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.

© 2005 Federation of European Microbiological Societies. Published by Elsevier B.V. All rights reserved.

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
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 (\textit{wzx}) and O-antigen polymerase (\textit{wzy}) [1]. The role of \textit{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 \textit{Wzy}. Both the \textit{wzx} and \textit{wzy} genes 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 \textit{E. coli} and \textit{Salmonella enterica}, the O-antigen gene clusters lie between the \textit{galF} and \textit{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 \textit{E. coli} O55, two of the colitose synthetic pathway genes, \textit{coli} and \textit{col2}, were found outside the O-antigen gene cluster, downstream of the \textit{gnd} gene, while the rest of the pathway genes were found within the O-antigen gene cluster between \textit{galF} and \textit{gnd} [6]. The location of \textit{coli} and \textit{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 \textit{Plesiomonas shigelloides} to \textit{E. coli} was also reported [11,12].

In this study, the region between \textit{galF} and \textit{his} operon of \textit{E. coli} O59 and O155 chromosome were sequenced, respectively, \textit{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 \textit{galF} and \textit{gnd}. However, the O unit flippase gene (\textit{wzx}) and an acetyltransferase gene were found outside the main O-antigen gene clusters, located between the \textit{gnd} gene and the \textit{his} operon, in both strains. Both \textit{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 \textit{E. coli} O59 and O155 gene clusters were recently generated from a common ancestor, and they are at a stage to have the \textit{wzx} genes adapted specifically to the respective O units and the \textit{wzx} and acetyltransferase gene moved to the main cluster. Finally, by screening against all 186 \textit{Shigella} and \textit{E. coli} O serotypes, genes specific for each of \textit{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'-ATTGTGGCGGCTAACAA-AGAAAT-3') and #1524 (5'-TAGTCGCGTNGGCC-TGGATTAAGTTCGC-3') [14] were used to amplify the region between \textit{galF} and \textit{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 \textit{gnd} gene and the \textit{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 \textit{wzx}, \textit{wzy} and \textit{wclD} genes from \textit{E. coli} O59 and O155

The \textit{wzx}, \textit{wzy} and \textit{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
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/C0995/996 and WL/C0991/992 were used for replacement of wzx; WL/C0997/998 and WL/C0993/994 for wzy in E. coli O59 and O155, respectively; WL/C0983/984 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 NeoI and BamHI or NeoI and EcoRI 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
Primer used in this work

| Primers      | Oligonucleotide sequences | Description                                      | Reference |
|--------------|---------------------------|--------------------------------------------------|-----------|
| WL-359 (F)   | 5'-CCTGGTTGTGATGTGAT(TC)CTGGA(TC) | From gnd, for amplification between gnd and hisI | This study|
|              | GAAG(ACT)TAGTGA-3'         |                                                  |           |
| WL-360 (R)   | 5'-CTGCGAGACCT(CT)ATCGG(CAG)(CC)AG | From hisI, for amplification between gnd and hisI | This study|
|              | ACCTGGCACAAA-3'            |                                                  |           |
| WL-991 (F)   | 5'-TATATATTTCTCTATTCTAAGTCTAC | wxx (O155)flanking DNA, for wxx (O155) replacement | This study|
|              | GTGAGATTAGTAGG(AAG)GAGG-3'  |                                                  |           |
| WL-992 (R)   | 5'-ACCCCCCTTAAATAGCTGATAAATAA | wxx (O155)flanking DNA, for wxx (O155) replacement | This study|
|              | TTGGCTAATATCTAAAAATAATCGGCCC-3' |                                                  |           |
| WL-993 (F)   | 5'-ATAAAACGCGCTTTTTCTGTA(AAA)AGGAAGAGGA | wxy (O155)flanking DNA, for wxy (O155) replacement | This study|
|              | AGTATGGAGAAAAAGCTACCTGG-3' |                                                  |           |
| WL-994 (R)   | 5'-AAAATTAGCGGTTACGATATAATAATACA | wxy (O155)flanking DNA, for wxy (O155) replacement | This study|
|              | TCTGATCTAATATACGATCGCCC-3' |                                                  |           |
| WL-995 (F)   | 5'-TTTGAATATCTTCATTAGGTCAAATATGTGATG | wxy (O59)flanking DNA, for wxy (O59) replacement | This study|
|              | AGATAATAAGTGAAGAAAAATACCTGG-3' |                                                  |           |
| WL-996 (R)   | 5'-CCATTTTTTTTTTTCTTGTACTGTATTTTT | wxy (O59)flanking DNA, for wxy (O59) replacement | This study|
|              | GGCACCTATAAAAATATCGGCCC-3' |                                                  |           |
| WL-997 (F)   | 5'-AATAAAACCTTTGTACCTTTATTATACAAA | wxy (O59)flanking DNA, for wxy (O59) replacement | This study|
|              | TCGTATCGGAGGCTAAAGGAGCTAAA |                                                  |           |
|              | TGGC-3' |                                                  |           |
| WL-998 (R)   | 5'-ATTTTACCAATATATTTAAAAAAGACGACATAA | wxy (O59)flanking DNA, for wxy (O59) replacement | This study|
|              | CGGATATTTGTA(AAA)AAATAATCGGCCC-3' |                                                  |           |
| WL-983 (F)   | 5'-GGAATTATTAAGGAGACTTTCTCCTATTTTA | wxy (O59)flanking DNA, for wxy (O59) replacement | This study|
|              | TGAAGATTTGGGAGCTAAGGAAGCTA(AAT)G-3' |                                                  |           |
| WL-984 (R)   | 5'-TCATCGGTTGACGCAACGACATCTTTATCCTCAGAAGCTCTATAAAAAATATCGGCCC-3' | wxy (O59)flanking DNA, for wxy (O59) replacement | This study|
| WL-971 (F)   | 5'-CGAGATCTCCATGGCTAATATATCAA | From wxx (O155), for amplification of wxx (O155) | This study|
|              | AATG-3' |                                                  |           |
| WL-972 (R)   | 5'-GAGATTGGAATCTCCATCCTTCTATAATA-3' | From wxx (O59), for amplification of wxx (O59) | This study|
| WL-973 (F)   | 5'-TAATCCATGGAATCATAATAATAATATATGAAATGCAATG-3' | From wxx (O59), for amplification of wxx (O59) | This study|
| WL-974 (R)   | 5'-TTTGGACTTCTTACTGATATTTTGGGAC-3' | From wxx (O59), for amplification of wxx (O59) | This study|
| WL-3947 (F)  | 5'-TAATCCATGGAATAAGGAAAAATTTCAAG-3' | From wyy (O59), for amplification of wyy (O59) | This study|
| WL-3948 (F)  | 5'-TCGCAGATCTTACATCTTTTCTCTATTACA | From wyy (O59), for amplification of wyy (O59) | This study|
| WL-3949 (F)  | 5'-TCATCGGTTGAGGGCGCTTTATTACAAAG-3' | From wyy (O59), for amplification of wyy (O59) | This study|
| WL-3950 (R)  | 5'-GCAGAATTTCATCTATTTGTGAGCAACATTGAC-3' | From wyy (O59), for amplification of wyy (O59) | This study|
| WL-3951 (F)  | 5'-TCATCGGTTGAGGGCGCTTTATTACAAAG-3' | From wyy (O59), for amplification of wyy (O59) | This study|
| WL-3952 (R)  | 5'-TCGCAGATCTTACATCTTTTCTTTAAAGAATAC-3' | From wyy (O59), for amplification of wyy (O59) | This study|
| WL-3953 (F)  | 5'-TGATCCGATGGGTGACATTGGAAGGAAAAG-3' | From wyy (O59), for amplification of wyy (O59) | This study|
| WL-3954 (R)  | 5'-TCGCCAGATCCTGAGAATATTAGTAG-3' | From wyy (O59), for amplification of wyy (O59) | This study|

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

The putative function of \textit{wzy} was further confirmed by comparison of LPS phenotype between the wild type and the mutant in which \textit{wzy} was replaced by a CAT gene. While H1412, the \textit{wzy} deficient mutant strain of \textit{E. coli} O59, produced only lipid-A/core part of the LPS and one oligosaccharide unit, the wild type \textit{E. coli} O59 and complement strain (H1125) produced normal LPS (Fig. 2(b)). The same was found for the mutant and wild type \textit{E. coli} O155 (Fig. 2(a)). This confirms the designation of \textit{wzy}, and the regions between \textit{galF} and \textit{gnd} in \textit{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 \textit{E. coli} O59 and O155 have not been elucidated, and the sequencing data suggest that they are substantively different. Both \textit{manB} and \textit{manC} are exclusively used for GDP-\textit{d}-mannose synthesis.

Table 3

| Orf No. | Gene name | No. of a.a. | %G+C content | Similar proteins (accession number) | %Identical/%Similar (residues) | Putative function |
|---------|-----------|------------|--------------|-------------------------------------|-------------------------------|------------------|
| orf1    | galF      | 254        | 50.0         | UTP-glucose-1-phosphate uridylyltransferase, \textit{Escherichia coli} (AB008676) | 100/100(254) | UTP-glucose-1-phosphate uridylyltransferase |
| orf2    | wclA      | 403        | 32.1         | Glycosyltransferase, \textit{Agrobacterium tumefaciens} str. C58 (AAC44363) | 26/50(411) | Glycosyltransferase |
| orf3    | wzy       | 389        | 28.6         | O-antigen polymerase, \textit{Escherichia coli} O6(AAJ426423) | 27/45(447) | O-antigen polymerase |
| orf4    | wclB      | 358        | 34.3         | Mannosyltransferase, \textit{Brucella melitensis} (AAC98618) | 38/58(376) | Mannosyltransferase |
| orf5    | wclC      | 347        | 37.2         | Glycosyltransferase, \textit{Escherichia coli} O6 (CAD19797) | 42/60(401) | Glycosyltransferase |
| orf6    | manC      | 465        | 37.1         | Mannose-1-phosphate guanylyltransferase, \textit{Escherichia coli} O113 (BAB36277) | 66/80(478) | Mannose-1-phosphate guanylyltransferase |
| orf7    | manB      | 476        | 37.2         | Phosphomannomutase, \textit{Shigella boydii} O5 (AAL27334) | 92/97(471) | Phosphomannomutase |
| orf8    | gnd       | 468        | 48.9         | 6-Phosphogluconate dehydrogenase, \textit{Escherichia coli} K12 (AAA24496) | 98/98(468) | 6-Phosphogluconate dehydrogenase |
| orf9    | wzx       | 427        | 30.6         | Wzx, \textit{Yersinia enterocolitica} type O8 (AAC60766) | 25/52(429) | O-antigen filippase |
| orf10   | wclD      | 152        | 31.4         | Serine acetyltransferase, \textit{Clostridium tetani} E88 (AAO34989) | 41/65(186) | Acetyltransferase |
| orf11   | ugd       | 388        | 42.7         | UDP-Glc-6-dehydrogenase Ugd, \textit{Escherichia coli} O113 (AAD50493) | 98/98(388) | UDP-Glc-6-dehydrogenase |
| orf12   | gla       | 334        | 42.2         | GlcA, \textit{Escherichia coli} O113 (AAD50494) | 99/99(334) | GlcA epimerase |
| orf13   | wzz       | 344        | 46.2         | O-antigen chain length determinant Wzz, \textit{Escherichia coli} O55 (AAC07656) | 92/96(345) | O-antigen chain length determinant |
| orf14   | hisI      | 115        | 56.0         | Phosphoribosyl-AMP cyclohydrolase \textit{Escherichia coli} O157 (BAA77339) | 99/100(203) | Phosphoribosyl-AMP cyclohydrolase |
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].

![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.](image)

### Table 4

| 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       | welE      | 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       | welF      | 350         | 26.6           | Glucosyltransferase, Methanosarcina mazei Goe1 (AAM30345) | 24/52(306) | Glucosyltransferase |
| orf5       | welG      | 272         | 32.1           | WbwC, Escherichia coli O104 (AAK64375) | 58/76(268) | Glucosyltransferase |
| 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       | welD      | 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 |
3.3. \textit{wzx} and acetyltransferase genes are located downstream of the main gene cluster in \textit{E. coli} O59 and O155

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

\textit{orf9} of \textit{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 \textit{orf9} and the putative Wzx proteins of \textit{Nostoc punctiforme} (GenBank entry ZP_00108476) and \textit{Yersinia enterocolitica} (GenBank entry AAC-60766) were analyzed using the BlockMaker program, six conserved motifs were revealed (of length 37, 9, 28, 49, 33 and 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 \textit{N. punctiforme} and \textit{Y. enterocolitica}, other distantly related Wzx proteins were retrieved after three iterations (\textit{E} value = 4 \times 10^{-3}). This indicates that \textit{orf9} is a putative \textit{wzx} gene.

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

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

To confirm that \textit{wzx} and \textit{wclD} genes are involved in O-antigen synthesis, we replaced \textit{wzx} of both O59 and O155 and \textit{wclD} gene of O59 with \textit{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 \textit{wclD} gene of O155 produced normal LPS (Fig. 2b), indicating the \textit{wclD} genes of both O59 and O155 have the same function. This indicates that \textit{wzx} and \textit{wclD} genes are required for the synthesis of O59 and O155 antigens.

3.4. \textit{wzx} genes of \textit{E. coli} O59 and O155 are evolutionary related

\textit{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 \textit{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 \textit{wzx} gene was found outside of the main gene cluster.

The region between \textit{gnd} and \textit{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 \textit{gnd}, which has a GC content of about 0.5, \textit{wzx} and \textit{wclD}, adjacent to \textit{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 (\textit{wzx} and \textit{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 \textit{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 (\textit{Ks}) and nonsynonymous substitution rate (\textit{Ka}) between the O59 and O155 genes using program K-estimator (Fig. 1) [29]. Comparison between the two \textit{wzx} genes gave a \textit{Ks} of 1.2397, a \textit{Ka} of 0.1685, and a \textit{Kal/Ks} ratio of 0.14, the two \textit{wclD} had a \textit{Ks} of 1.1854, a \textit{Ka} of 0.1873, and a \textit{Kal/Ks} ratio of 0.16. Comparison of 67 house-keeping genes present in both \textit{E. coli} K-12 and \textit{S. enterica} serovar gave average values for \textit{Ks} and \textit{Ka} of 0.94 and 0.039, respectively, and an average \textit{Kal/Ks} ration of 0.04 [30]. The much higher \textit{Kal/Ks} ratio of \textit{wzx} and \textit{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 \textit{wzx} and \textit{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 \textit{galF} and \textit{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 wclD 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 flipase. 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) Comparative analysis of the O-antigen gene clusters of E.coli O157, O59 and O155. Journal of Bacteriology 184, 1669–1677.

[10] Lan, R. and Reeves, P.R. (2000) Intrasppecies 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-antigene gene clusters of Escherichia coli (Shigella) Sonnei and Plesiomonas shigelloides O17: Sonnei gained its current plasmid-borne O-antigene 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-antigene cluster prior to gaining its current plasmid-borne O-antigene 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. (1990) 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., Ettwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M. and
Sonnhammer, J.M. (1994) The Pfam protein families database. Nucleic Acids Research 30, 276–280.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Research 22, 4673–4680.

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.

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.

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.

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.

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.

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.

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.

Comeron, J.M. (1999) K-Estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals. Bioinformatics 15, 763–764.

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.