Journal of Medical Microbiology (2007), **56**, 177–184 **DOI** 10.1099/jmm.0.46775-0

Relationship between O-antigen subtypes, bacterial surface structures and O-antigen gene clusters in *Escherichia coli* O123 strains carrying genes for Shiga toxins and intimin

Lothar Beutin,[1] Quan Wang,[2,3] Dieter Naumann,[4] Weiqing Han,[2,3] Gladys Krause,[1] Luciana Leomil,[5] Lei Wang[2,3] and Lu Feng[2,3]

Correspondence
Lu Feng
fenglu63@nankai.edu.cn

[1]National Reference Laboratory for *Escherichia coli*, Centre for Infectiology and Pathogen Characterization, Federal Institute for Risk Assessment (BfR), Diedersdorfer Weg 1, D-12277 Berlin, Germany

[2]TEDA School of Biological Sciences and Biotechnology, Nankai University, 23 HongDa Street, TEDA, Tianjin 300457, P. R. China

[3]Tianjin Key Laboratory for Microbial Functional Genomics, TEDA College, Nankai University, 23 HongDa Street, TEDA, Tianjin 300457, P. R. China

[4]Robert Koch Institute, P13, Nordufer 20, D-13353 Berlin, Germany

[5]Departamento de Microbiologia, Instituto de Ciências Biomédicas II, Universidade de São Paulo, 05508-900, São Paulo, SP, Brazil

Escherichia coli O123 strains express a broad spectrum of phenotypes, H serotypes and virulence markers and are able to colonize and to cause disease in different hosts including humans. In this study, two subtypes of *E. coli* O123 antigen (group I and group II) have been identified based on their cross-reactions with other *E. coli* O antigens. Investigation of the relationship between O123 group I and group II strains by O serotyping and Fourier transform infrared spectroscopy of whole bacteria revealed surface structural differences between these two groups of *E. coli* O123 strains. Nucleotide sequence analysis of the O-antigen gene clusters of two *E. coli* O123 strains representing O123 group I and group II revealed no change at the amino acid level. These findings indicate that the differences in the surface structures of group I and group II strains are not related to genetic heterogeneity in their O-antigen gene clusters. A PCR assay based on O123 antigen-specific *wzx* and *wzy* genes was developed and found to be suitable for reliable detection of all subtypes of *E. coli* O123 strains, which bears an advantage over traditional serological detection.

Received 12 June 2006
Accepted 14 September 2006

INTRODUCTION

The O antigen, which consists of many repeats of an oligosaccharide unit (O unit), is the outer component of LPS in the surface of Gram-negative bacteria (Reeves & Wang, 2002). Due to the presence of different sugars and sugar linkages, it is one of the most variable cell constituents which plays an important role in bacterial evasion of host defence systems (Reeves, 1995).

Abbreviations: AEEC, attaching and effacing *Escherichia coli*; FT-IR, Fourier transform infrared; STEC, Shiga toxin-producing *Escherichia coli*.

The GenBank/EMBL/DDBJ accession numbers for the sequences of the *E. coli* O123 strains 43w and CB9827 O-antigen gene clusters are DQ676933 and DQ676934, respectively.

Genes involved in O-antigen synthesis are normally clustered between two housekeeping genes, *galF* and *gnd*, in *Escherichia coli*, and are commonly classified into three main classes: sugar biosynthetic pathway genes, sugar transferase genes and O-antigen processing genes including the flippase (Wzx) and polymerase (Wzy) genes (Reeves & Wang, 2002). Genes encoding Wzx and Wzy are normally specific to different O antigens, and can be used for the development of a PCR assay for the identification and detection of individual strains. Most O-antigen gene clusters have a low G+C content, and it has been proposed that the O-antigen gene cluster was acquired by transfer from other low G+C content species (Wang & Reeves, 2000).

The *E. coli* serogroup O123 was defined by Orskov in 1952 with prototype strain 43w isolated from blood of a calf with

septicaemia (Orskov, 1952; Orskov *et al.*, 1977). Interest in *E. coli* O123 strains has recently been regained because of their roles as Shiga toxin-producing *E. coli* (STEC) and as attaching and effacing *E. coli* (AEEC) from animals and humans. STEC O123 strains have been isolated from healthy sheep (Beutin *et al.*, 1993; Djordjevic *et al.*, 2004) and diarrhoeic cattle and humans (Beutin *et al.*, 2004; Leomil *et al.*, 2003; Mercado *et al.*, 2004). Non-motile STEC O123 and O123 : H10 strains from sheep were associated with production of stx_{1c}- and stx_{2d}-variant toxins and most of these strains were negative for intimin (Beutin *et al.*, 2004; Brett *et al.*, 2003; Ramachandran *et al.*, 2001). Stx1- and Stx2-producing O123 : H2 and O123 : H11 strains positive for intimins β1 and ε have been isolated from diarrhoeic cattle and human patients (Beutin *et al.*, 2004; Leomil *et al.*, 2003; Mercado *et al.*, 2004). Stx-negative *E. coli* O123 strains producing intimin (AEEC) have been isolated as causative agents of diarrhoea in calves and piglets (Malik *et al.*, 2006; Orden *et al.*, 1998). Current data indicate that *E. coli* O123 strains express a broad spectrum of phenotypes, H serotypes and virulence markers and are able to colonize and to cause disease in different hosts including humans.

We previously reported that *E. coli* O123 strains from pigs were different from the prototype O123 strains by showing an unusual cross-reaction with *E. coli* O4 specific antiserum that had not been described before (Malik *et al.*, 2006). We became interested in characterizing the differences between two O-antigen types of *E. coli* O123; therefore, in this report, we analysed the representative *E. coli* O123 strains of two O-antigen types for their serological cross-reactions by O serotyping, for alterations in their surface structures by using Fourier transform-infrared (FT-IR) spectroscopy, and for their O-antigen-encoding genes by nucleotide sequencing. A PCR assay specific for all types of *E. coli* O123 strains was developed on the basis of *E. coli* O123 *wzx* and *wzy* genes in the O-antigen gene cluster and found to be useful for the detection of *E. coli* O123 strains.

METHODS

Bacterial strains. *E. coli* O123 reference strain 43w (laboratory stock no. G1169) was obtained from the Institute of Medical and Veterinary Science, Adelaide, Australia (IMVS). Other *E. coli* and *Shigella* reference strains used were as previously described (Feng *et al.*, 2004c). The clinical isolates of *E. coli* O123, O4 and O12 listed in Table 1 were from the strain collection of the Federal Institute for Risk Assessment (BfR), Berlin. Serotyping of O and H antigens was performed in titration assays as described by Orskov & Orskov (1984). The strength of the agglutination reaction was indicated as the reciprocal value of the highest dilution of antiserum causing visible agglutination of bacteria (Ewing, 1986; Orskov, 1952; Orskov *et al.*, 1977). Rabbit antisera against *E. coli* O and H antigens were produced at the BfR, according to standard methods (Orskov & Orskov, 1984). Subtyping of flagellar (*fliC*) genes by PCR was performed as described previously (Beutin *et al.*, 2005).

Phenotypical and genotypical characterization of virulence markers. The strains were investigated for production of cytotoxins with the Vero cell toxicity test and for cytolysins (haemolysins) on washed blood agar plates (Beutin *et al.*, 2004). Subtyping of Shiga toxin, haemolysin and intimin genes was performed as described previously (Beutin *et al.*, 2005).

FT-IR spectroscopy. For FT-IR spectroscopy the *E. coli* strains were subcultured on Caso agar (Merck) for 24 h as overnight cultures applying a four quadrant streak pattern. Cells were harvested and prepared for FT-IR measurements as already described by Helm *et al.* (1991a, b). Briefly, 35 µl of a bacterial cell suspension in distilled water was transferred to a ZnSe optical plate and dried to a transparent film under mild vaccum (2.5–7.5 kPa). All spectra between 500 and 4000 cm^{-1} were recorded on an IFS 28/B Fourier transform infrared spectrometer (Bruker Optik) equipped with a deuterated triglycerine sulphate detector. Nominal physical resolution was set to 6 cm^{-1}, a Blackman/Harris apodization was used for Fourier transformation and a zerofilling factor of 4 was applied to yield an encoding interval of approximately one data point per wavenumber. Spectral data were collected and evaluated with OPUS 3.0 software (Opus Software). Data processing included calculation of second derivatives to minimize baseline problems and to enhance apparent resolution. For hierarchical cluster analysis, an unsupervised classification technique, the Ward's algorithm (Helm *et al.*, 1991a, b), was used to construct the dendrograms. As a distance measure, Pearson's product moment correlation coefficient was used as defined previously (Helm *et al.*, 1991a, b).

Construction of DNaseI shotgun bank, and sequencing and analysis of the O-antigen gene cluster. Chromosomal DNA was prepared as previously described (Bastin & Reeves, 1995). The primer pair 5′-ATTGGTAGCTGTAAGCCAAGGGCGGTAGCGT-3′ and 5′-CACTGCCATACCGACGACGCCGATCTGTTGCTTGG-3′, based on JUMPstart sequence and *gnd*, respectively (Wang & Reeves, 1998), was used to amplify the O-antigen gene clusters of *E. coli* O123 type I strain 43w and *E. coli* O123 type II strain CB9827. 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. Shotgun banks for each strain were constructed as described previously (Wang & Reeves, 1998). Sequencing was carried out using an ABI 3730 automated DNA sequencer and sequence data were analysed using computer programs as described previously (Feng *et al.*, 2004a).

Specificity and sensitivity tests of O-serogroup specific PCR assay. Chromosomal DNA was prepared from each of 186 reference strains to represent the broadest range of O antigens of *E. coli* (including *Shigella*) and used to make DNA pools as described by Feng *et al.* (2004c). A total of 13 pools were made, each containing DNA from 12–19 strains (Feng *et al.*, 2004c). Primer pairs based on *wzx* and *wzy* genes of *E. coli* O123 (Table 3) were used to screen the DNA pools. The PCR cycles used were as follows: denaturation at 95 °C for 30 s, annealing at 50 °C for 30 s, extension at 72 °C for 1 min, 30 cycles. Template DNA from 48 clinical *E. coli* isolates including 22 *E. coli* O123 and 26 other O-serogroup strains (data not shown) was prepared as described by Guo *et al.* (2004) and screened using the same primers in a double-blind test.

To test the sensitivity of the PCR assay, 10-fold serial dilutions of chromosomal DNA from *E. coli* O123 strain 43w were amplified using all of the primer pairs designed. Primer pairs wl-2400/wl-2401 and wl-2406/wl-2407 from *E. coli* O123 were also used to screen *E. coli* O123 strain 43w in pork and water samples. Raw pork was purchased from three local butchers, weighed into 20 g portions, and stored at −40 °C before use. Serial 10-fold dilutions (10^{-3}–10^{-9}) from the full-grown culture of *E. coli* O123 strain 43w were added to each portion of pork or 20 ml Milli-Q water. The concentration of the O123 cells in pork or water samples was determined by checking the c.f.u. on agar plates.

Table 1. Properties of *E. coli* strains used in this work

Strain	Country*	Source, disease†	Virulence markers‡	Serotype	Agglutination with O-specific antisera (immunizing strain)§			
					O123 group I (43w)	O4 (U4-41)	O123 group II (CB9827)	O12 (Bi 626-42)
43w	DK, 1952	Calf, b, S	None	O123 : H16	3200	< 50	1600	800
CB1654	D, 1991	Human, f, D	*eae*$^+$	O123 : H19	3200	< 50	1600	1600
CB6793	D, 1997	Human, f, D	E-*hlyA*, *stx*$_1$, *eae*-β1	O123 : H11	3200	< 50	1600	1600
CB6805	D, 1997	Human, f, D	E-*hlyA*, *stx*$_1$, *eae*-β1	O123 : H11	6400	< 50	3200	1600
CB9716	D, 2003	Human, f, D	*eae*-ε	O123 : H25	6400	< 50	3200	1600
CB671	D, 1989	Calf, f, D	E-*hlyA*, *eae*-β1	O123 : H2	800	1600	3200	< 50
CB1930	D, 1992	Human, f, D	*eae*$^+$	O123 : H rough	400	400	800	< 50
CB4350	AUS, 1995	Human, f, D	E-*hlyA*, *stx*$_{1c}$ + *stx*$_{2d}$	O123 : H10	800	1600	3200	< 50
CB7157	D, 1997	Human, f, D	E-*hlyA*, *stx*$_{1c}$ + *stx*$_{2d}$	O123 : H10	400	800	1600	< 50
CB7815	D, 1997	Human, f, D	E-*hlyA*, *stx*$_1$, *eae*-ε	O123 : H2	1600	1600	3200	< 50
CB8817	D, 1997	Human, f, D	*stx*$_1$	O123 : H16	800	1600	3200	50
CB8900	N, 2001	Sheep, f, AS	E-*hlyA*, *stx*$_{1c}$ + *stx*$_{2d}$	O123 : H10	800	800	1600	< 50
CB9192	HU, 2002	Piglet, c, FD	*eae*-β1	O123 : H11	800	1600	3200	< 50
CB9200	HU, 2002	Piglet, c, FD	*eae*-β1	O123 : H11	800	1600	3200	< 50
CB9368	D, 2002	Human, f, D	*eae*-β1	O123 : H40	800	1600	3200	< 50
CB9588	D, 2003	Human, f, D	E-*hlyA*, *stx*$_{1c}$ + *stx*$_{2d}$	O123 : H10	800	1600	3200	< 50
CB9779	D, 2003	Human, f, D	*eae*$^+$	O123 : H45	800	800	3200	< 50
CB9790	D, 2003	Human, f, D	*eae*$^+$	O123 : H45	800	800	1600	< 50
CB9827	BR, 2002	Calf, f, D	E-*hlyA*, *stx*$_1$ + *stx*$_2$, *eae*-ε	O123 : H2	1600	1600	6400	< 50
CB9828	BR, 2002	Calf, f, D	*stx*$_1$ + *stx*$_2$, *eae*-ε	O123 : H2	800	800	1600	< 50
CB9829	BR, 2002	Calf, f, D	E-*hlyA*, *stx*$_1$ + *stx*$_2$, *eae*-ε	O123 : H2	800	1600	3200	< 50
CB9830	BR, 2002	Calf, f, D	E-*hlyA*, *stx*$_1$ + *stx*$_2$, *eae*-ε	O123 : H2	800	1600	3200	50
CB3037	D, 1993	Human, f, ND	alpha-*hlyA*	O4 : H5	50	6400	100	100
CB3039	D, 1993	Human, f, ND	alpha-*hlyA*	O4 : H5	50	6400	100	100
CB6932	D, 1997	Human, f, D	alpha-*hlyA*	O4 : H5	50	3200	50	50
CB7087	D, 1997	Human, f, HUS	alpha-*hlyA*	O4 : H5	50	3200	100	200
CB8211	D, 1999	Human, f, D	alpha-*hlyA*	O4 : H5	50	6400	100	200
J96	USA, 1981	Human, u, UTI	alpha-*hlyA*	O4 : K6 : H5	50	6400	100	100
U4-41	DK, 1941	Human, u, UTI	alpha-*hlyA*	O4 : K3 : H5	50	6400	200	100
Bi 626-42	DK, 1941	Human, p	None	O12 : K5 : NM	200	400	50	6400

*Origin of strains and year of isolation: DK, Denmark; D, Germany; AUS, Australia; N, Norway; HU, Hungary; BR, Brazil; USA, United States of America.

†b, Blood; f, faeces; c, colon biopsy; p, peritoneum; u, urine; S, septicaemia; D, diarrhoea; AS, asymptomatic; FD, fatal diarrhoea; ND, no data; HUS, haemolytic uraemic syndrome; UTI, urinary tract infection.

‡All strains positive for shiga-toxin (*stx*)-genes showed Vero cell toxicity. Subtyping of *stx*-genes and *stx*-gene nomenclature were performed as described previously (Beutin *et al.*, 2005). All strains positive for the E-*hlyA* gene showed an enterohaemolytic phenotype and all carrying an alpha-*hlyA* gene showed an alpha-haemolytic phenotype as previously described (Beutin *et al.*, 2005). Subtyping of intimin (*eae*) genes and nomenclature of *eae*-genes were performed and listed as described by Krause *et al.* (2005). Strains with *eae*-gene that could not be subtyped into the hitherto described types were designated *eae*$^+$.

§Reciprocal values of the highest dilution of antiserum which caused visible agglutination of bacteria.

Samples spiked with different concentrations of *E. coli* O123 cells were homogenized in 200 ml Luria broth culture. The homogenized samples were cultured at 200 r.p.m. at 37 °C for 12 h and chilled at 4 °C. The culture was passed through a six-chamber filter bag, and 3 ml filtrate was collected for use as the template DNA in the PCR as described above (Guo *et al.*, 2004).

RESULTS AND DISCUSSION

Properties of *E. coli* O123 strains

We have investigated 22 *E. coli* O123 strains that were isolated from diarrhoeic humans, calves, pigs and healthy

sheep. The strains, except *E. coli* O123 reference strain 43w, were from Europe, Latin America and Australia and had been isolated between 1989 and 2003 (Table 1). All these *E. coli* O123 strains except 43w carried one or more virulence markers associated with AEEC (*eae* gene), with STEC [*stx* gene(s)] or with enterohaemorrhagic *E. coli* (EHEC) [*stx*, *eae* and E-*hlyA* (EHEC haemolysin) genes]. *E. coli* O123 strains could be associated with multiple H types such as H2, H10, H11, H16, H19, H25, H40 and H45. Taken together, *E. coli* O123 strains presented as a group of human and animal pathogenic strains with a broad host spectrum and a high diversity in their virulence attributes.

By O serotyping, the *E. coli* O123 strains fall into two groups by their cross-reactions with antisera directed to *E. coli* O antigens other than O123. The prototype O123 strain 43w was previously reported to cross-react with *E. coli* O12 specific antiserum (Ewing, 1986; Orskov, 1952; Orskov *et al.*, 1977). This cross-reaction was confirmed in this study (O titres 1 : 800–1 : 1600) with 43w and also found with four other *E. coli* O123 strains (CB1654, CB6793, CB6805 and CB9716) (Table 1). These *E. coli* O123 strains were designated 'O123 group I' and none of these agglutinated significantly with *E. coli* O4 specific antiserum (agglutination titres $<1:50$). In contrast, all other *E. coli* O123 strains from this study did not agglutinate with *E. coli* O12 antiserum (agglutination titres $<1:50$), but did agglutinate with *E. coli* O4 specific antiserum (here called 'O123 group II' strains). The two groups of O123 strains differed in their agglutination reactions with the O123 specific antiserum made with strain 43w (O123 group I) as immunizing antigen. With this serum O123 group I strains showed higher agglutination titres (1 : 3200–1 : 6400) than O123 group II strains (titres 1 : 400–1 : 1600) (Table 1). To further analyse the specificity of cross-reactions obtained with the two groups of O123 strains we produced an O123 specific antiserum with *E. coli* O123 strain CB9827 as a prototype strain for the O123 group II strains. O antiserum made against CB9827 showed similar agglutination titres with *E. coli* O123 group I and group II strains but did not cause significant agglutination (titres 1 : 50–1 : 100) with *E. coli* O4 strains (Table 1).

FT-IR spectroscopy

FT-IR spectroscopy of intact micro-organisms provides information on the composition and structure of the whole cell (Naumann *et al.*, 1991; Naumann, 2000). These spectra are complex spectroscopic patterns encoding the signals of thousands of bands that cannot be resolved easily. Thus pattern recognition techniques are generally used to extract the essential information. It has been shown by several groups that FT-IR may provide information useful for discrimination at the genus, species and even strain level (Helm *et al.*, 1991b; Horbach *et al.*, 1988; Kirschner *et al.*, 2001; Maquelin *et al.*, 2003; Tintelnot *et al.*, 2000). Since all cell components depend on the expression of smaller or larger parts of the genome, the FT-IR spectra of micro-organisms display specifically a complete phenotypic and genetic fingerprint of the cells under study.

The purpose of this FT-IR analysis was to group the *E. coli* isolates according to their O-antigenic properties which reside in the structure of O-specific side chains of LPS. Since it was expected that differences in O antigen would primarily be expressed in the spectral range where the vibrational modes of cellular carbohydrates dominate the experimentally observed spectral contour, data analysis was directed to the spectral range of 900–1200 cm^{-1} (Helm *et al.*, 1991b). Using the information contained in this spectral range as input for cluster analysis a distinct clustering was observed according to the dendrogram shown in Fig. 1. This dendrogram clearly showed the presence of three main clusters according to the grouping scheme suggested by O serotyping. It was particularly notable that the two groups of O123 strains are spectroscopically more related than the O4 group to both, suggesting that the two underlying carbohydrate structures found at the cell surface of the two O123 group strains were indeed more similar to each other than to the carbohydrate structure of the O4 strains.

E. coli O123 group I and group II strains share the same O-antigen gene cluster

A sequence of 15 783 bases between JUMPStart and *gnd* was obtained from each of *E. coli* O123 strains 43w (group I) and CB9827 (group II), and each contained 16 genes (excluding *gnd*) with transcription direction from JUMPStart to *gnd* (Fig. 2).

The two sequences shared 99.94 % identity at DNA level and 100 % at protein level, and were regarded as identical. Therefore, their genes were given the same names and the functions of the genes were predicted based on their similarity to those from available databases (Table 2).

Sugar biosynthesis pathway genes. Orfs 1 and 2 shared high level identity (78–92 %) to many other known RmlB and RmlA proteins from other *E. coli*, *Shigella* and *Salmonella enterica* strains, respectively. Orf4 shared 60 % identity to VioA of *E. coli* O7. RmlB, RmlA and VioA were proposed as dTDP-D-4-amino-4,6-dideoxy-glucose biosynthesis pathway enzymes (Marolda *et al.*, 1999). Therefore, *orf1*, *orf2* and *orf4* were identified as *rmlB*, *rmlA* and *vioA*, and named accordingly.

Orfs 12, 13 and 14 shared 74, 53 and 62 % identity to WbvB (FnlA), WbvR (QnlA) and WbvD (QnlB) of the *Vibrio cholerae* O37 O antigen gene cluster, respectively, which were identified as UDP-L-*N*-acetyl-quinovosamine biosynthesis pathway enzymes experimentally (Kneidinger *et al.*, 2003). Therefore, *orf12*, *orf13* and *orf14* were identified as *fnlA*, *qnlA* and *qnlB*, and named accordingly.

Sugar transferase genes. Orf9 and Orf11 belonged to glycosyltransferase family 2 (PF00535, *E* value $= 9.0 \times e^{-25}$) and glycosyltransferase family 1 (PF00534, *E*

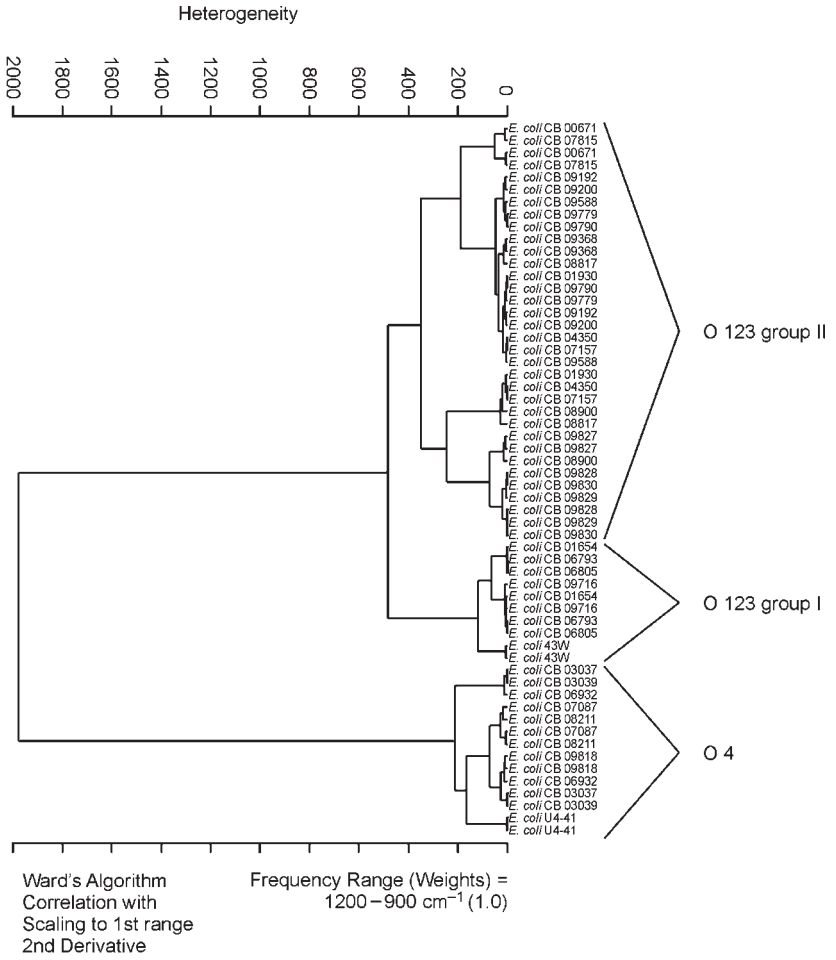

Fig. 1. FT-IR classification of the *E. coli* isolates. Cluster analysis was performed using the second derivatives of the spectra and the spectral information contained in the spectral range between 900 and 1200 cm^{-1}. Ward's algorithm was used for calculating the interrelationship between the strains. The numbers at the side of the dendrogram correspond to the different strains measured by FT-IR spectroscopy. All strains have been measured twice from independent cultivations.

value $= 1.1 \times e^{-12}$), and shared 50 and 63 % similarity to glycosyltransferases of *Clostridium acetobutylicum* and *Shigella boydii* type 13, respectively. Therefore, *orf9* and *orf11* were proposed to be glycosyltransferase genes and named *wfbE* and *wfbF*, respectively. Orf15 belonged to glycosyltransferase family 1 (PF00534, *E* value $= 5.7 \times e^{-8}$), and shared 55 % identity to WbwH of *S. boydii* type 13, which was a putative L-*N*-acetyl-quinovosamine transferase (Feng *et al.*, 2004a). Therefore, *orf15* was proposed to be a glycosyltransferase gene and named *wbwH*.

O-antigen processing genes. Both *orf3* and *orf10* were found to encode predicted membrane proteins. Orf3 had 10 predicted transmembrane segments, which is a typical topological character of Wzx proteins (Liu *et al.*, 1996). Orf3 belonged to the protein family PF01943 (*E* value $= 6.3 \times e^{-22}$), and members of this family including RfbX are involved in the export of the O antigen and teichoic acid. It also shared 66 and 65 % similarity to the Wzx proteins of *E. coli* O7 and *E. coli* O113, respectively. Therefore, *orf3* was proposed to be an O-unit flippase gene (*wzx*) and named accordingly. Orf10 had eight predicted transmembrane segments and a large periplasmic loop of 113 amino acid residues, which is a typical topological character of Wzy proteins (Daniels *et al.*, 1998). It also shared 48 % similarity to the Wzy protein of *E. coli* O121. Therefore, *orf10* was proposed to be an O-antigen polymerase gene (*wzy*) and named accordingly.

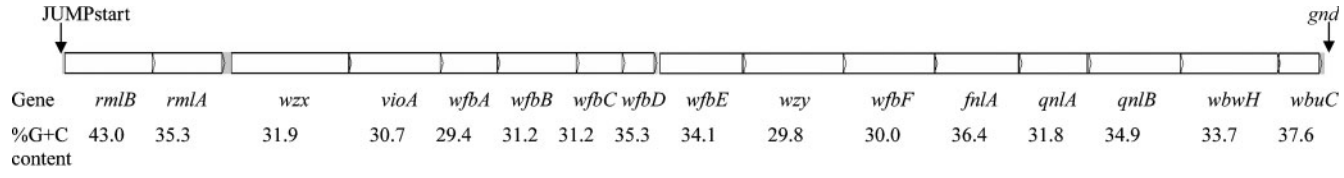

Fig. 2. O-antigen gene cluster organization of both *E. coli* O123 group I and group II. All the genes are transcribed in the direction from JUMPStart to *gnd*.

Table 2. Putative genes in the *E. coli* O123 O antigen gene cluster

Gene name	Location in sequence	G+C content (mol%)	Similar protein(s), strain(s) (GenBank accession no.)	% Identity/% similarity (residues)	Putative function of protein
rmlB	108–1181	43.0	RmlB *E. coli* K-12 (AAB88398)	90/93 (349)	dTDP-D-Glucose 4,6-dehydratase
rmlA	1182–2051	35.3	RmlA *E. coli* K-12 (AAB88400)	80/89 (289)	Glucose-1-phosphate thymidylyltransferase
wzx	2169–3614	31.9	Wzx *E. coli* O7 (AAD44153)	44/66 (474)	O unit flippase
vioA	3619–4749	30.7	VioA *E. coli* O7 (AAD44154)	60/76 (370)	dTDP-Viosamine synthetase
wfbA	4746–5453	29.4	WbnG *S. dysenteriae* type 7 (AAR97959)	38/60 (231)	Unknown
wfbB	5457–6431	31.2	Conserved hypothetical protein *C. violaceum* (AAQ61692)	26/45 (286)	Unknown
wfbC	6431–6997	31.2	Acetyltransferase *B. cereus* ATCC 14579 (AAP10400)	33/56 (177)	Acetyltransferase
wfbD	6994–7395	35.3	Monoamine oxidase C family protein *C. crescentus* CB15 (AAK22689)	48/70 (131)	Unknown
wfbE	7469–8494	34.1	Glycosyltransferase *Clostridium acetobutylicum* (AAK80301)	31/50 (243)	Glycosyltransferase
wzy	8478–9728	29.8	O antigen polymerase *E. coli* O121 (AAO39700)	23/48 (368)	O-antigen polymerase
wfbF	9725–10 846	30.0	Glycosyltransferase *S. boydii* type 13 (AAR24273)	40/63 (302)	Glycosyltransferase
fnlA	10 859–11 890	36.4	L-*N*-Acetyl-quinovosamine synthase *S. boydii* type 13 (AAR24274)	84/92 (342)	L-*N*-Acetyl-quinovosamine synthase
qnlA	11 883–12 752	31.8	L-*N*-Acetyl-quinovosamine synthase *S. boydii* type 13 (AAR24275)	54/71 (287)	L-*N*-Acetyl-quinovosamine synthase
qnlB	12 730–13 884	34.9	L-*N*-Acetyl-quinovosamine synthase *S. boydii* type 13 (AAR24276)	72/86 (383)	L-*N*-Acetyl-quinovosamine synthase
wbwH	13 884–15 077	33.7	Glycosyltransferase *S. boydii* type 13 (AAR24277)	55/71 (397)	Glycosyltransferase
wbuC	15 096–15 593	37.6	WbuC *E. coli* O26 (AAN60465)	42/62 (131)	Unknown

Other genes. Orf7 belonged to the acetyltransferase (GNAT) family PF00583 (E value $= 5.6 \times e^{-7}$), and it also shared 56 % similarity to the acetyltransferase of *Bacillus cereus* ATCC 14579. Therefore, Orf7 was proposed to be an acetyltransferase. Orf5 shared 60 % identity to WbnG of *Shigella dysenteriae* type 7, which was proposed to be a glycine transferase (Feng *et al.*, 2004b). Orf8 belonged to the protein family PF01575 (E value $= 5.6 \times e^{-7}$), members of which include monoamine oxidase C dehydratase. It also shared 60–70 % similarity to hydratase or dehydratase with monoamine oxidase C-like domain of many bacteria strains. No function can be assigned to *orf6* by searching currently available databases. As the *E. coli* O123 O-antigen structure has not been identified, only general functions of *orfs* 5–8 could be proposed, and *orfs* 5–8 were named *wfbA*, *wfbB*, *wfbC* and *wfbD*, respectively.

Orf16 shared 67 and 62 % similarity to a proposed remnant gene product, WbuC, of *S. boydii* type 13 (Feng *et al.*, 2004a) and *E. coli* O26 (D'Souza *et al.*, 2002). *orf16* was proposed to be non-functional and named *wbuC*.

Possible factors responsible for the differences between the O antigens of the two *E. coli* O123 groups

The data above indicated that the differences observed between the *E. coli* O123 group I and group II strains were not attributable to the genes encoded by the O-antigen gene cluster but to other functions responsible for the structures of the O antigen in *E. coli* O123 strains. These must be encoded by genes located outside of the O-antigen gene clusters between *galF* and *gnd* in *E. coli* O123 strains and their functions need to be explored. It was reported that the O antigen of *E. coli* O4 has a side branch UDP-D-Glu residue, which was proposed to be added after synthesis and translocation of the O unit, and the genes responsible for the side chain transferase were located outside of the O-antigen gene cluster (D'Souza *et al.*, 2005). Therefore, we proposed that the cross-reaction between *E. coli* O123 group II strains and *E. coli* O4 specific antiserum may be due to the side chain of the O antigen in both strains, although the O-antigen structures of *E. coli* O123 group I and group II strains were not characterized. The differences found between the *E. coli* O123 groups by serotyping could be related to differences in their surface structures (presumably the addition of side branch residues), which was also indicated by the results from FT-IR spectroscopy.

Identification of *E. coli* O123 O-antigen specific genes

Four primer pairs based on the proposed O unit processing genes *wzx* and *wzy* (two pairs for each) were designed (Table 3). All the primers were used to screen DNA pools containing representatives of the 186 known O antigens of typical *E. coli* and *Shigella* strains, and no PCR products were detected from other pools, except the pools containing *E. coli* O123, which gave the expected PCR products. All of the

Table 3. Primers used for genetic detection of all types of *E. coli* O123 strains

Gene	Base positions of gene	Forward primer/reverse primer	Length of PCR fragment (bp)
wzx	2169–3614	wl-2400 (459–478)/wl-2401 (1107–1126)	668
		wl-2402 (638–657)/wl-2403 (1206–1227)	590
wzy	8478–9728	wl-2404 (509–528)/wl-2405 (1007–1026)	518
		wl-2406 (126–145)/wl-2407 (994–1013)	888

primer pairs were further used for a double-blind test on *E. coli* clinical isolates including 22 *E. coli* O123 group I and group II strains (Table 1) and 26 *E. coli* strains of other O serogroups (data not shown). All of the *E. coli* O123 strains gave the expected PCR products corresponding to the primer pairs used, and no PCR products were obtained from other O-serotype strains. Therefore, all of these primer pairs are highly specific and can be used for the identification and detection of both serological groups of *E. coli* O123 strains.

Sensitivity of the serogroup specific PCR assay

All the four primer pairs of *E. coli* O123 were used to amplify a dilution series of chromosomal DNA from *E. coli* O123 type I strain 43w, and obtained a sensitivity of 1 pg μl^{-1}. The primer pairs of wl-2400/wl-2401 and wl-2406/wl-2407 of *E. coli* O123 were used to screen *E. coli* O123 strain 43w in pork and water samples. By this, as few as 10^3 c.f.u. g^{-1} were detected in samples which were examined directly and 0.1 c.f.u. g^{-1} could be detected in samples which were further incubated at 37 °C for 12 h. In contrast to traditional serological testing, which is slow, labour-intensive and subject to interference by cross-reactions, the specific PCR assay is specific, sensitive and could be easily performed to detect *E. coli* O123 group I and group II strains belonging to different clonal groups, and sero- and pathotypes, and changes of the bacterial surface which may cause difficulties in O serotyping are avoided.

ACKNOWLEDGEMENTS

The authors are grateful to Béla Nagy, Budapest, Hungary, for supplying porcine pathogen *Escherichia coli* O123 strains. This work was supported by the NSFC General Program (30370023, 30370339) and the NSFC Key Program (30530010). L. L. was supported from Brazil by 'Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)', number BEX 1686/03-8 and by 'Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)' grant no. 03/01823-9. The authors are also grateful to Maren Stämmler for carefully helping in the FT-IR measurements.

REFERENCES

Bastin, D. A. & Reeves, P. R. (1995). Sequence and analysis of the O antigen gene (*rfb*) cluster of *Escherichia coli* O111. *Gene* **164**, 17–23.

Beutin, L., Geier, D., Steinruck, H., Zimmermann, S. & Scheutz, F. (1993). Prevalence and some properties of verotoxin (Shiga-like toxin)-producing *Escherichia coli* in seven different species of healthy domestic animals. *J Clin Microbiol* **31**, 2483–2488.

Beutin, L., Krause, G., Zimmermann, S., Kaulfuss, S. & Gleier, K. (2004). Characterization of Shiga toxin-producing *Escherichia coli* strains isolated from human patients in Germany over a 3-year period. *J Clin Microbiol* **42**, 1099–1108.

Beutin, L., Tao, J., Feng, L., Krause, G., Zimmermann, S., Gleier, K., Xia, Q. & Wang, L. (2005). Sequence analysis of the *Escherichia coli* O15 antigen gene cluster and development of a PCR assay for rapid detection of intestinal and extraintestinal pathogenic *E. coli* O15 strains. *J Clin Microbiol* **43**, 703–710.

Brett, K. N., Ramachandran, V., Hornitzky, M. A., Bettelheim, K. A., Walker, M. J. & Djordjevic, S. P. (2003). stx_{1c} is the most common Shiga toxin 1 subtype among Shiga toxin-producing *Escherichia coli* isolates from sheep but not among isolates from cattle. *J Clin Microbiol* **41**, 926–936.

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

Djordjevic, S. P., Ramachandran, V., Bettelheim, K. A., Vanselow, B. A., Holst, P., Bailey, G. & Hornitzky, M. A. (2004). Serotypes and virulence gene profiles of shiga toxin-producing *Escherichia coli* strains isolated from feces of pasture-fed and lot-fed sheep. *Appl Environ Microbiol* **70**, 3910–3917.

D'Souza, J. M., Wang, L. & Reeves, P. R. (2002). Sequence of the *Escherichia coli* O26 antigen gene cluster and identification of O26 specific genes. *Gene* **297**, 123–127.

D'Souza, J. M., Samuel, G. N. & Reeves, P. R. (2005). Evolutionary origins and sequence of the *Escherichia coli* O4 O-antigen gene cluster. *FEMS Microbiol Lett* **244**, 27–32.

Ewing, W. H. (1986). *Edwards and Ewing's Identification of the Enterobacteriaceae*, 4th edn. Amsterdam: Elsevier.

Feng, L., Senchenkova, S. N., Yang, J., Shashkov, A. S., Tao, J., Guo, H., Zhao, G., Knirel, Y. A., Reeves, P. & Wang, L. (2004a). Structural and genetic characterization of the *Shigella boydii* type 13 O antigen. *J Bacteriol* **186**, 383–392.

Feng, L., Tao, J., Guo, H., Xu, J., Li, Y., Rezwan, F., Reeves, P. R. & Wang, L. (2004b). Structure of the *Shigella dysenteriae* 7 O antigen gene cluster and identification of its antigen specific genes. *Microb Pathog* **36**, 109–115.

Feng, L., Wang, W., Tao, J., Guo, H., Krause, G., Beutin, L. & Wang, L. (2004c). Identification of *Escherichia coli* O114 O-antigen gene cluster and development of an O114 serogroup-specific PCR assay. *J Clin Microbiol* **42**, 3799–3804.

Guo, H., Feng, L., Tao, J., Zhang, C. & Wang, L. (2004). Identification of *Escherichia coli* O172 O-antigen gene cluster and development of a serogroup-specific PCR assay. *J Appl Microbiol* **97**, 181–190.

Helm, D., Labischinski, H. & Naumann, D. (1991a). Elaboration of a procedure for identification of bacteria using Fourier-transform infrared spectral libraries: a stepwise correlation approach. *J Microbiol Methods* **14**, 127–142.

Helm, D., Labischinski, H., Schallehn, G. & Naumann, D. (1991b). Classification and identification of bacteria by Fourier-transform infrared spectroscopy. *J Gen Microbiol* **137**, 69–79.

Horbach, I., Naumann, D. & Fehrenbach, F. J. (1988). Simultaneous infections with different serogroups of *Legionella pneumophila* investigated by routine methods and Fourier transform infrared spectroscopy. *J Clin Microbiol* **26**, 1106–1110.

Kirschner, C., Maquelin, K., Pina, P., Ngo Thi, N. A., Choo-Smith, L. P., Sockalingum, G. D., Sandt, C., Ami, D., Orsini, F. & other authors (2001). Classification and identification of enterococci: a comparative phenotypic, genotypic, and vibrational spectroscopic study. *J Clin Microbiol* **39**, 1763–1770.

Kneidinger, B., Larocque, S., Brisson, J., Cadotte, N. & Lam, J. S. (2003). Biosynthesis of 2-acetamido-2,6-dideoxy-L-hexoses in bacteria follows a pattern distinct from those of the pathways of 6-deoxy-L-hexoses. *Biochem J* **371**, 989–995.

Krause, G., Zimmermann, S. & Beutin, L. (2005). Investigation of domestic animals and pets as a reservoir for intimin- (*eae*) gene positive *Escherichia coli* types. *Vet Microbiol* **106**, 87–95.

Leomil, L., Aidar-Ugrinovich, L., Guth, B. E., Irino, K., Vettorato, M. P., Onuma, D. L. & de Castro, A. F. (2003). Frequency of Shiga toxin-producing *Escherichia coli* (STEC) isolates among diarrheic and non-diarrheic calves in Brazil. *Vet Microbiol* **97**, 103–109.

Liu, D., Cole, R. & Reeves, P. R. (1996). An O-antigen processing function for Wzx(RfbX): a promising candidate for O-unit flippase. *J Bacteriol* **178**, 2102–2107.

Malik, A., Toth, I., Beutin, L., Schmidt, H., Taminiau, B., Dow, M. A., Morabito, S., Oswald, E., Mainil, J. & Nagy, B. (2006). Serotypes and intimin types of intestinal and faecal strains of *eae*+ *Escherichia coli* from weaned pigs. *Vet Microbiol* **114**, 82–93.

Maquelin, K., Kirschner, C., Choo-Smith, L. P., Ngo-Thi, N. A., van Vreeswijk, T., Stammler, M., Endtz, H. P., Bruining, H. A., Naumann, D. & Puppels, G. J. (2003). Prospective study of the performance of vibrational spectroscopies for rapid identification of bacterial and fungal pathogens recovered from blood cultures. *J Clin Microbiol* **41**, 324–329.

Marolda, C. L., Feldman, M. F. & Valvano, M. A. (1999). Genetic organization of the O7-specific lipopolysaccharide biosynthesis cluster of *Escherichia coli* VW187 (O7 : K1). *Microbiology* **145**, 2485–2495.

Mercado, E. C., Gioffre, A., Rodriguez, S. M., Cataldi, A., Irino, K., Elizondo, A. M., Cipolla, A. L., Romano, M. I., Malena, R. & Mendez, M. A. (2004). Non-O157 Shiga toxin-producing *Escherichia coli* isolated from diarrhoeic calves in Argentina. *J Vet Med B Infect Dis Vet Public Health* **51**, 82–88.

Naumann, D. (2000). Infrared spectroscopy in microbiology. In *Encyclopedia of Analytical Chemistry*, pp. 102–131. Edited by R. A. Meyers. Chichester: Wiley.

Naumann, D., Helm, D. & Labischinski, H. (1991). Microbiological characterizations by FT-IR spectroscopy. *Nature* **351**, 81–82.

Orden, J. A., Ruiz-Santa-Quiteria, J. A., Cid, D., Garcia, S., Sanz, R. & de la Fuente, R. (1998). Verotoxin-producing *Escherichia coli* (VTEC) and *eae*-positive non-VTEC in 1-30-days-old diarrhoeic dairy calves. *Vet Microbiol* **63**, 239–248.

Orskov, F. (1952). Antigenic relationships between o groups 1-112 of *E. coli* and Wramby's o groups 26w-43w. 10 new o groups: 114-123. *Acta Pathol Microbiol Scand* **31**, 51–56.

Orskov, F. & Orskov, I. (1984). *Serotyping of Escherichia coli*. London: Academic Press.

Orskov, I. F., Orskov, F. B., Jann, B. & Jann, K. (1977). Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. *Bacteriol Rev* **41**, 667–710.

Ramachandran, V., Hornitzky, M. A., Bettelheim, K. A., Walker, M. J. & Djordjevic, S. P. (2001). The common ovine Shiga toxin 2-containing *Escherichia coli* serotypes and human isolates of the same serotypes possess a Stx2d toxin type. *J Clin Microbiol* **39**, 1932–1937.

Reeves, P. R. (1995). Role of O-antigen variation in the immune response. *Trends Microbiol* **3**, 381–386.

Reeves, P. R. & Wang, L. (2002). Genomic organization of LPS-specific loci. *Curr Top Microbiol Immunol* **264**, 109–135.

Tintelnot, K., Haase, G., Seibold, M., Bergmann, F., Staemmler, M., Franz, T. & Naumann, D. (2000). Evaluation of phenotypic markers for selection and identification of *Candida dubliniensis*. *J Clin Microbiol* **38**, 1599–1608.

Wang, L. & Reeves, P. R. (1998). Organization of *Escherichia coli* O157 O antigen gene cluster and identification of its specific genes. *Infect Immun* **66**, 3545–3551.

Wang, L. & Reeves, P. R. (2000). The *Escherichia coli* O111 and *Salmonella enterica* O35 gene clusters: gene clusters encoding the same colitose-containing O antigen are highly conserved. *J Bacteriol* **182**, 5256–5261.