Localization and molecular characterization of putative O antigen gene clusters of *Providencia* species

Olga G. Ovchinnikova,[1,2]† Bin Liu,[1]† Dan Guo,[1] Nina A. Kocharova,[2] Alexander S. Shashkov,[2] Miao Chen,[1] Lu Feng,[1,3] Antoni Rozalski,[4] Yuriy A. Knirel[2] and Lei Wang[1,3]

Correspondence
Lei Wang
wanglei@nankai.edu.cn
Yuriy A. Knirel
yknirel@gmail.com

[1]TEDA School of Biological Sciences and Biotechnology, Nankai University, 23 Hongda Street, TEDA, 300457 Tianjin, PR China

[2]N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Leninsky Prospekt 47, 119991 Moscow, Russia

[3]Tianjin Key Laboratory of Microbial Functional Genomics, Nankai University, 23 Hongda Street, TEDA, 300457 Tianjin, PR China

[4]Department of Immunobiology of Bacteria, Institute of Microbiology, Biotechnology and Immunology, University of Lodz, Banacha 12/16, 90-237 Lodz, Poland

Enterobacteria of the genus *Providencia* are opportunistic human pathogens associated with urinary tract and wound infections, as well as enteric diseases. The lipopolysaccharide (LPS) O antigen confers major antigenic variability upon the cell surface and is used for serotyping of Gram-negative bacteria. Recently, *Providencia* O antigen structures have been extensively studied, but no data on the location and organization of the O antigen gene cluster have been reported. In this study, the four *Providencia* genome sequences available were analysed, and the putative O antigen gene cluster was identified in the polymorphic locus between the *cpxA* and *yibK* genes. This finding provided the necessary information for designing primers, and cloning and sequencing the O antigen gene clusters from five more *Providencia alcalifaciens* strains. The gene functions predicted *in silico* were in agreement with the known O antigen structures; furthermore, annotation of the genes involved in the three-step synthesis of GDP-colitose (*gmd*, *colD* and *colC*) was supported by cloning and biochemical characterization of the corresponding enzymes. In one strain (*P. alcalifaciens* O39), no polysaccharide product of the gene cluster in the *cpxA–yibK* locus was found, and hence genes for synthesis of the existing O antigen are located elsewhere in the genome. In addition to the putative O antigen synthesis genes, homologues of *wza*, *wzb*, *wzc* and (in three strains) *wzi*, required for the surface expression of capsular polysaccharides, were found upstream of *yibK* in all species except *Providencia rustigianii*, suggesting that the LPS of these species may be attributed to the so-called K LPS (K_{LPS}). The data obtained open a way for development of a PCR-based typing method for identification of *Providencia* isolates.

†These authors contributed equally to this work.

INTRODUCTION

The genus *Providencia* of the family *Enterobacteriaceae* currently consists of eight species (Juneja & Lazzaro, 2009; Somvanshi *et al.*, 2006; O'Hara *et al.*, 2000), among which *Providencia stuartii*, *Providencia rettgeri*, *Providencia rustigianii* and *Providencia alcalifaciens* are the species that most commonly cause human infection. *P. stuartii* is often found in complicated urinary tract infections in patients with chronic indwelling urinary catheters (Rahav *et al.*, 1994). *Providencia* species, especially *P. rustigianii* and *P. alcalifaciens*, are an infrequent cause of traveller's diarrhea (Yoh

et al., 2005) and foodborne gastroenteritis (Murata *et al.*, 2001). *P. rettgeri* has been recognized as a pathogen involved in urinary tract infections (O'Hara *et al.*, 2000) and has also been reported to cause ocular infections (Koreishi *et al.*, 2006). Nosocomial infection due to *Providencia* represents an emerging problem because of the increasing prevalence of antibiotic resistance secondary to the presence of extended-spectrum beta-lactamase (Tumbarello *et al.*, 2004).

Lipopolysaccharide (LPS) is a key component of the outer membrane of the cell wall and is considered to be a virulence factor of Gram-negative bacteria, including *Providencia*. The full LPS molecule is composed of lipid A, which anchors the molecule to the outer membrane, core oligosaccharide and O polysaccharide (O antigen), which is built up of oligosaccharide repeats (O units). The O antigen is one of the most variable cell constituents and is a major determinant of immunospecificity.

The combined serological classification scheme of *P. stuartii*, *P. rustigianii* and *P. alcalifaciens* used in epidemiology is based on O antigens and flagella H antigens; it includes 63 O-serogroups and 30 H-serogroups (Ewing, 1986). Recently, it has been found that strains representing serogroups O58 : H9 and O59 : H18 do not belong to the genus *Providencia* and must be reclassified to *Morganella morganii* (A. Rozalski, unpublished data). An early report described five K antigens in strains belonging to serogroups O14 (two strains), O15, O25 and O46 (Ewing *et al.*, 1954), but no later studies have been undertaken to confirm the presence of a capsule in *P. alcalifaciens*, *P. rustigianii* and *P. stuartii*. A separate serological classification scheme based on 34 O antigens, 26 H antigens and one K antigen has been proposed for *P. rettgeri* (Ewing, 1986).

Serological relationships are observed between different *Providencia* strains as well as between strains of *Providencia* and *Proteus*, *Morganella*, *Escherichia coli*, *Salmonella enterica* and *Shigella*. To establish the uniqueness of a particular O-serogroup or to identify the presence of common epitopes, more than 30 new O polysaccharide structures of *Providencia* O-serogroups have been determined (Knirel, 2011). Most O polysaccharides are acidic heteropolymers with an O unit ranging from a di- to a heptasaccharide and containing various unusual monosaccharides and non-sugar components.

The genes involved in O antigen biosynthesis are normally located in a gene cluster that maps between two conserved genes in the chromosome. The diversity of O polysaccharide structures is mainly due to genetic variations in the O antigen gene cluster (OGC). Genes within the OGC encode proteins that mostly fall into three major groups. Proteins of the first group are involved in the biosynthesis, which proceeds in the cytosol, of nucleotide sugar precursors specific to the O antigen. Enzymes of the second group, namely glycosyltransferases, sequentially transfer the precursor sugars to form an O unit or a polysaccharide on a

carrier lipid, undecaprenyl diphosphate (UndPP), which is anchored into the inner membrane facing the cytoplasmic side. The third group consists of integral membrane proteins for O antigen processing/assembly. There are four distinct processes for synthesis and translocation of the O antigen, i.e. the Wzx/Wzy-, ATP-binding cassette (ABC) transporter-, synthase- and Wzk-dependent processes (Valvano *et al.*, 2011). Among them, the first pathway occurs in the synthesis of the majority of O antigens, especially in heteropolymeric O antigens. In the Wzx/Wzy-dependent process, once the assembly of the UndPP-linked O units is completed, they are thought to be 'flipped' across the inner membrane by a flippase (Wzx), where they subsequently become polymerized into a long O polysaccharide chain by O antigen polymerase (Wzy).

The OGC is located at different genomic loci in different species. It occurs between the housekeeping genes *galF* and *gnd* in a number of *Enterobacteriaceae*, including *E. coli*, *S. enterica* (Reeves, 1994) and *Enterobacter* (now *Cronobacter*) *sakazakii* (Mullane *et al.*, 2008). In *Pseudomonas aeruginosa*, the OGC is located between *himD* and *tyrB* (Raymond *et al.*, 2002), in *Vibrio cholerae* between *gmhD* and *rjg* (Sozhamannan *et al.*, 1999), in *Yersinia* spp. between *hemH* and *gsk* (Pacinelli *et al.*, 2002; Zhang *et al.*, 1996), and in *Proteus mirabilis* between *cpxA* and *secB* (Wang *et al.*, 2010). However, in *Providencia*, no OGC has been found and characterized so far.

In this study, we revealed a polymorphic locus tentatively associated with O antigen biosynthesis in four *Providencia* species (*P. alcalifaciens*, *P. stuartii*, *P. rustigianii* and *P. rettgeri*). The locus also contained genes required for the surface expression of capsular polysaccharide (CPS). *In silico*-predicted functions of genes involved in the three-step synthesis of GDP-colitose were supported by cloning and biochemical characterization of their products. A relationship between the gene clusters of the *Providencia* O antigens and *E. coli* K antigens is discussed.

METHODS

Construction of a random shotgun bank. Chromosomal DNA was prepared as described by Bastin & Reeves (1995). Primers wl-35627 (5′-CAA TTT TCT GGT TTA CCC TCG CAC T-3′) and wl-35631 (5′-TCT GGA CCA AAC ATT AAA TAA TCA TCT T-3′), based on the *cpxA* and *yibK* genes, respectively, were used to amplify the OGCs in *P. alcalifaciens* O19, O28, O36, O39 and O44 with the Expand Long Template PCR system (TaKaRa Biotechnology). Each PCR cycle consisted of denaturation at 95 °C for 30 s, annealing at 55 °C for 45 s and extension at 68 °C for 15 min. The PCR products were sheared at speed code 8 (20 cycles) to the desired molecular mass of 1000–2000 kb using a HydroShear apparatus (GeneMachines). The resulting DNA fragments were cloned into vector pUC18 to produce a shotgun bank. The OGC of *P. alcalifaciens* O6 was sequenced by primer walking.

Sequencing and analysis. Sequencing was carried out with an ABI 3730 automated DNA sequencer by the Tianjin Biochip Corporation. The Staden package (Staden, 1996) and the program Artemis (Rutherford *et al.*, 2000) were used for sequence assembly and gene

annotation, respectively. CD-Search (Marchler-Bauer & Bryant, 2004) was performed to search for conserved motifs. BLAST (Altschul *et al.*, 1997) was used to search databases, including the GenBank and Pfam protein motif databases (Bateman *et al.*, 2002), for possible gene functions. The program TMHMM 2.0 was used for identification of potential transmembrane segments (http://www.cbs.dtu.dk/services/TMHMM/).

Cloning and plasmid construction. The *gmd*, *colD* and *colC* from *P. alcalifaciens* O39 were amplified by PCR using the primer pairs wl-39915 (5′-CAT GCC ATG GCC ATG AAA ATA GCA TTA ATCA CA-3′)/wl-39916 (5′-CCG CTC GAG TAA CTC TAA TGA TAA AGA AAC ATC-3′), wl-39923 (5′-CAT GCC ATG GAA ATG GTC TCC AAA AAC TTC T-3′)/wl-39924 (5′-CCG CTC GAG ATT AGA CAG TAG TGA TTG TAT AAT A-3′) and wl-39919 (5′-CAT GCC ATG GCC ATG AAG TTA TTT ATT ACC GGC-3′)/wl-39920 (5′-CCG CTC GAG TAA GAC CTC TGT TAG GTA AAA AT-3′), respectively (restriction sites underlined). A total of 30 cycles were performed using conditions as follows: denaturation at 95 °C for 30 s, annealing at 45 °C for 45 s and extension at 72 °C for 1 min, in a final volume of 50 µl. The PCR products were digested with *NcoI*/*XhoI* and ligated into pET28a⁺ to construct pLW1593, pLW1595 and pLW1594 (containing Gmd, ColD and ColC, respectively). The presence of the inserts in plasmids was confirmed by sequencing using an ABI 3730 automated DNA sequencer.

Protein expression and purification. *E. coli* BL21 (DE3) carrying pLW1593, pLW1594 or pLW1595 was grown in Luria–Bertani (LB) medium containing 50 µg kanamycin ml⁻¹ overnight at 37 °C with shaking. The overnight culture was inoculated into 500 ml fresh medium and grown to OD_{600} 0.6. Expression of Gmd and ColC was induced by 0.1 mM IPTG at 30 °C for 4 h, and expression of ColD was induced by 0.1 mM IPTG at 18 °C for 12 h. The His_6-tagged fusion proteins were purified by nickel ion affinity chromatography with a Chelating Sepharose Fast Flow column (GE Healthcare), essentially as described by Ren *et al.* (2010). Eluted proteins were dialysed overnight (50 mM Tris-HCl, pH 8.0, 50 mM EDTA, pH 8.0, 1 % glycine, 10 %, v/v, glycerol) at 4 °C. The purity of proteins was checked by Coomassie blue-stained SDS-PAGE analysis on a 12 % gel. Protein concentration was determined by the Bradford method. Purified proteins were stored at −40 °C.

Enzyme activity assays. The reaction mixture (20 µl) for purified Gmd (2 µg) contained 0.5 mM GDP-D-Man and 50 mM Tris-HCl buffer (pH 7.5). To assay the ColD activity, the above mixture was supplemented with 50 mM L-glutamate and 2 µg purified ColD (total volume 20 µl). To assay the ColC activity, the Gmd/ColD reaction mixture (20 µl) was amended with 4 mM NADH, 50 mM Tris-HCl (pH 7.5) and 2 µg purified ColC (total volume 40 µl). Reactions were carried out at 37 °C for 20 min before being terminated by adding an equal volume of chloroform.

For NMR monitoring (McNally *et al.*, 2006) of the reaction with Gmd, the reaction mixture (250 µl) containing 4 mM GDP-D-Man, 20 µg purified Gmd and 50 mM Tris-HCl buffer (pH 7.5) in a 9:1 H_2O/D_2O mixture was prepared in a Shigemi (Japan) microtube, and the ¹H NMR spectra were run at 10 min intervals as described below.

HPLC and electrospray ionization-MS (ESI-MS) analysis. Reactions were monitored by HPLC using a Shimadzu LC-20A system with a Venusil MP C18 column (5 µm particle size, 4.6 mm × 250 mm) (Agela Technologies). The reaction products were purified by HPLC on the same column using 4 % acetonitrile and 96 % 50 mM triethylammonium acetate (pH 6.8) as the mobile phase at a flow rate of 0.6 ml min⁻¹ and were monitored by UV detection at 254 nm. Fractions containing the expected products were collected, lyophilized and redissolved in water before injecting into a

Finnigan LCQ Advantage MAX ion trap mass spectrometer (Thermo Electron) in negative mode (4.5 kV, 250 °C) for ESI-MS analysis. For MS2 and MS3 analyses, nitrogen was used as the collision gas and helium as the auxiliary gas, and the collision energy was typically set to 20–30 eV.

NMR spectroscopy. ¹H, ¹³C and ³¹P NMR spectra were recorded at 30 °C on a Bruker Avance II 600 spectrometer (Germany). Chemical shifts were referenced to internal sodium 3-trimethylsilylpropanoate-2,2,3,3-d_4 (δ_H 0.00), external acetone (δ_C 31.45) and external aqueous 85 % H_3PO_4 (δ_P 0). 2D NMR spectra were obtained using standard Bruker software, and the Bruker TOPSPIN 2.1 program was employed to acquire and process the NMR data. A mixing time of 100 ms was used in the total correlation spectroscopy (TOCSY) experiment.

A sample of GDP-Col was deuterium-exchanged by freeze-drying from 99.9 % D_2O and measured as a solution in 99.96 % D_2O. NMR monitoring of the Gmd reaction was performed in a 9:1 H_2O/D_2O mixture.

RESULTS

Sequence analysis of putative O antigen gene clusters of *Providencia* strains

Analysis of two draft genome sequences of *P. alcalifaciens* DSM 30120 (serogroup O19) and *P. stuartii* ATCC 25827 (serogroup O47) (GenBank accession nos ABXW00000000 and ABJD00000000, respectively) suggested that the OGC may be located between *cpxA* and *yibK*. These two genes are conserved housekeeping genes encoding a two-component system sensor kinase and tRNA/rRNA methyltransferase, respectively. Then, the *cpxA*–*yibK* region was analysed in two other available *Providencia* genomes (*P. rustigianii* DSM 4541 and *P. rettgeri* DSM 1131; GenBank accession nos ABXV00000000 and ACCI00000000, respectively) and sequenced from five more *P. alcalifaciens* strains belonging to serogroups O6, O28, O36, O39 and O44.

Fig. 1 shows the organization of the putative OGCs of eight strains studied and the known O antigen structures of seven strains from serogroups O6 (Ovchinnikova *et al.*, 2007), O19 (Kocharova *et al.*, 2004, 2008b), O28 (Ovchinnikova *et al.*, 2011a), O36 (Kocharova *et al.*, 2007), O39 (our unpublished data), O44 (Kocharova *et al.*, 2005) and O47 (Ovchinnikova *et al.*, 2004). The OGCs vary in size and contain 10–21 ORFs, all of which have the same transcriptional direction from *cpxA* to *yibK*.

The ORFs were assigned putative functions using BLASTP, based on similarities to LPS and sugar biosynthetic genes in other bacteria. Each OGC was found to comprise a specific set of sugar biosynthetic pathway genes, glycosyltransferase genes and O antigen assembly genes. From seven strains studied with known O antigen structures, the *in silico*-predicted gene functions were in agreement with the structures for six strains, whereas strain O39 had essentially the same OGC as strain O6 (sharing >99 % overall DNA identity) but a different O antigen structure (Fig. 1). In addition, some other genes which are probably involved in

biosynthesis of CPS (K antigen) rather than O antigen, were identified within the OGCs.

The genes within the *cpxA–yibK* region were annotated (Fig. 1, Supplementary Table S1), and their putative functions are described below. The organization of the OGCs of *P. rustigianii* and *P. rettgeri* are shown in Fig. 1 to demonstrate the heterogeneity of the locus, but they are not included in the analysis, since their O antigen structures remain unknown.

Sugar biosynthetic pathway genes

The nucleotide precursors of common sugars, such as UDP-D-Glc, UDP-D-Gal and UDP-D-GlcNAc, are a part of other pathways generally present in bacteria, and usually the genes involved in their biosynthesis are not within the OGC. However, in *Providencia*, a gene with high-level homology to *galE*, responsible for the conversion of UDP-D-Glc to UDP-D-Gal, was found at the 3′ end of the *cpxA–yibK* region. Strikingly, this gene occurred in all strains, whether galactose was an O antigen component or not. In the sequenced strains, a putative *galE* gene was also found outside the *cpxA–yibK* region, and the reason for its reduplication in the OGC is unclear.

Homologues of the *gne* and *ugd* genes were present within the *cpxA–yibK* region of *Providencia* strains whose O antigens contain GalNAc (O44) and GlcA (O6, O28, O44 and O47), respectively. *gne* encodes UDP-*N*-acetylglucosamine 4-epimerase, capable of synthesizing UDP-D-GalNAc from UDP-D-GlcNAc, and *ugd* is the gene for UDP-glucose 6-dehydrogenase, catalysing conversion of UDP-D-Glc to UDP-D-GlcA.

The O antigens of *P. alcalifaciens* O19, *P. stuartii* O47 and *P. alcalifaciens* O36 contain 3-acetamido-3,6-dideoxy-D-galactose (D-Fuc3NAc), L-rhamnose (L-Rha) and 6-deoxy-L-talose (L-6dTal), respectively, which are synthesized by a combined pathway. It starts with glucose 1-phosphate, which is activated by the glucose-1-phosphate thymidylyltransferase RmlA to form dTDP-D-glucose, and dehydrated by the dTDP-D-glucose 4,6-dehydratase RmlB to form a common intermediate, dTDP-6-deoxy-D-*xylo*-hexos-4-ulose. As expected, the full set of *rmlABCD* genes required for synthesis of dTDP-L-Rha (Köplin *et al.*, 1993) was found in the O47 OGC. In addition to *rmlA*, the O19 OGC contains the *fdtABC* genes for isomerization of dTDP-6-deoxy-D-*xylo*-hexos-4-ulose to dTDP-6-deoxy-D-*xylo*-hexos-3-ulose, followed by its transformation into dTDP-3-amino-3,6-dideoxy-D-*xylo*-hexose and acetylation to dTDP-D-Fuc3NAc (Pfoestl *et al.*, 2003). The O36 OGC contains *rmlA*, *rmlC* (for dTDP-6-deoxy-D-*xylo*-hexos-4-ulose 3,5-isomerase) and *tll* (for dTDP-6-deoxy-L-*lyxo*-hexos-4-ulose reductase), which are responsible for synthesis of dTDP-L-6dTal (Nakano *et al.*, 2000). However, no gene within the OGCs of O19 and O36 strains showed homology with *rmlB*, suggesting that this gene is located outside the *cpxA–yibK* region. Indeed, analysis of the available *Providencia* genomes

revealed homologues of *rmlA* (GenBank accession nos EEB46719, EDU58095, EFB72495, EFE52173) and *rmlB* (GenBank accession nos EEB46720, EDU58094, EFB72496, EFE52174) within the enterobacterial common antigen (ECA) gene cluster, as these gene products are involved in the early steps of the synthesis of 4-acetamido-4,6-dideoxygalactose (D-Fuc4Nac), a component of ECA (Erbel *et al.*, 2003). Therefore, *rmlA* is always reduplicated in the OGC, whereas *rmlB* is sometimes reduplicated and sometimes not.

manA (type I phosphomannose isomerase), *manB* (phosphomannomutase) and *manC* (GDP-mannose pyrophosphorylase) are required for biosynthesis of GDP-D-Man, the precursor of the D-Man present in the O47 antigen. In the *Enterobacteriaceae*, *manA* is generally present as part of the mannose (Man) metabolism pathway, whereas *manB* and *manC* are specific to GDP-D-Man synthesis (Samuel & Reeves, 2003), and accordingly both were found within the O47 OGC.

The O44 antigen contains two 6-deoxy sugars, namely L-fucose (L-Fuc) and 6-deoxy-L-glucose (L-quinovose; L-Qui), and L-Fuc is also present in the O28 antigen. The biosynthesis of GDP-L-Fuc requires the *manBC* genes, *gmd* for GDP-D-mannose 4,6-dehydratase, and *fcl* for bifunctional GDP-6-deoxy-D-*lyxo*-hexos-4-ulose 3,5-epimerase/4-reductase (Andrianopoulos *et al.*, 1998; Wu *et al.*, 2001), all of which were found within the OGCs of both strain O28 and strain O44. The biosynthesis of L-Qui in bacteria has not been characterized so far, but one can propose a pathway similar to that of L-Fuc. Therefore, one more enzyme, such as bifunctional GDP-6-deoxy-D-*lyxo*-hexos-4-ulose 3,5-epimerase/4-reductase (alternatively two enzymes with separate functions) or GDP-L-Fuc 4-epimerase, should be encoded within the OGC. However, no such gene was found in the fully annotated *cpxA–yibK* region of strain O44, and the origin of the L-Qui precursor remains obscure.

The O6 antigen contains another unusual component, 3,6-dideoxy-L-*xylo*-hexose (colitose; Col). The GDP-Col pathway requires five genes, *manB*, *manC*, *gmd*, *colD* and *colC* (Fig. 2) (Beyer *et al.*, 2003; Alam *et al.*, 2004), all of which were found in the O6 OGC. *colD* encodes pyridoxal 5′-phosphate (PLP)-dependent GDP-6-deoxy-D-*lyxo*-hexos-4-ulose 3-deoxygenase, which catalyses the C-3 deoxygenation of the Gmd product in the presence of L-glutamate. The resultant GDP-3,6-dideoxy-D-*threo*-hexos-4-ulose is converted into GDP-Col by bifunctional GDP-3,6-dideoxy-D-*threo*-hexos-4-ulose 5-epimerase/4-reductase (ColC).

In *E. coli* and *S. enterica*, the OGCs that contain *gmd* generally also have the *gmm* gene, encoding GDP-D-Man mannosyl hydrolase. Gmm removes GDP-D-Man from the sugar biosynthesis pathway and is thought to regulate the synthesis of GDP-sugars derived from GDP-D-Man (Samuel & Reeves, 2003). In the OGCs of *P. alcalifaciens* O6 and O44, *gmd* was accompanied by *gmm* as well.

P. alcalifaciens O19 →2)-β-Fuc3NAc4Ac-(1→3)-α-GlcNAc-(1→4)-α-Gal-(1→4)-β-Gal-(1→3)-β-GlcNAc-(1→
(*S*)-pyr-(2–4,6)⌐

dTDP-Fuc3NAc OAc   pyruvate   UDP-Gal

cpxA   rmlA $\frac{A\ C}{fdt}\ B$ AT wzx GT PT GT wzy GT GT galE wza wzb wzc yibK

P. stuartii O47 →2)-β-Gal-(1→4)-β-Man6Ac-(1→3)-β-Man-(1→4)-β-GlcA-(1→3)-α-GlcNAc-(1→
α-L-Rha-(1→3)⌐

dTDP-L-Rha   GDP-Man UDP-GlcA   UDP-Gal

cpxA   wzi $\frac{B\ \ \ D\ \ \ A\ \ C}{rml}$ GT wzy GT wzx GT GT manC manB ugd GT wza wzb wzc galE yibK

P. rustigianii DSM 454

cpxA   ansB tagD wzx tagF GT GT wzy GT gne galE yibK

P. rettgeri DSM 1131

cpxA   wzi   rmlB rmlA $\frac{A\ C}{fdt}\ B$ $\frac{D\ C}{rml}$ GT wzy GT GT wza wzb wzc galE tagF GT PPDK GAT CCT wzx yibK

P. alcalifaciens O28 →3)-β-GlcNAc-(1→3)-α-L-Fuc-(1→3)-β-GlcNAc-(1→
α-GlcA-(1→3)-α-L-Fuc-(1→4)⌐

GDP-L-Fuc   UDP-GlcA   UDP-Gal

cpxA   GT wzy GT GT gmd fcl gmm manC GT manB ugd wzx galE wza wzb wzc yibK

P. alcalifaciens O36 →7)-β-Kdo-(2→3)-α-L-6dTal-(1→3)-α-GlcNAc-(1→
AcO-2⌐ (~70 %)

dTDP-L-6dTal   OAc   CMP-Kdo   UDP-Gal

cpxA $\frac{A\ \ \ \ C}{rml}$ tll GT GT AT wzx wzy $\frac{A\ \ \ B\ \ \ D}{kds}$ yibK galE wza wzb wzc yibK

P. alcalifaciens O44 →4)-α-GalNAc-(1→3)-α-L-Fuc-(1→3)-α-Glc-(1→4)-α-L-Qui-(1→3)-α-GlcNAc-(1→
β-GlcA-(1→4)⌐

UDP-GalNAc   GDP-L-Fuc, GDP-L-Qui? UDP-GlcA   UDP-Gal

cpxA   wzi   int   wzx GT GT wzy GT GT gne gmd fcl gmm manC GT manB ugd wza wzb wzc galE yibK

P. alcalifaciens O6 →4)-β-GlcA-(1→3)-β-GlcNAc-(1→
α-Col-(1→2)-β-Gal-(1→3)-β-GlcNAc-(1→6)⌐

GDP-Col   UDP-GlcA UDP-Gal

cpxA   wzx GT gmd colC colD gmm manC manB GT wzy GT ugd GT galE wza wzb wzc yibK

P. alcalifaciens O39 →4)-α-GalNAc-(1→6)-α-Glc-(1→4)-β-GlcA-(1→3)-β-GalNAc-(1→
β-Qui4N(alaAla)-(1→6)⌐

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 kb

Fig. 1. Organization of the putative *Providencia* OGCs, together with the O antigen structures. The OGCs reside between the housekeeping genes *cpxA* and *yibK* (hatched). Transferase genes are indicated by upper-case type: *GT*, glycosyltransferase; *AT*, acetyltransferase; *PT*, pyruvyltransferase; *GAT*, glutamine amidotransferase; *CCT*, CTP:phosphocholine cytidylyltransferase. *orf18* in the OGC of *P. rettgeri* encodes a cluster protein with glycosyltransferase and pyruvate phosphate dikinase (*PPDK*) predicted functions. Shaded areas represent intergenic regions. In the O antigen structures, all monosaccharides are in the pyranose form and have the D configuration, unless stated otherwise. alaAla, *N*-(1-carboxyethyl)alanine (alanopine); (*S*)-pyr, pyruvic acid (*S*)-acetal.

A distinguishing feature of the O36 antigen is the presence of 3-deoxy-D-*manno*-oct-2-ulosonic acid (Kdo), which is an essential constituent of the LPS core, providing the linkage between lipid A and the carbohydrate moiety of the LPS, but is a rare component of O antigens (Knirel *et al.*, 2011). Accordingly, the O36 OGC contains predicted *kdsD* (arabinose 5-phosphate isomerase), *kdsA* (Kdo 8-phosphate synthase) and *kdsB* (Kdo cytidylyltransferase) genes, which are responsible for the synthesis of CMP-Kdo from D-ribulose 5-phosphate and phosphoenolpyruvate (Gronow & Brade, 2001; Raetz & Whitfield, 2002; Schnaitman & Klena, 1993).

Sugar transferase genes

In most *Enterobacteriaceae* members studied, the first sugar of the O unit is either GlcNAc or GalNAc (Katzenellenbogen *et al.*, 2005; Carlin *et al.*, 1984; Stenutz *et al.*, 2006). Four *Providencia* O antigens (O9, O14, O19, O34) have been studied in this respect and have also been found to employ one of these monosaccharides as the first sugar (Kocharova *et al.*, 2008a, b; Ovchinnikova *et al.*, 2011b; Kondakova *et al.*, 2007). Furthermore, GlcNAc or GalNAc is present in the O antigens of most other *Providencia* serogroups, including all those shown in Fig. 1, and can be suggested to occupy the same position in their O units.

O unit synthesis is initiated by transfer of the first sugar 1-phosphate from the UDP precursor to the lipid acceptor undecaprenyl phosphate (UndP), catalysed by WecA (Valvano *et al.*, 2011). In *Enterobacteriaceae*, the *wecA* gene is located in the ECA gene cluster rather than the OGC (Valvano, 2003). In the available draft *Providencia* genome, homologues of *wecA* (GenBank accession nos EEB46724, EDU58090, EFB72500 and EFE52178) were also found in the ECA gene cluster and were not reduplicated in the OGC.

Other glycosyltransferases are specific for sugar donors, sugar acceptors, and the type of linkage between them. To synthesize UndPP-linked tri-, penta- and hexasaccharide O units, two, four and five individual glycosyltransferases are expected, respectively. The corresponding number of glycosyltransferase genes was found in the putative OGCs of *Providencia* strains with a known O antigen structure (Fig. 1). The occurrence of five annotated glycosyltransferase genes in the *P. alcalifaciens* O19 genome sequence appeared to be at variance with the pentasaccharide O unit. However, the sequencing of the O19 OGC between *cpxA* and *yibK* revealed that two predicted proteins (GenBank accession nos

EEB44290 and EEB44291) are encoded by one ORF, which had been broken, probably, as a result of low sequencing accuracy (similarly, the proteins with the GenBank accession nos EEB44288 and EEB44286 are parts of one protein). Hence, the O19 OGC in fact contains four glycosyltransferase genes, which is in agreement with the O unit structure.

Orf5 in the O36 OGC shows little homology to any glycosyltransferase, but 28 % identity and 46 % similarity to the CPS biosynthesis protein of *Burkholderia phytofirmans*, and also a similar percentage identity to KpsS of a number of other bacteria. Based on a putative involvement of KpsS in a diacylglycerophosphate-Kdo modification of CPS (see Discussion), we suggest that Orf5 is Kdo transferase.

O antigen processing genes

Wzx and Wzy are highly hydrophobic membrane proteins, usually sharing low sequence identities with their homologues. The *wzx* and *wzy* genes were found in all eight *Providencia* OGCs studied. The topology models of the putative Wzx proteins derived from *in silico* prediction contain 10–14 well-proportioned transmembrane segments, which is typical of O unit flippases (Valvano *et al.*, 2011). The putative Wzy proteins have 9–12 predicted transmembrane segments, with a large periplasmic loop of 36–88 amino acid residues. One or two large periplasmic loops are observed for a number of O antigen polymerases, and seem to be important in the recognition of the O unit and/or for catalytic activity (Daniels *et al.*, 1998; Islam *et al.*, 2010, 2011). Hence, the Wzx/Wzy-dependent process is suggested to mediate the synthesis of all *Providencia* O antigens studied.

The third component of the Wzy/Wzx-dependent processing is the integral inner-membrane protein Wzz, which modulates O antigen chain-length distribution. Usually, it maps outside the OGC. In the four *Providencia* genomes sequenced, a homologue of the *wzz* (*fepE*) gene (GenBank accession nos EEB45532, EDU60377, EFB71782 and EFE53677) was found within a cluster of genes associated with ferric enterochelin synthesis and transport.

Other genes

Putative *O*-acetyltransferase genes were found in the OGCs of *P. alcalifaciens* O19 and O36, and are suggested to be responsible for the 4-*O*-acetylation of UDP-D-Fuc3NAc and 2-*O*-acetylation of dTDP-L-6dTal, respectively. The
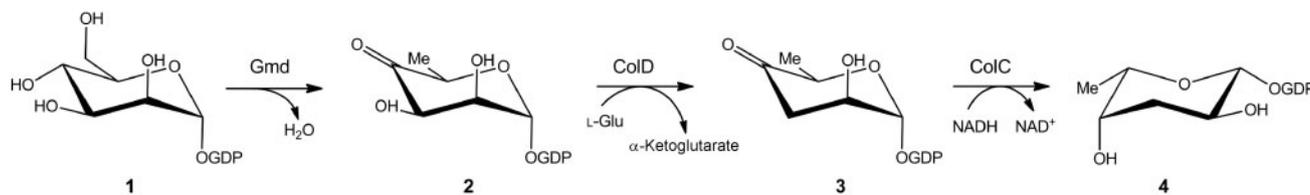
Fig. 2. Biosynthetic pathway of GDP-Col. Numbers refer to the compounds listed in the legend to Fig. 3.

O19 OGC also contains a homologue of pyruvyltransferase, which is consistent with the presence of a pyruvic acid acetal on a GlcN residue in the O19 antigen.

In addition to the full-length flanking *yibK*, the O36 OGC contains a defective *yibK* homologue, which lacked the 5′ end due to a nonsense mutation. A homologue of the *int* gene encoding a putative integrase was found in the O44 OGC.

Another common feature of the putative OGCs of the *Providencia* species studied, except for *P. rustigianii*, is the presence of a conserved block of homologues of the *wza*, *wzb* and *wzc* genes, which are generally found in *E. coli* group 1 CPS synthesis clusters (Whitfield, 2006). The *wzc* encodes an integral inner-membrane tyrosine autokinase; transphosphorylation of Wzc and its dephosphorylation by the cognate low-molecular-mass phosphatase Wzb are essential for the maintenance of both polymerization and CPS export. Wzc forms a complex with the outer-membrane protein Wza; the latter acts as a channel through which CPS is translocated to the surface. In contrast to O antigen processing, the export of CPS is coupled to Wzx/Wzy-dependent polymerization, and mutations in the *wza* or *wzc* gene result in similar acapsular phenotypes of mutants (Collins *et al.*, 2007).

In the OGCs of three strains (O44, O47 and *P. rettgeri* DSM 1131), a *wzi* gene homologue was found downstream next to *cpxA*. Wzi is an outer-membrane protein unique to group 1 capsules and is involved in modulation of surface attachment of the capsule (Rahn *et al.*, 2003; Whitfield, 2006). In *P. alcalifaciens* O19 and *P. rustigianii*, no gene showed homology with *wzi* either in the OGC or elsewhere in the genome.

Functional characterization of GDP-Col biosynthesis genes

As mentioned above, *P. alcalifaciens* O6 and O39 have essentially identical OGCs, which include a set of genes for synthesis of GDP-Col from GDP-Man, namely *gmd*, *colD* and *colC* (Fig. 2). Sequence alignment revealed that ColC and ColD from *P. alcalifaciens* O6 and O39 have 100 % protein identity, and that Gmd from two strains only has a single amino acid residue substitution. To confirm the functions of these genes, the three encoded proteins from *P. alcalifaciens* O39 were overexpressed in *E. coli* BL21 in a

soluble form with a C-terminal His$_6$-tag and purified by nickel affinity chromatography (Supplementary Fig. S1).

The 4,6-dehydratase activity of Gmd was verified by incubation of GDP-D-Man and Gmd followed by HPLC analysis. The initial substrate GDP-D-Man (**1**) (Fig. 3a) was converted completely to a novel product, **2**, migrating as a broad peak (Fig. 3b). The negative-mode ESI mass spectrum of **2** (Supplementary Fig. S2a) showed an $[M-H]^-$ ion peak at *m/z* 586 and thus a loss of 18 atomic mass units (amu) compared with **1**, whose $[M-H]^-$ ion peak was detected at *m/z* 604. By performing MS/MS on the ion peak at *m/z* 586, we observed a peak at *m/z* 303 that could be attributed to 6-deoxyhexos-4-ulose 1-diphosphate.

^{1}H NMR spectroscopy was used to monitor the Gmd reaction directly in an NMR tube and for identification of **2** in the reaction mixture (Fig. 4a, b). A decrease in content of **1** was observed within the first few minutes, and was accompanied by the appearance of GDP-6-deoxy-D-*lyxo*-hexos-4-ulose (**2**) and a minor product, **2′**, in a 3 : 1 ratio. The latter showed essentially the same correlation pattern in the 2D NMR spectra, but all ^{1}H NMR signals were shifted upfield as compared with **2**. Therefore, it is suggested that compound **2′** is a hydrated form of **2** (Supplementary Table S2). This conclusion and our assignment of the ^{1}H NMR chemical shifts of **2′** are at variance with the data of Ramm and co-workers, who assigned to **2′** the structure of a 3-keto isomer of **2** (Ramm *et al.*, 2004).

Adding ColD and L-glutamate to the reaction mixture containing **2** and **2′** resulted in the formation of a new product, **3**, also eluting as a broad peak, suggesting the conservation of the keto group (Fig. 3c). The function of ColD requires L-glutamate and PLP (Beyer *et al.*, 2003; Alam *et al.*, 2004), and indeed no conversion of **2** was observed in the absence of L-glutamate. However, exogenous PLP was not necessary for formation of **3**, suggesting that ColD was purified as a complex with endogenous PLP. This was confirmed by the UV absorbance at 326 nm, which also indicated that the saturation of ColD with the cofactor was ~10 %. ESI-MS analysis of **3** showed an $[M-H]^-$ ion peak at *m/z* 570 attributed to GDP-3,6-dideoxyhexos-4-ulose. MS/MS fragmentation of this ion, in addition to the GDP moiety fragments, yielded an ion peak at *m/z* 287, which corresponds to 3,6-dideoxyhexos-4-ulose 1-diphosphate (Supplementary Fig. S2b).
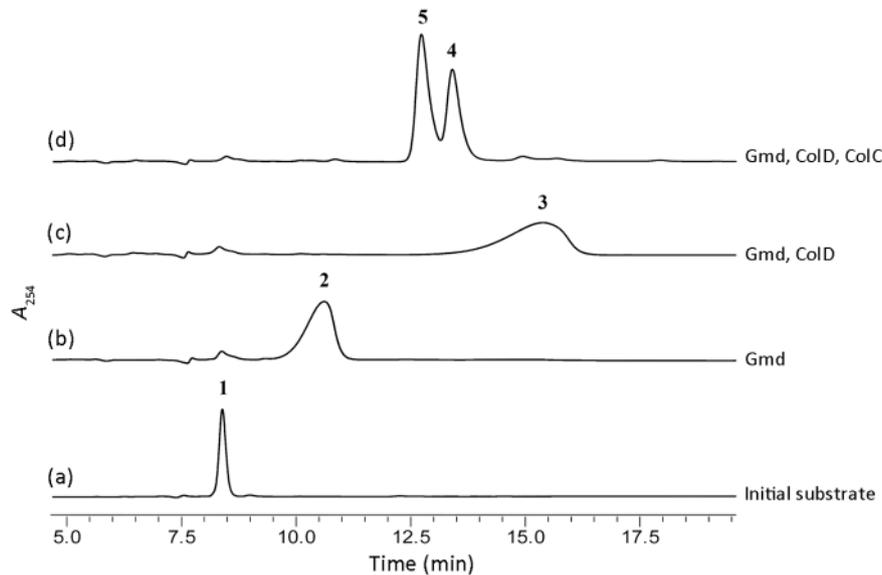
Fig. 3. HPLC analysis of the initial substrate (a) and the enzymatic reaction products (b–d) catalysed by the proteins indicated on the right-hand side. **1**, GDP-D-Man; **2**, GDP-6-deoxy-D-*lyxo*-hexos-4-ulose; **3**, GDP-3,6-dideoxy-D-*threo*-hexos-4-ulose; **4**, GDP-Col; **5**, NAD$^+$. NADH eluted at 41.5 min (not shown).

Adding ColC and NADH to the reaction mixture resulted in formation of a new product, **4**, accompanied by the conversion of NADH to NAD$^+$ (Fig. 3d). The ESI-MS spectrum of **4** contained an $[M-H]^-$ ion peak at *m/z* 572, which exhibited a gain of 2 amu compared with **3** and thus indicated the reduction of the 4-keto group. MS/MS analysis of **4** revealed a fragment ion peak at *m/z* 289 for 3,6-dideoxyhexose 1-diphosphate (Supplementary Fig. S2c). The product **4** was purified by HPLC and characterized by ^{1}H, ^{13}C and ^{31}P NMR spectroscopy (Supplementary Table S3, Fig. 4c and Supplementary Fig. S3), which confirmed the identity of **4** as GDP-Col.

Therefore, the products of *gmd*, *colD* and *colC* possess the predicted activities.

DISCUSSION

Among the closely related genera *Providencia*, *Proteus* and *Morganella*, the OGC has been characterized only in *Proteus*, where it maps between the *cpxA* and *secB* genes (Wang *et al.*, 2010). The present paper describes, for the first time to our knowledge, the genetic organization of putative O antigen loci retrieved from available databases for four strains of various *Providencia* species, and sequences from four more strains of *P. alcalifaciens*. The data obtained demonstrate that in all strains the OGC occupies the same locus downstream of *cpxA*. Homologues of serine *O*-acetyltransferase, glycerol-3-phosphate dehydrogenase, and preprotein translocase subunit (*secB*) genes were found downstream of *yibK* in both *Proteus* and *Providencia* genomes, and are not likely associated with O antigen biosynthesis. Therefore, we suggest that in *Providencia*, the OGC is flanked by *yibK* rather than *secB*.

The overall G+C mol% of the OGC of the *Providencia* strains examined ranged from 31 to 36, which is lower than the mean genome G+C content of approximately 40 mol% for *P. rettgeri* and 41 mol% for the other three species. Such variation in G+C content is typical for OGCs and some other polysaccharide gene clusters of a number of bacteria. The simplest explanation for this phenomenon is that the clusters originated in another species and were then captured by lateral transfer, but it remains unclear why the G+C content of the clusters is almost always lower than the rest of the genome (Cunneen & Reeves, 2011).

The consistency of the OGC of *Providencia* was confirmed by: (i) the serogroup-specific polymorphism of the *cpxA–yibK* locus demonstrated in six strains; (ii) the absence of another gene cluster, which may be associated with synthesis of O antigen or another polysaccharide different from ECA, as demonstrated by analysis of the full genome sequences of *P. alcalifaciens* O19 and *P. stuartii* O47; (iii) the correspondence of the gene functions inferred by *in silico* analysis with the known O antigen structures; (iv) a biochemical confirmation of the functions of the genes in a three-step sugar (Col) biosynthetic pathway.

Surprisingly, in one strain (O39), the putative OGC does not correspond to the O antigen structure. In particular, neither *rmlA*, *rmlB* and *vioA*, expected for dTDP-D-Qui4N synthesis (Qui4N=4-amino-4,6-dideoxyglucose), nor *gne* for UDP-GalNAc synthesis, was found within the OGC. Instead, homologues of *manB*, *manC* and *gmd* are present,
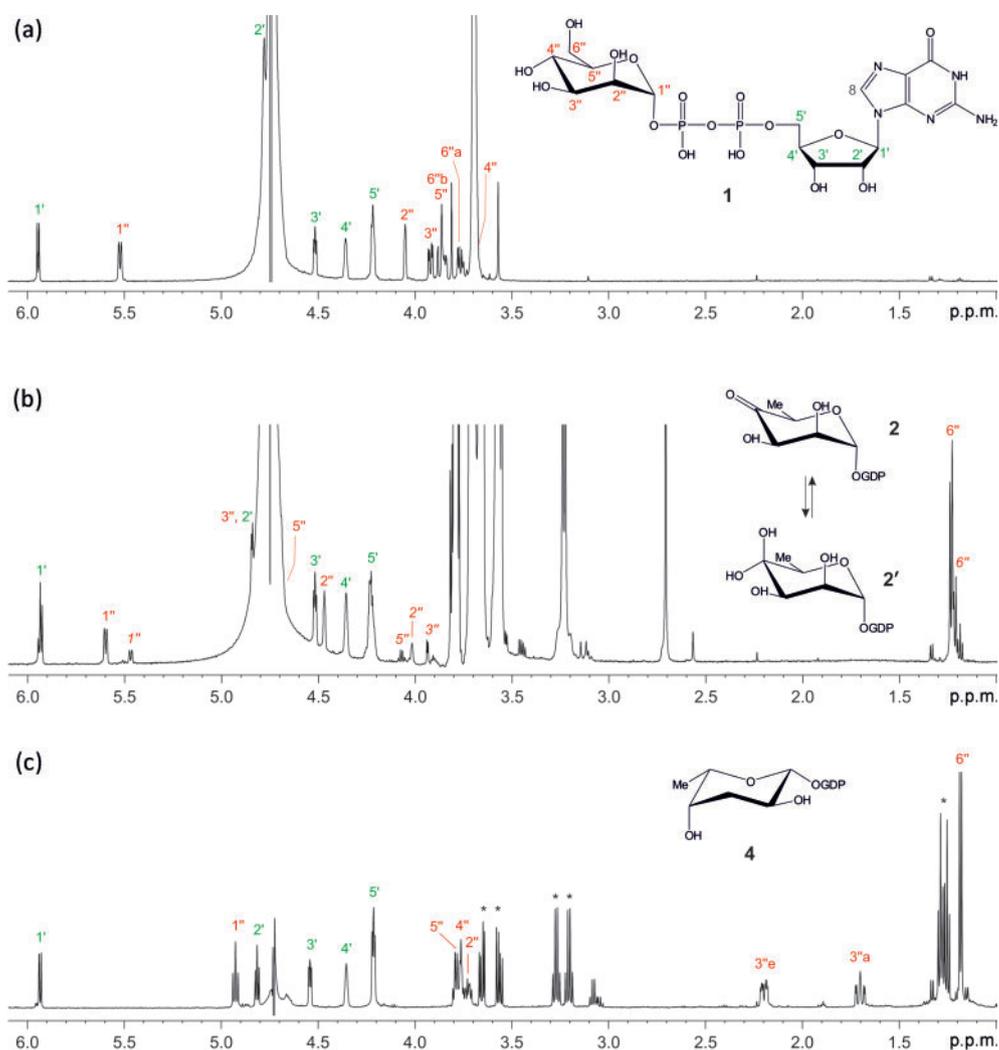
Fig. 4. ^{1}H NMR spectra of the initial substrate GDP-Man (**1**) (a), the Gmd product GDP-6-deoxy-D-*lyxo*-hexos-4-ulose (**2**) (in equilibrium with the *gem*-diol form **2′**) (b), and the final product GDP-Col (**4**) (c). The spectra for **1** and **2** were run in the reaction buffer (10 % D_2O, 50 mM Tris-HCl, pH 7.5), and the spectrum of **4** isolated was measured in D_2O. The spectra were assigned using ^{1}H, ^{1}H correlation spectroscopy (COSY) and TOCSY experiments. Numerals refer to protons in sugar (double primed) and nucleotide (primed) moieties; annotations for the *gem*-diol form **2′** are italicized. The signal for the H-8 proton of guanidine is not shown. Signals of contaminating substances from the HPLC buffer are marked with asterisks.

indicating that another sugar precursor(s) is synthesized from GDP-Man. Two more sugar nucleotide synthesis genes present in the OGC (*orf4* and *orf5*) could be involved in two further steps of this pathway. Orf4 belongs to the NAD-dependent epimerase/dehydratase family and showed the highest identity with GDP-6-deoxy-D-*lyxo*-hex-4-ulose 3,5-epimerase/4-reductase (GDP-L-fucose synthase) of a number of strains, and also a high homology (50 % identity and 70 % similarity) with the putative GDP-Col biosynthetic protein of *S. enterica*. Orf5 belongs to the aspartate aminotransferase superfamily of PLP-dependent enzymes and has identity of up to 64 % with aminotransferases from bacteria of different genera. In addition, a high homology was observed with the predicted

pyridoxamine 5′-phosphate (PMP)-dependent dehydrase of *E. coli* O111. Therefore, both Orf4 and Orf5 share high similarity with proteins that are involved in the biosynthetic pathways of different sugars and which have different functions, although we could not assign any role to these enzymes in the synthesis of the O39 antigen.

To solve the problem, a biochemical confirmation of *gmd* and elucidation of the functions of *orf4* and *orf5* was performed using GDP-D-Man as substrate. Cloning and overexpression of the three genes followed by purification and characterization of the corresponding protein products showed their involvement in the three-step conversion of GDP-D-Man to GDP-Col. Therefore, *orf4* and *orf5*

function as *colC* and *colD*, respectively, encoding enzymes that catalyse 3-deoxygenation of the Gmd product (ColD) and 4-reduction/5-epimerization of the ColD product (ColC).

A search among *Providencia* O antigens of known structure retrieved only one O antigen that contains Col, namely the O6 antigen (Ovchinnikova *et al.*, 2007). The *cpxA–yibK* region of *P. alcalifaciens* O6 was sequenced by primer walking, and the predicted gene functions were found to be in full agreement with the O6 antigen structure. In particular, all genes of the GDP-Col pathway, including *manB*, *manC*, *gmd* (accompanied by *gmm*), *colC* and *colD*, were present. The GDP-Col pathway has been characterized biochemically only in *Yersinia pseudotuberculosis* (Beyer *et al.*, 2003; Alam *et al.*, 2004), and our studies showed that *P. alcalifaciens* uses the same pathway for synthesis of this monosaccharide.

Most remarkably, the O6 OGC was essentially identical to that in the *cpxA–yibK* region of strain O39 (sharing >99 % overall DNA identity). We suggest that *P. alcalifaciens* O39 originated from *P. alcalifaciens* O6 by inactivation of the OGC in the *cpxA–yibK* locus and by the simultaneous gain of another gene cluster for synthesis of the existing O39 antigen at a different locus. The location of the alternative OGC could not be determined in this work, as no full genome sequence of *P. alcalifaciens* O39 is available.

The O36 antigen is distinguished by the presence of Kdo, which is widespread in Gram-negative bacteria, including *Providencia*, as a component of the LPS core and *E. coli* group 2 and 3 CPSs, but rarely occurs in O antigens. Four genes, *kdsABCD*, for synthesis of CMP-Kdo, a donor sugar nucleotide for LPS core assembly, are part of the common enterobacterial genome and are scattered around the chromosome (Reeves & Wang, 2002). Homologues of two of them, *kdsB* and *kdsD*, are present in the *E. coli* group 2 and 3 CPS gene clusters. In the OGC of *P. alcalifaciens* O36, homologues of three genes (*kdsABD*) are present, which evidently duplicate the common CMP-Kdo synthesis genes. Interestingly, this is not the case for *C. sakazakii*, in which the gene clusters for Kdo-containing O antigens contain no genes for CMP-Kdo synthesis (Shashkov *et al.*, 2011).

Among putative glycosyltransferase genes of strain O36, *orf5* shows similarity to *kpsS* and its homologues encoding CPS biosynthesis proteins in various bacteria. This gene is generally found in the group 2 and 3 CPS synthesis clusters (Whitfield, 2006). Earlier studies of *kpsS* mutants gave rise to speculation that KpsS is involved in the addition of the diacylglycerophosphate-Kdo modification to the CPS. However, this has not been confirmed by later studies in *E. coli*, which demonstrate that *kpsS* mutants, although they do not express the capsule on the cell surface, accumulate full-length lipidated intracellular CPS. To date, no biochemical evidence for the exact function of KpsS is available

(Vimr & Steenbergen, 2009). Two glycosyltransferase genes are expected for the synthesis of the O36 antigen, but the full annotation of the O36 OGC revealed only one gene with clear homology to glycosyltransferase genes and left only *orf5*, a homolog of *kpsS*, as the second glycosyltransferase gene necessary for Kdo transfer.

In addition to the O antigen synthesis genes, homologues of *wza*, *wzb* and *wzc* were found upstream of *yibK* in all *Providencia* species except *P. rustigianii*. These genes are required for surface expression of *E. coli* group 1 and 4 CPSs, which can be expressed in two forms: as a capsule or linked to core-lipid A of LPS (Whitfield, 2006). The latter is called K_{LPS} to distinguish it from the normal LPS of the same strain, whose polysaccharide chain is encoded by the OGC at a different locus. Some strains express K_{LPS} as the only serotype-specific polysaccharide, and in this case the term 'O antigen capsule' is sometimes used. In group 1, the CPS gene cluster occupies the same locus between *galF* and *gnd* as the OGC, and differs from the latter by the presence of *wzi*, *wza*, *wzb* and *wzc* genes only. The group 4 gene cluster is in another location and is not distinguished from the OGC, since *wza*, *wzb* and *wzc*, together with the *ymcABCD* genes that are also essential for capsular 4 assembly, are located outside the CPS gene cluster in the so-called '22-minute locus'; in addition, the group 4 CPS producers have no *wzi* gene in the cluster or elsewhere.

The presence of the *wza*, *wzb* and *wzc* genes and the absence of the *ymcABCD* genes in the OGCs places the LPS of *Providencia* O6, O19, O28, O36, O39, O47 and O44 close to group 1 K_{LPS} of *E. coli*. In addition, the group 1-specific *wzi* gene was found in the OGCs of three *Providencia* serogroups.

A polysaccharide chain-length determinator gene, *wzz* (*fepE*), was found in four *Providencia* genomes, within a cluster of genes associated with ferric enterochelin synthesis and transport; accordingly, the corresponding strains produce LPS species having both short and long polysaccharide chains (Ovchinnikova *et al.*, 2011a, b). This is at variance with *E. coli* group 1 capsular producers, which synthesize the K_{LPS} with only one or a few repeating units. However, the modality and the longer chain length can be imparted to group 1 K_{LPS} simply by introduction of a heterologous *wzz* gene (Whitfield, 2006). Hence, the presence of *wzz* may not make a big difference between *E. coli* group 1 K_{LPS} on the one hand and Wzz-containing group 4 K_{LPS} of *E. coli* and LPS of *Providencia* on the other.

To sum up, the molecular organization of the putative OGCs was characterized in nine *Providencia* strains, in one of which the cluster was found to be inactive. Based on the presence of the *wza*, *wzb* and *wzc* genes, the LPS of seven strains is attributed to K_{LPS}. Localization of the OGCs opens the way to the development of a PCR typing method for identification of *Providencia* isolates based on the serogroup-specific *wzx* and *wzy* O antigen processing genes.

ACKNOWLEDGMENTS

REFERENCES

Alam, J., Beyer, N. & Liu, H.-W. (2004). Biosynthesis of colitose: expression, purification, and mechanistic characterization of GDP-4-keto-6-deoxy-D-mannose-3-dehydrase (ColD) and GDP-L-colitose synthase (ColC). *Biochemistry* **43**, 16450–16460.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Andrianopoulos, K., Wang, L. & Reeves, P. R. (1998). Identification of the fucose synthetase gene in the colanic acid gene cluster of *Escherichia coli* K-12. *J Bacteriol* **180**, 998–1001.

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.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M. & Sonnhammer, E. L. (2002). The Pfam protein families database. *Nucleic Acids Res* **30**, 276–280.

Beyer, N., Alam, J., Hallis, T. M., Guo, Z. & Liu, H.-W. (2003). The biosynthesis of GDP-L-colitose: C-3 deoxygenation is catalyzed by a unique coenzyme B$_6$-dependent enzyme. *J Am Chem Soc* **125**, 5584–5585.

Carlin, N. I. A., Lindberg, A. A., Bock, K. & Bundle, D. R. (1984). The *Shigella flexneri* O-antigenic polysaccharide chain. Nature of the biological repeating unit. *Eur J Biochem* **139**, 189–194.

Collins, R. F., Beis, K., Dong, C., Botting, C. H., McDonnell, C., Ford, R. C., Clarke, B. R., Whitfield, C. & Naismith, J. H. (2007). The 3D structure of a periplasm-spanning platform required for assembly of group 1 capsular polysaccharides in *Escherichia coli*. *Proc Natl Acad Sci U S A* **104**, 2390–2395.

Cunneen, M. M. & Reeves, P. R. (2011). Evolution of lipopolysaccharide biosynthesis genes. In *Bacterial Lipopolysaccharides: Structure, Chemical Synthesis, Biogenesis and Interaction with Host Cells*, pp. 339–370. Edited by Y. A. Knirel & M. A. Valvano. Vienna: Springer.

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

Erbel, P. J. A., Barr, K., Gao, N., Gerwig, G. J., Rick, P. D. & Gardner, K. H. (2003). Identification and biosynthesis of cyclic enterobacterial common antigen in *Escherichia coli*. *J Bacteriol* **185**, 1995–2004.

Ewing, W. H. (1986). The tribe Proteae. In *Identification of Enterobacteriaceae*, pp. 454–459. Edited by P. R. Edward. New York: Elsevier.

Ewing, W. H., Tanner, K. E. & Dennard, D. A. (1954). The *Providence* group: an intermediate group of enteric bacteria. *J Infect Dis* **94**, 134–140.

Gronow, S. & Brade, H. (2001). Lipopolysaccharide biosynthesis: which steps do bacteria need to survive? *J Endotoxin Res* **7**, 3–23.

Islam, S. T., Taylor, V. L., Qi, M. & Lam, J. S. (2010). Membrane topology mapping of the O-antigen flippase (Wzx), polymerase (Wzy), and ligase (WaaL) from *Pseudomonas aeruginosa* PAO1 reveals novel domain architectures. *mBio* **1**, e00189-00110–e00189-e00119.

Islam, S. T., Gold, A. C., Taylor, V. L., Anderson, E. M., Ford, R. C. & Lam, J. S. (2011). Dual conserved periplasmic loops possess essential charge characteristics that support a catch-and-release mechanism of O-antigen polymerization by Wzy in *Pseudomonas aeruginosa* PAO1. *J Biol Chem* **286**, 20600–20605.

Juneja, P. & Lazzaro, B. P. (2009). *Providencia sneebia* sp. nov. and *Providencia burhodogranariea* sp. nov., isolated from wild *Drosophila melanogaster*. *Int J Syst Evol Microbiol* **59**, 1108–1111.

Katzenellenbogen, E., Kocharova, N. A., Zatonsky, G. V., Shashkov, A. S., Bogulska, M. & Knirel, Y. A. (2005). Structures of the biological repeating units in the O-chain polysaccharides of *Hafnia alvei* strains having a typical lipopolysaccharide outer core region. *FEMS Immunol Med Microbiol* **45**, 269–278.

Knirel, Y. A. (2011). Structure of O-antigens. In *Bacterial Lipopolysaccharides: Structure, Chemical Synthesis, Biogenesis and Interaction with Host Cells*, pp. 41–115. Edited by Y. A. Knirel & M. A. Valvano. Vienna: Springer.

Knirel, Y. A., Shevelev, S. D. & Perepelov, A. V. (2011). Higher aldulosonic acids: components of bacterial glycans. *Mendeleev Commun* **21**, 173–182.

Kocharova, N. A., Maszewska, A., Zatonsky, G. V., Torzewska, A., Bystrova, O. V., Shashkov, A. S., Knirel, Y. A. & Rozalski, A. (2004). Structure of the O-polysaccharide of *Providencia alcalifaciens* O19. *Carbohydr Res* **339**, 415–419.

Kocharova, N. A., Ovchinnikova, O. G., Toukach, F. V., Torzewska, A., Shashkov, A. S., Knirel, Y. A. & Rozalski, A. (2005). The O-polysaccharide from the lipopolysaccharide of *Providencia stuartii* O44 contains L-quinovose, a 6-deoxy sugar rarely occurring in bacterial polysaccharides. *Carbohydr Res* **340**, 1419–1423.

Kocharova, N. A., Ovchinnikova, O. G., Torzewska, A., Shashkov, A. S., Knirel, Y. A. & Rozalski, A. (2007). The structure of the O-polysaccharide from the lipopolysaccharide of *Providencia alcalifaciens* O36 containing 3-deoxy-D-*manno*-oct-2-ulosonic acid. *Carbohydr Res* **342**, 665–670.

Kocharova, N. A., Kondakova, A. N., Vinogradov, E., Ovchinnikova, O. G., Lindner, B., Shashkov, A. S., Rozalski, A. & Knirel, Y. A. (2008a). Full structure of the carbohydrate chain of the lipopolysaccharide of *Providencia rustigianii* O34. *Chemistry* **14**, 6184–6191.

Kocharova, N. A., Vinogradov, E., Kondakova, A. N., Shashkov, A. S., Rozalski, A. & Knirel, Y. A. (2008b). The full structure of the carbohydrate chain of the lipopolysaccharide of *Providencia alcalifaciens* O19. *J Carbohydr Chem* **27**, 320–331.

Kondakova, A. N., Vinogradov, E., Lindner, B., Kocharova, N. A., Rozalski, A. & Knirel, Y. A. (2007). Mass-spectrometric studies of *Providencia* SR-form lipopolysaccharides and elucidation of the biological repeating unit structure of *Providencia rustigianii* O14-polysaccharide. *J Carbohydr Chem* **26**, 497–512.

Köplin, R., Wang, G., Hötte, B., Priefer, U. B. & Pühler, A. (1993). A 3.9-kb DNA region of *Xanthomonas campestris* pv. *campestris* that is necessary for lipopolysaccharide production encodes a set of enzymes involved in the synthesis of dTDP-rhamnose. *J Bacteriol* **175**, 7786–7792.

Koreishi, A. F., Schechter, B. A. & Karp, C. L. (2006). Ocular infections caused by *Providencia rettgeri*. *Ophthalmology* **113**, 1463–1466.

Marchler-Bauer, A. & Bryant, S. H. (2004). CD-Search: protein domain annotations on the fly. *Nucleic Acids Res* **32** (Web Server issue), W327–W331.

McNally, D. J., Schoenhofen, I. C., Mulrooney, E. F., Whitfield, D. M., Vinogradov, E., Lam, J. S., Logan, S. M. & Brisson, J.-R. (2006). Identification of labile UDP-ketosugars in *Helicobacter pylori*, *Campylobacter jejuni* and *Pseudomonas aeruginosa*: key metabolites used to make glycan virulence factors. *ChemBioChem* **7**, 1865–1868.

Mullane, N., O'Gaora, P., Nally, J. E., Iversen, C., Whyte, P., Wall, P. G. & Fanning, S. (2008). Molecular analysis of the *Enterobacter sakazakii* O-antigen gene locus. *Appl Environ Microbiol* **74**, 3783–3794.

Murata, T., Iida, T., Shiomi, Y., Tagomori, K., Akeda, Y., Yanagihara, I., Mushiake, S., Ishiguro, F. & Honda, T. (2001). A large outbreak of foodborne infection attributed to *Providencia alcalifaciens*. *J Infect Dis* **184**, 1050–1055.

Nakano, Y., Suzuki, N., Yoshida, Y., Nezu, T., Yamashita, Y. & Koga, T. (2000). Thymidine diphosphate-6-deoxy-L-*lyxo*-4-hexulose reductase synthesizing dTDP-6-deoxy-L-talose from *Actinobacillus actinomycetemcomitans*. *J Biol Chem* **275**, 6806–6812.

O'Hara, C. M., Brenner, F. W. & Miller, J. M. (2000). Classification, identification, and clinical significance of *Proteus*, *Providencia*, and *Morganella*. *Clin Microbiol Rev* **13**, 534–546.

Ovchinnikova, O. G., Kocharova, N. A., Bakinovskiy, L. V., Torzewska, A., Shashkov, A. S., Knirel, Y. A. & Rozalski, A. (2004). The structure of the O-polysaccharide from the lipopolysaccharide of *Providencia stuartii* O47. *Carbohydr Res* **339**, 2621–2626.

Ovchinnikova, O. G., Kocharova, N. A., Wykrota, M., Shashkov, A. S., Knirel, Y. A. & Rozalski, A. (2007). Structure of a colitose-containing O-polysaccharide from the lipopolysaccharide of *Providencia alcalifaciens* O6. *Carbohydr Res* **342**, 2144–2148.

Ovchinnikova, O. G., Kocharova, N. A., Kondakova, A. N., Bialczak-Kokot, M., Shashkov, A. S., Knirel, Y. A. & Rozalski, A. (2011a). Structure of the O-polysaccharide from the lipopolysaccharide of *Providencia alcalifaciens* O28. *Carbohydr Res* **346**, 2638–2641.

Ovchinnikova, O. G., Kocharova, N. A., Shashkov, A. S., Arbatsky, N. P., Rozalski, A. & Knirel, Y. A. (2011b). Elucidation of the full O-polysaccharide structure and identification of the core type of the lipopolysaccharide of *Providencia alcalifaciens* O9. *Carbohydr Res* **346**, 644–650.

Pacinelli, E., Wang, L. & Reeves, P. R. (2002). Relationship of *Yersinia pseudotuberculosis* O antigens IA, IIA, and IVB: the IIA gene cluster was derived from that of IVB. *Infect Immun* **70**, 3271–3276.

Pfoestl, A., Hofinger, A., Kosma, P. & Messner, P. (2003). Biosynthesis of dTDP-3-acetamido-3,6-dideoxy-α-D-galactose in *Aneurinibacillus thermoaerophilus* L420-91T. *J Biol Chem* **278**, 26410–26417.

Raetz, C. R. & Whitfield, C. (2002). Lipopolysaccharide endotoxins. *Annu Rev Biochem* **71**, 635–700.

Rahav, G., Pinco, E., Silbaq, F. & Bercovier, H. (1994). Molecular epidemiology of catheter-associated bacteriuria in nursing home patients. *J Clin Microbiol* **32**, 1031–1034.

Rahn, A., Beis, K., Naismith, J. H. & Whitfield, C. (2003). A novel outer membrane protein, Wzi, is involved in surface assembly of the *Escherichia coli* K30 group 1 capsule. *J Bacteriol* **185**, 5882–5890.

Ramm, M., Wolfender, J.-L., Queiroz, E. F., Hostettmann, K. & Hamburger, M. (2004). Rapid analysis of nucleotide-activated sugars by high-performance liquid chromatography coupled with diode-array detection, electrospray ionization mass spectrometry and nuclear magnetic resonance. *J Chromatogr A* **1034**, 139–148.

Raymond, C. K., Sims, E. H., Kas, A., Spencer, D. H., Kutyavin, T. V., Ivey, R. G., Zhou, Y., Kaul, R., Clendenning, J. B. & Olson, M. V. (2002). Genetic variation at the O-antigen biosynthetic locus in *Pseudomonas aeruginosa*. *J Bacteriol* **184**, 3614–3622.

Reeves, P. R. (1994). Biosynthesis and assembly of lipopolysaccharide. In *Bacterial Cell Wall: New Comprehensive Biochemistry*, vol. 27, pp. 281–317. Edited by A. Neuberger & L. L. M. van Deenen. New York: Elsevier.

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

Ren, Y., Perepelov, A. V., Wang, H., Zhang, H., Knirel, Y. A., Wang, L. & Chen, W. (2010). Biochemical characterization of GDP-L-fucose *de novo* synthesis pathway in fungus *Mortierella alpina*. *Biochem Biophys Res Commun* **391**, 1663–1669.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944–945.

Samuel, G. & Reeves, P. R. (2003). Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr Res* **338**, 2503–2519.

Schnaitman, C. A. & Klena, J. D. (1993). Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiol Rev* **57**, 655–682.

Shashkov, A. S., Arbatsky, N. P. & Knirel, Y. A. (2011). Structures and genetics of Kdo-containing O-antigens of *Cronobacter sakazakii* G2706 and G2704, the reference strains of serotypes O5 and O6. *Carbohydr Res* **346**, 1924–1929.

Somvanshi, V. S., Lang, E., Sträubler, B., Spröer, C., Schumann, P., Ganguly, S., Saxena, A. K. & Stackebrandt, E. (2006). *Providencia vermicola* sp. nov., isolated from infective juveniles of the entomopathogenic nematode *Steinernema thermophilum*. *Int J Syst Evol Microbiol* **56**, 629–633.

Sozhamannan, S., Deng, Y. K., Li, M., Sulakvelidze, A., Kaper, J. B., Johnson, J. A., Nair, G. B. & Morris, J. G., Jr (1999). Cloning and sequencing of the genes downstream of the *wbf* gene cluster of *Vibrio cholerae* serogroup O139 and analysis of the junction genes in other serogroups. *Infect Immun* **67**, 5033–5040.

Staden, R. (1996). The Staden sequence analysis package. *Mol Biotechnol* **5**, 233–241.

Stenutz, R., Weintraub, A. & Widmalm, G. (2006). The structures of *Escherichia coli* O-polysaccharide antigens. *FEMS Microbiol Rev* **30**, 382–403.

Tumbarello, M., Citton, R., Spanu, T., Sanguinetti, M., Romano, L., Fadda, G. & Cauda, R. (2004). ESBL-producing multidrug-resistant *Providencia stuartii* infections in a university hospital. *J Antimicrob Chemother* **53**, 277–282.

Valvano, M. A. (2003). Export of O-specific lipopolysaccharide. *Front Biosci* **8**, s452–s471.

Valvano, M. A., Furlong, S. E. & Patel, K. B. (2011). Genetics, biosynthesis and assembly of O-antigen. In *Bacterial Lipopolysaccharides: Structure, Chemical Synthesis, Biogenesis and Interaction with Host Cells*, pp. 275–310. Edited by Y. A. Knirel & M. A. Valvano. Vienna: Springer.

Vimr, E. R. & Steenbergen, S. M. (2009). Early molecular-recognition events in the synthesis and export of group 2 capsular polysaccharides. *Microbiology* **155**, 9–15.

Wang, Q., Torzewska, A., Ruan, X., Wang, X., Rozalski, A., Shao, Z., Guo, X., Zhou, H., Feng, L. & Wang, L. (2010). Molecular and genetic analyses of the putative *Proteus* O antigen gene locus. *Appl Environ Microbiol* **76**, 5471–5478.

Whitfield, C. (2006). Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu Rev Biochem* **75**, 39–68.

Wu, B., Zhang, Y. & Wang, P. G. (2001). Identification and characterization of GDP-D-mannose 4,6-dehydratase and GDP-L-fucose

synthetase in a GDP-L-fucose biosynthetic gene cluster from *Helicobacter pylori*. *Biochem Biophys Res Commun* **285**, 364–371.

Yoh, M., Matsuyama, J., Ohnishi, M., Takagi, K., Miyagi, H., Mori, K., Park, K.-S., Ono, T. & Honda, T. (2005). Importance of *Providencia* species as a major cause of travellers' diarrhoea. *J Med Microbiol* **54**, 1077–1082.

Zhang, L., Toivanen, P. & Skurnik, M. (1996). The gene cluster directing O-antigen biosynthesis in *Yersinia enterocolitica* serotype O : 8: identification of the genes for mannose and galactose biosynthesis and the gene for the O-antigen polymerase. *Microbiology* **142**, 277–288.

Edited by: P. W. O'Toole