, 3398–3402.
[20] Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M. and

Sonnhammer, J.M. () The Pfam protein families database. Nucleic Acids Research 30, 276–280.
[21] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: impoving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specifc gap penalties and weight matrix choice. Nucleic Acids Research 22, 4673–4680.
[22] Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proceedings of the National Academy of Sciences USA 97, 6640–6645.
[23] Yu, D., Ellis, H.M., Lee, E.C., Jenkins, N.A., Copeland, N.G. and Court, D.L. (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. Proceedings of the National Academy of Sciences USA 91, 5978–5983.
[24] Wang, L. and Reeves, P.R. (1994) Involvement of the galactosyl-1-phosphate transferase encoded by the *Salmonella enterica rfbP* gene in O antigen subunit processing. Journal of Bacteriology 176, 4348–4356.
[25] Pupo, G.M., Karaolis, D.K.R., Lan, R. and Reeves, P.R. (1997) Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and *mdh* sequence studies. Infection and Immunity 65, 2685–2692.
[26] Jensen, S.O. and Reeves, P.R. (2001) Molecular evolution of the GDP-mannose pathway genes (*manB* and *manC*) in *Salmonella enterica*. Microbiology 147, 599–610.
[27] Franco, V.A., Liu, D. and Reeves, P.R. (1998) The Wzz (Cld) protein in *Escherichia coli*: amino acid sequence variation determines O antigen chain length specificity. Journal of Bacteriology 180, 2670–2675.
[28] Bruggemann, H., Baumer, S., Fricke, W.F., Wiezer, A., Liesegang, H., Decker, I., Herzberg, C., Martinez-Arias, R., Merkl, R., Henne, A. and Gottschalk, G. (2003) The genome sequence of Clostridium tetani, the causative agent of tetanus disease. PNAS 100, 1316–1321.
[29] Comeron, J.M. (1999) *K*-Estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals. Bioinformatics 15, 763–764.
[30] Thampapillai, G., Lan, R. and Reeves, P.R. (1994) Molecular evolution in the *gnd* locus of *Salmonella enterica*. Molecular Biology and Evolution 11, 813–828.