REVIEW ARTICLE

Structure and genetics of *Escherichia coli* O antigens

Bin Liu[1,2,3,†], Axel Furevi[4,†], Andrei V. Perepelov[5,†], Xi Guo[1,2,3,†], Hengchun Cao[1,2,3], Quan Wang[1,2,3,‡], Peter R. Reeves[6], Yuriy A. Knirel[5], Lei Wang[1,2,3,*] and Göran Widmalm[4,*]

[1]TEDA Institute of Biological Sciences and Biotechnology, Nankai University, 23 Hongda Street, TEDA, Tianjing 300457, China, [2]The Key Laboratory of Molecular Microbiology and Technology, Ministry of Education, 23 Hongda Street, TEDA, Tianjin 300457 , China, [3]Tianjin Key Laboratory of Microbial Functional Genomics, 23 Hongda Street, TEDA, Tianjin 300457, China, [4]Department of Organic Chemistry, Arrhenius Laboratory, Svante Arrhenius väg 16C, Stockholm University, S-106 91 Stockholm, Sweden, [5]N.D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Leninsky Prospect, 47, Moscow, Russia and [6]School of Molecular and Microbial Bioscience, University of Sydney, 2 Butilin Ave, Darlington NSW 2008, Sydney, Australia

[*]Corresponding author: TEDA Institute of Biological Sciences and Biotechnology, Nankai University, 23 Hongda Street, TEDA, Tianjin 300457, China. Tel: +86 22 66229588; E-mail: wanglei@nankai.edu.cn & Department of Organic Chemistry, Arrhenius Laboratory, Svante Arrhenius väg 16C, Stockholm University, S-106 91 Stockholm, Sweden. Tel: +46 721 464748; E-mail: goran.widmalm@su.se

One sentence summary: This review summarizes the diversity of *E. coli* O antigen structures and gene clusters, and gives a systematic analysis of the genetic basis for structural diversity.

[‡]Current address: Department of Immunology, School of Basic Medical Sciences, Tianjin Medical University, 22 Qixiangtai Road, Heping District, Tianjin 300070, China.

[†]These authors contributed equally to this work.

Editor: Chris Whitfield

ABSTRACT

Escherichia coli includes clonal groups of both commensal and pathogenic strains, with some of the latter causing serious infectious diseases. O antigen variation is current standard in defining strains for taxonomy and epidemiology, providing the basis for many serotyping schemes for Gram-negative bacteria. This review covers the diversity in *E. coli* O antigen structures and gene clusters, and the genetic basis for the structural diversity. Of the 187 formally defined O antigens, six (O31, O47, O67, O72, O94 and O122) have since been removed and three (O34, O89 and O144) strains do not produce any O antigen. Therefore, structures are presented for 176 of the 181 *E. coli* O antigens, some of which include subgroups. Most (93%) of these O antigens are synthesized via the Wzx/Wzy pathway, 11 via the ABC transporter pathway, with O20, O57 and O60 still uncharacterized due to failure to find their O antigen gene clusters. Biosynthetic pathways are given for 38 of the 49 sugars found in *E. coli* O antigens, and several pairs or groups of the *E. coli* antigens that have related structures show close relationships of the O antigen gene clusters within clades, thereby highlighting the genetic basis of the evolution of diversity.

Keywords: *Escherichia coli*; O antigen; gene cluster; serogroup; diversity; structure

INTRODUCTION

Escherichia coli is the predominant facultative anaerobe of the human colonic flora (Kaper, Nataro and Mobley 2004). It typically colonizes the gastrointestinal tract of human infants within a few hours after birth and co-exists with its human host in good health with mutual benefit for decades (Bäckhed *et al*. 2015; Dicks *et al*. 2018). However, *E. coli* includes both commensal and pathogenic clones. Although the commensal *E. coli* rarely cause disease, except in immunocompromised hosts or where the normal gastrointestinal barriers are breached, some *E. coli* clones are adapted to pathogenic niches and cause a broad spectrum of diseases by acquiring specific virulence attributes (Kaper, Nataro and Mobley 2004). Eight *E. coli* pathotypes causing disease in humans have been described, which include six extensively studied intestinal pathotypes and two pathotypes causing extraintestinal infections commonly named extraintestinal pathogenic *E. coli* (ExPEC) (Russo and Johnson 2000). The six intestinal pathotypes include enteropathogenic *E. coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC) and diffusely adherent *E. coli* (DAEC) (Nataro and Kaper 1998); ExPEC include uropathogenic *E. coli* (UPEC) and meningitis-associated *E. coli* (NMEC). Pathogenic *E. coli* are responsible for numerous large outbreaks of infant diarrhea, bloody diarrhea, cystitis, pyelonephritis, meningitis and so on and *E. coli* are increasingly identified as the pathogens causing infectious diseases, such as the novel Shiga toxin-producing *E. coli* strain O104:H4, which resulted in a widespread and severe foodborne disease outbreak in Germany in 2011 (Bielaszewska, Mellmann, Zhang et al. 2011; Muniesa *et al*. 2012).

Lipopolysaccharide (LPS) is found exclusively in the outer leaflet of the Gram-negative outer membrane and performs various biological functions. LPS typically consists of three structural regions: lipid A, a hydrophobic glycolipid which anchors LPS in the bacterial membrane; core oligosaccharide, a non-repeating oligosaccharide which commonly contains sugars such as heptose and keto-deoxyoctulosonate (Kdo); and the O antigen, a polysaccharide composed of multiple oligosaccharide repeating units (O units) each with two to seven residues from a broad range of common or rare sugars and their derivatives (Valvano 2003; Merino, Gonzalez and Tomás 2016). The O antigen is one of the most variable cell constituents, due to variation in the sugars present in the O unit and the linkages within and between O units. The diversity allows the various clones in the species to each present a surface that offers a selective advantage in its specific niche, which probably accounts for the maintenance of O antigen diversity (Reeves 1992; Wang, Wang and Reeves 2010). For instance, a serovar of *Salmonella enterica* may better escape intestinal amoebae predators in a particular intestinal environment by virtue of the specific O-antigen it possesses (Wildschutte and Lawrence 2007). The O antigen is exposed on the cell surface and is usually highly immunogenic, as well as being subject to intense selection by the host immune system and bacteriophages (Reeves and Wang 2002).

The O antigen is also an important virulence factor (Pluschke *et al*. 1983; Achtman and Pluschke 1986), and loss of O antigen makes many pathogens serum sensitive or otherwise seriously impaired in virulence, which has been shown for many species, (Pluschke *et al*. 1983; Bengoechea, Najdenski and Skurnik 2004; West *et al*. 2005; Plainvert *et al*. 2007; Raynaud *et al*. 2007; March

et al. 2013; Sarkar *et al*. 2014; Caboni *et al*. 2015). For example, O antigens could promote immune suppression by enabling UPEC to dampen the induction of cytokines and chemokines in epithelial cells and contribute to the protection against phagocytosis and killing by neutrophils and monocytes (Lüthje and Brauner 2014; Vila *et al*. 2016). Plainvert *et al*. showed that O antigen plays a key role in the virulence of NMEC (Plainvert *et al*. 2007). A recent study also showed that the O antigen could influence the inhibitory ability of LPS on enzymatic and bactericidal activity of lysozyme, a key element of innate immunity characterized by antibacterial activity against ExPEC (Bao *et al*. 2018).

The variability in O antigen structure provides the major basis for the serotyping schemes of many Gram-negative bacteria and is the most widely used method for identifying strains for epidemiological purposes, making O serotyping one of the most important components in typing organisms for taxonomy and epidemiology, and a basic tool used in outbreak investigations and surveillance. In the 1940s, Fritz Kauffmann established the serotyping scheme for *E. coli* which now includes O antigens numbered 1 to 187 (Kauffman 1947; Scheutz *et al*. 2004), but note that O31, O47, O67, O72, O94 and O122 are no longer recognized, some being duplicate names for an O antigen and others were in organisms that were reclassified into other genera. This gives 181 O antigens in *E. coli* at present. Also, a few O antigens contain subgroups, such as O9 and O9a in what was the O9 O antigen, O18ab and O18ac in the O18 O antigen, O28ab and O28ac in the O28 O antigen and O112ab and O112ac in the O112 O antigen. As each subgroup possesses a unique O antigen and gene cluster, they, in fact, represent independent serogroups. Thus, we discuss 185 *E. coli* O serogroups here.

The genes for O antigen synthesis are usually present as a gene cluster at a specific locus. In the majority of *E. coli* and *Shigella*, the O antigen gene cluster is located between two housekeeping genes, *galF* and *gnd*. There are three major classes of genes commonly found in O antigen gene clusters: nucleotide sugar precursor synthesis genes for sugars that are specific to the particular polysaccharide, glycosyltransferase genes that are specific for the donor and acceptor sugars and generate a specific linkage between them, and O unit processing genes for O unit translocation and polymerization. It should be mentioned that a few synthesis genes for sugars also found in other structures or used in metabolism, are usually at other loci, such as *glmU* which is responsible for the synthesis of UDP-GlcNAc from glucosamine-1-phosphate (Guan *et al*. 2009). There are three pathways known for assembly and synthesis of an O antigen, the Wzx/Wzy, ABC transporter and synthase pathways each involving a specific gene or set of genes for key steps in processing the O units (Bronner, Clarke and Whitfield 1994; Keenleyside and Whitefield 1996; Daniels, Vindurampulle and Morona 1998; Linton and Higgins 1998; Samuel and Reeves 2003).

All three O antigen synthesis pathways are initiated by the transfer of a sugar phosphate from an NDP-sugar to undecaprenyl phosphate (Und-P) (Reeves and Wang 2002) and in most species the gene cluster includes an initial transferase (IT) gene for this step; cf. polyprenol phosphate glycosyltransferases (GTs) (Valvano 2003; Lehrer et al. 2007; Eichler and Imperiali 2018). In most *E. coli* and *Shigella*, and most other Enterobacteriaceae, WecA, encoded by the *wecA* gene in the enterobacterial common antigen (ECA) gene cluster, is the IT, transferring a GlcNAc-P residue from UDP-GlcNAc to Und-P (Alexander and Valvano 1994; Rick, Hubbard and Barr 1994; Yao and Valvano 1994). WecA

belongs to the Polyprenol phosphate: *N*-acetylhexosamine-1-phosphate transferases family (Hakulinen *et al.* 2017) and a predicted topological model for WecA (Hakulinen *et al.* 2017) suggests that it contains 11 transmembrane segments, five cytosolic loops and five periplasmic loops (Lehrer *et al.* 2007). The reaction mechanism of "WecA" from *Thermotoga maritima,* was characterized and a conserved aspartate residue (D72) ensures the deprotonation of the hydroxyl group of the Und-P substrate, which is required to allow the nucleophilic attack of the phosphate oxyanion of Und-P on the β-phosphate of UDP-GlcNAc, leading to the formation of the Und-PP-GlcNAc and the release of UMP (Al-Dabbagh *et al.* 2016). The Gnu epimerase, with its gene located upstream of the O antigen cluster, converts Und-PP-GlcNAc to Und-PP-GalNAc when GalNAc is the initial sugar (Rush *et al.* 2010; Cunneen *et al.* 2013).

Most *E. coli* O antigens are synthesized by the Wzx/Wzy pathway, in which sugars are sequentially transferred from the respective sugar nucleotides, initially Und-PP-GlcNAc or Und-PP-GalNAc, by GTs and subsequently to further additions to form the O units. In the ABC transporter pathway, the O antigen is built direcly on an Und-PP-GlcNAc residue. An adaptor is first generated on Und-PP-GlcNAc by one or more GTs, followed by the sugars for an O unit being added sequentially to the adaptor by other GTs, and this process for O-unit addition is repeated many times to give the complete polymer.

In the Wzx/Wzy pathway, the Und-PP-linked O units are translocated by the flippase protein, Wzx, across the inner membrane, to the periplasmic face of the inner membrane, where the O unit is polymerized by the polymerase protein, Wzy, to generate the polymer, with the number of O units in the final O antigen being regulated by the chain length determinant, Wzz (Mulford and Osborn 1983; McGrath and Osborn 1991; Reeves and Wang 2002). Wzx belongs to the polysaccharide transport (PST) family, with 12 or 14 transmembrane segments (TMS) (Paulsen, Beness and Saier 1997). To date, there are no Wzx tertiary structures elucidated; however, studies of Wzx from *E. coli* O157 and *P. aeruginosa* PAO1 identified charged residues essential of translocation, some of which are located in the TMS (Marolda *et al.* 2010; Islam and Lam 2013). Crystal analysis of the MOP (multidrug/oligosaccharidyl-lipid/polysaccharide) flippase MurJ revealed several regions and sites important for the function and suggested that MurJ-mediated flipping occurs via an alternating access mechanism, which is informative for understanding the mechanism of Wzx-mediated flipping (Kuk, Mashalidis and Lee 2017). The sequence variability is believed to be required to accommodate the wide range of available O-unit structures. Early studies by Marolda *et al.* pointed out experimentally that Wzx is specific only for the first sugar of an O-unit substrate (Marolda, Vicarioli and Valvano 2004). But this finding was later shown to be due to over-expression of Wzx, and Wzx does have a preference for its O-unit rather than the first sugar (Hong Y and Reeves 2014). Further work is needed to determine whether the Wzx substrate preference is always strict, as there are several instances of structurally diverse O antigens having the same Wzx, for example in *P. aeruginosa* O2/O5/O16/O18/O20 serogroups (Lam *et al.* 2011). The Wzy polymerase also exhibits low sequence conservation both between and within bacterial species, and the number of TMS predicted in Wzy proteins ranges from 10 to 14. There are no published X-ray crystal structures of Wzy to enhance our understanding of how this protein functions. The number of O units in each LPS molecule is variable with a very interesting "modal" distribution, in which most O antigen polymers have rather similar lengths. This distribution is determined by the Wzz protein. A detailed assessment of chain lengths in *E. coli* O111 showed that most of the molecules had from 9 to 16 O units with few having more or less than that (Bastin *et al.* 1993). Collins *et al.* (2017) recently proposed a model for a Wzx/Wzy complex based on Cryoelectron Microscopy. They also proposed a mechanism for chain length determination based on modeling using the new structure, in which the growing O antigen chain is bound to the surface of Wzz. In their words, polysaccharide chain extension by Wzy would continue until either the polysaccharide-binding capacity of Wzz was reached, or until the Wzy escaped from the Wzz association. The latter 'escape event' would become much less likely (with time) as the size of the polysaccharide chain increases and association of the polysaccharide chain with Wzz becomes stabilized. They noted that this mechanism would incorporate aspects of both molecular ruler and molecular stopwatch mechanisms proposed by Bastin *et al.* in 1993 as alternative mechansms for chain-length determination (Collins *et al.* 2017).

In the ABC transporter pathway, translocation of the Und-PP-O antigen is carried out by an ABC transporter. A typical O antigen ABC transporter is composed of two transmembrane domains (TMDs, also designated Wzm) and two nucleotide-binding domains (NBDs, also designated Wzt) encoded by *wzm* and *wzt*, respectively, with TMDs forming the translocation channel and NBDs hydrolyzing ATP to drive the transport cycle (Cuthbertson, Kimber and Whitfield 2007; Whitfield and Trent 2014; Simpson *et al.* 2015). A recent structural and functional analysis showed a novel processive O antigen translocation mechanism mediated by the ABC transproter. In this process, the Wzm/Wzt forms a transmembrane channel sufficiently wide to accommodate a linear polysaccharide, and the NBD and a periplasmic loop form "gate helices" at the cytosolic and periplasmic membrane interfaces, likely serving as substrate entry and exit points (Bi *et al.* 2018). In some systems, the ABC transporter recognizes the modified terminus via a carbohydrate-binding domain (CBD) fused to the C terminus of its NBD to accomplish transport. This is the case for *E. coli* O9a and *K. pneumoniae* O12, whose O antigen chains are capped at their non-reducing (growing) termini by methyl-phosphate and Kdo residues, respectively (Clarke *et al.* 2009; Mann *et al.* 2016; Liston, Mann and Whitfield 2017). In other systems represented by *K. pneumoniae* serogroup O2a lacking a capping residue, export of the O antigen is obligatorily coupled to chain elongation, and the ABC transporter is essential for establishing the O antigen (Kos, Cuthbertson and Whitfield 2009; Liston, Mann and Whitfield 2017)

The synthase pathway has only been reported in bacteria for the plasmid-encoded *S. enterica* O54 O antigen (Keenleyside *et al.* 1994). The *S. enterica* O54 O antigen is a homopolymer of *N*-acetylmannosamine (ManNAc). Similarly, to the ABC transporter pathway, the synthesis is initiated by WecA and the first ManNAc residue is added to the Und-PP-GlcNAc primer by the non-processive transferase WbbE, forming the adaptor. Next, the second transferase, WbbF, is proposed to perform the chain-extension steps. However, the exact mechanism of export mediated by WbbF as well as the process of chain termination remain unclear.

After translocation and polymerization mediated by one of the three pathways above, the resultant O polysaccharides are then transferred from the Und-PP moiety, and attached to lipid A-core by the ligase WaaL (Han *et al.* 2011; Ruan *et al.* 2012) to give an LPS molecule ready for transport to the outermembrane via the Lpt pathway (Okuda *et al.* 2016).

A review in 2006 described the structures of *E. coli* O-polysaccharide antigens based on 75 published *E. coli* O antigen structures at that time (Stenutz, Weintraub and Widmalm 2006) and the currently known structures are included in the database ECODAB (www.casper.organ.su.se/ECODAB/) (Lundborg, Modhukur and Widmalm 2010; Rojas-Macias *et al.* 2015). Two recent reviews cover most of the O antigen gene cluster sequences (DebRoy *et al.* 2016; Iguchi *et al.* 2015a). This review covers the diversity in both structures and gene clusters of *E. coli,* and the genetic basis of the structural diversity. The structures are shown in Table 1 and the gene clusters in Fig. S1 (Supporting Information).

E. coli O antigen structures

The structures for 197 *E. coli* O antigens (including 21 subgroups) have been elucidated to date (Table 1). Note that the "O antigen" includes only the polysaccharide composed of O units (or single O unit if the polymerase is missing) and (in some cases) the terminating residue for the ABC transporter pathway. More than half of the O units are branched (58%) and the remaining are linear (42%). The backbones contain two to six residues. The most common (51%) has four residues in the backbone, and most others have three or five residues (41%), while only a few backbones contain two or six residues (Table 2). All seven two-residue backbones (O20, O52, O92, O95, O97, O101 and O162) have O units that use the ABC transporter dependent pathway, except O20 whose gene cluster is not at the *galF-gnd* locus and has not yet been characterized. Most branched O units have a one-residue side-branch (85%), but a few have two residues, viz. O29, O36, O55, O102, O105, O124, O164 and O183. The most common *E. coli* O unit contains a four-residue backbone with a single residue side-branch.

There are 49 different sugars found in the *E. coli* O antigens (Table 3) and the 38 hexoses account for 97% of the sugar occurrences. There are four kinds of nonose (Neu, 8eLeg, Leg and Pse) present in one or more of the O antigens of O24, O56, O104, O145, O108, O161, O136, two kinds of pentose (D-Rib*f* and D-Xul*f*) present in one or more of O5, O20, O105, O114, O153, O178, O97 and one heptose (6d-D-*man*Hep) is present in O52. Most of the monosaccharides are in the pyranose form (97%) with only D-Gal and D-Fuc also found in the furanose form; D-Rib and D-Xul occur only in the furanose form. Most (79%) of the sugars have the D-configuration and the others (21%) have the L-configuration. Among the hexoses, the most common are D-GlcNAc, D-Gal, D-Glc, L-Rha, D-GalNAc and D-Man, which account for 18%, 13%, 13%, 12%, 11% and 8%, respectively, of the sugars. In addition, L-Fuc, L-FucNAc, D-GlcA, and D-GalA are present more than 10 times, and D-Gal*f*, D-Qui3NAc, D-ManNAc, D-Qui4NAc, D-GalNAcA, D-Rha, 6d-L-Tal, Col, L-QuiNAc and D-Fuc3NAc present 3 to 9 times. Seven kinds of hexoses viz. D-FucNAc, L-RhaNAc, D-Fuc4NAc, D-Fuc*f*, D-Rha4NAc, D-ManNAc3NAcA and L-IdoA are rare, and present only in O45, O3, O10, O52, O157, O180 and O112ab, respectively. It is also notable that some sugars including Col, L-RhaNAc, D-Fuc4NAc, L-RhaNAc3NAc (or L-RhaNAc3NFo) are only present in side-branches. A total of 19 non-sugar constituents are found in *E. coli* O antigens, of which the acetyl group is the most common constituent (70%), others being e.g. amide-linked amino acids such as glycine (O91), *N*-acetyl-glycine (O121), L-alanine (O167), *N*-acetyl-L-serine (O114), D-*allo*-threonine (O110) and *N*-acetyl-D-*allo*-threonine (O63), pyruvate groups substituting positions 4 and 6 of *N*-acetyl-D-glucosamine (O112ac and

O149) or D-galactose (O156) as well as lactyl groups substituting L-rhamnose (O58), D-glucose (O124 and O146) or *N*-acetyl-D-glucosamine (O69 and O150).

Structure analysis usually determines the linkages present in the repeating unit including that made during polymerization, but commonly does not distinguish the initial sugar, as neither the terminal sugar nor the linkage to the core are identified because they are present only once. Thus, the structure as presented is sometimes a frameshift of the repeat unit. In the absence of an alternative IT in the gene cluster and near universal presence of D-GlcNAc and or D-GalNAc, which is transferred as 1-phosphate to Und-P by IT to initiate O-antigen synthesis, it is usually assumed that one or the other is the first sugar.

According to this and based on the 195 known structures (excluding O101 and O162, as discussed below), D-GlcNAc can be unambiguously confirmed as the initial sugar for 101 (52%) of the known structures, as there are no D-GalNAc residues. For 37 of them there is more than one D-GlcNAc, so the specific initial D-GlcNAc cannot be decided by considering just the primary structure of the repeating unit. D-GalNAc is confirmed to be the initial sugar for 59 (30%) of the O antigens as no D-GlcNAc is present, but the initial residue cannot be determined in 23 O antigens with more than one D-GalNAc.

Among the 16 O antigens that contain just one of each, D-GlcNAc is determined to be the initial sugar in 12 of them (O6, O23, O35, O48, O49, O65, O80, O98, O113, O138, O143 and O181), as the *gne* gene, whose product catalyzes the isomerization from UDP-D-GlcNAc to UDP-D-GalNAc, is present in the corresponding gene cluster, and D-GalNAc can be deduced as the initial sugar of the remaining four (O33, O55, O160 and O178), as the *gnu* gene, whose product converts Und-PP-GlcNAc to Und-PP-GalNAc when GalNAc is the initial sugar, is located just upstream of *galF*. For O103, O107 and O116, with structures possessing both D-GlcNAc and D-GalNAc simultaneously, it is proposed that D-GlcNAc should act as the initial sugar, as each gene cluster contains the *gne* gene but the *gnu* gene was not present upstream of *galF,* as is common for O antigens with GlcNAc as first sugar (Wang *et al.* 2002; Cunneen *et al.* 2013). However, theoretically, the possibility of having D-GalNAc as the first sugar of the O-unit could not be excluded, as the *gnu* homolog was not screened at the whole genome level.

The O101 and O162 O units both have two structural forms. The O101 unit contains →6)-D-GlcNAc-(α1→4)-D-GalNAc-(α1→ (form 1) and →4)-D-GalNAc-(α1→4)-D-GalNAc-(β1→ (form 2). The O antigen of O162 has an additional sugar, 4d-D-*ara*Hex attached to the GlcNAc of form 1, and to the second GalNAc of form 2. Although there is a *gne* gene within each gene cluster, it is hard to infer which sugar is the first, as these two serogroups utilize the ABC transporter pathway for the O antigen synthesis.

Neither D-GlcNAc nor D-GalNAc are present in the O8, O9, O9a, O20, O45, O52, O60, O62, O92, O95, O97 and O99 O units. All except O20, O45 and O60 are known to be synthesized by the ABC transporter dependent pathway; O20 and O60 are discussed below and the O45 O antigen contains D-Glc, 6d-L-Tal and D-FucNAc, with D-FucNAc proposed as the initial sugar, as is the case for *Pseudomonas aeruginosa* (Rocchetta *et al.* 1998; Dean *et al.* 1999; Raymond *et al.* 2002). The O45-antigen is unique in *E. coli* in having an IT gene in the gene cluster. The *wbhQ* gene encodes a product with 53% identity to WbpL of *P. aeruginosa*, and is proposed to have the same function using D-FucNAc as first sugar (Rocchetta *et al.* 1998).

Table 1. O antigen structures of *E. coli* serogroups.

Serogroup	Structure	Reference
O1A O1A1	→3)-LRha-(α1→3)-LRha-(α1→3)-LRha-(β1→4)-DGlcNAc-(β1→ | DManNAc-(β1→2)	Baumann et al.(1991) Jann et al.(1992a)
O1B	→3)-LRha-(α1→2)-LRha-(α1→2)-DGal-(α1→3)-DGlcNAc-(β1→ | DManNAc-(β1→2)	Gupta et al.(1992)
O1C	→3)-LRha-(α1→2)-LRha-(α1→3)-DGal-(α1→3)-DGlcNAc-(β1→ | DManNAc-(β1→2)	Gupta et al.(1992)
O2	→3)-LRha-(α1→2)-LRha-(α1→3)-LRha-(β1→4)-DGlcNAc-(β1→ | DFuc3NAc-(α1→2)	Jansson et al.(1987a)
O3	LRhaNAc-(β1→4) DGlc-(α1→4) | | →3)-DGlcNAc-(β1→3)-DGal-(α1→3)-DGlcNAc-(β1→	Medina et al.(1994)
O4	→2)-LRha-(α1→6)-DGlc-(α1→3)-LFucNAc-(α1→3)-DGlcNAc-(β1→	Jann et al.(1993)
O4	→2)-LRha-(α1→6)-DGlc-(α1→3)-LFucNAc-(α1→3)-DGlcNAc-(β1→ | DGlc-(α1→3)	Jansson et al.(1984a)
O5ab	→4)-DQui3NAc-(β1→3)-DRib*f*-(β1→4)-DGal-(β1→3)-DGalNAc-(α1→	MacLean et al.(1997)
O5ac	→2)-DQui3NAc-(β1→3)-DRib*f*-(β1→4)-DGal-(β1→3)-DGalNAc-(α1→	Urbina et al.(2005)
O6	→4)-DGalNAc-(α1→3)-DMan-(β1→4)-DMan-(β1→3)-DGlcNAc-(α1→ | DGlcNAc-(β1→2)	Jann et al.(1994a)
O6	→4)-DGalNAc-(α1→3)-DMan-(β1→4)-DMan-(β1→3)-DGlcNAc-(α1→ | DGlc-(β1→2)	Jansson et al.(1984b) Jann et al.(1994a)
O7	→3)-DQui4NAc-(β1→2)-DMan-(α1→4)-DGal-(β1→3)-DGlcNAc-(α1→ | LRha-(α1→3)	L'vov et al.(1984)
O8	→3)-DMan-(β1→2)-DMan-(α1→2)-DMan-(α1→	Jansson et al.(1985)
O9	→3)-DMan-(α1→3)-DMan-(α1→2)-DMan-(α1→2)-DMan-(α1→2)-DMan-(α1→	Prehm et al.(1976) Curvall et al.(1973)
O9a	→3)-DMan-(α1→3)-DMan-(α1→2)-DMan-(α1→2)-DMan-(α1→	Parolis et al.(1986)
O10[*i*]	→3)-LRha-(α1→3)-LRha-(α1→3)-DGal-(α1→3)-DGlcNAc-(β1→ | DFuc4NAcyl-(α1→2)	Kenne et al.(1986)
O11	→4)-DGalNAc-(α1→6)-DMan-(α1→3)-LFuc-(α1→3)-DGalNAc-(β1→ | LFuc-(α1→3)	Li et al.(2011)
O12	→2)-DGlc-(β1→6)-DGlcNAc-(α1→3)-LFucNAc-(α1→3)-DGlcNAc-(β1→	Senchenkova et al.(2016a)
O13	→2)-LRha-(α1→2)-LRha-(α1→3)-LRha-(α1→3)-DGlcNAc6Ac-(β1→ | DGlc-(α1→2)	Perepelov et al.(2010a)
O14[*ii*]		
O15	→2)-DGal-(β1→3)-LFucNAc-(α1→3)-DGlcNAc-(β1→	Perepelov et al.(2011a)
O16	DGlc-(α1→6) | →2)-DGal*f*-(β1→6)-DGlc-(α1→3)-LRha2Ac-(α1→3)-DGlcNAc-(α1→	Stevenson et al.(1994)
O16	→2)-DGal*f*-(β1→6)-DGlc-(α1→3)-LRha-(α1→3)-DGlcNAc-(α1→	Batley et al.(1997)
O16	→2)-DGal*f*-(β1→6)-DGlc-(α1→3)-LRha2Ac-(α1→3)-DGlcNAc-(α1→	Jann et al.(1994b)
O17	DGlc-(α1→6) | →6)-DMan-(α1→2)-DMan-(α1→2)-DMan-(β1→3)-DGlcNAc-(α1→	Masoud et al.(1996) Batley et al.(1997)
O18A O18ac	→2)-LRha-(α1→6)-DGlc-(α1→4)-DGal-(α1→3)-DGlcNAc-(α1→ | DGlcNAc-(β1→3)	Jansson et al.(1989)
O18A1	DGlc-(α1→6) | →2)-LRha-(α1→6)-DGlc-(α1→4)-DGal-(α1→3)-DGlcNAc-(α1→ | DGlcNAc-(β1→3)	Jann et al.(1992b)

Table 1. Continued

O18A1A2	DGlc-(α1→6) | →2)-LRha-(α1→6)-DGlc-(α1→4)-DGal-(α1→3)-DGlcNAc-(α1→ | DGlc/DGlcNAc-(β1→3)	Zaccheus et al.(2012)
O18B O18ab	→3)-LRha-(α1→6)-DGlc-(α1→4)-DGal-(α1→3)-DGlcNAc-(α1→ | DGlc-(β1→3)	Jann et al.(1992b)
O18B1	DGlc-(α1→4) | →3)-LRha-(α1→6)-DGlc-(α1→4)-DGal-(α1→3)-DGlcNAc-(α1→ | DGlc-(β1→3)	Jann et al.(1992b)
O19ab	→2)-LRha-(α1→2)-LRha-(α1→2)-LRha-(α1→2)-DGlc-(α1→3)-DGlcNAc6Ac-(α1→	Perepelov et al.(2011b)
O20ab	→2)-DRibf-(β1→4)-DGal-(α1→	Vasil'ev et al.(1976)
O20ac	→2)-DRibf-(β1→4)-DGal-(α1→ | DGal-(α1→3)	Vasil'ev et al.(1982)
O21	DGal-(β1→4) | →3)-DGal-(β1→4)-DGlc-(β1→3)-DGalNAc-(β1→ | DGlcNAc-(β1→2)	Staaf et al.(1999a)
O22	→6)-DGlc-(α1→4)-DGlcA-(β1→4)-DGalNAc3Ac-(β1→3)-DGal-(α1→3)-DGalNAc-(β1→	Bartelt et al.(1994)
O23A	DGlc-(α1→6) | →6)-DGlc-(α1→4)-DGal-(β1→3)-DGalNAc-(α1→3)-DGlcNAc-(β1→ | DGlcNAc-(β1→3)	Bartelt et al.(1993)
O24	→7)-Neu5Ac-(α2→3)-DGlc-(β1→3)-DGalNAc-(β1→ | DGlc-(α1→2)	Gamian et al.(1994) Torgov et al.(1995)
O25 Strain F171	DGlc-(β1→6) | →4)-DGlc-(α1→3)-LFucNAc-(α1→3)-DGlcNAc-(β1→ | LRha-(α1→3)	Kenne et al.(1983) Fundin et al.(2003)
O26	→3)-LRha-(α1→4)-LFucNAc-(α1→3)-DGlcNAc-(β1→	Manca et al.(1996)
O27	DGlcA3Ac/4Ac-(β1→3) DGlc-(β1→6) | | →4)-DGal-(α1→6)-DGlc-(β1→3)-DGalNAc-(β1→ | DGlc-(β1→2)	Perepelov et al.(2018a)
O28ab	→3)-DGlc-(β1→3)-DGro-(1-P-4)-DGlcNAc-(β1→3)-DGlcNAc-(α1→	Shashkov et al.(2016a)
O28ac	→2)-DGro-(1-P-4)-DGlcNAc-(β1→3)-DGalf2Ac-(β1→3)-DGlcNAc-(α1→	Rundlöf et al.(1996)
O29	→3)-DGro-(1-P-6)-DGlc-(β1→4)-LFucNAc-(α1→3)-DGlcNAc-(β1→ | DGlc-(α1→6)-DGal-(α1→3)	Perepelov et al.(2006)
O30	→4)-DGlcA-(β1→4)-DGlcA-(β1→3)-DGlcNAc-(α1→ | DGlcNAc-(β1→2)	Perepelov et al.(2014a)
O32	→6)-DGal-(α1→4)-DGlcA-(β1→6)-DGal-(β1→4)-DGal2Ac/3Ac-(β1→4)-DGlcNAc-(β1→	Knirel et al.(2016a)
O33	→2)-DGal-(β1→4)-DGlcNAc-(β1→3)-DGal-(β1→3)-DGalNAc-(β1→ | Gro-(2-P-3)	Senchenkova et al.(2018)
O34[iii]		
O35	→3)-LRha-(α1→2)-LRha-(α1→3)-LRha-(α1→2)-LRha-(α1→3)-DGlcNAc-(β1→ | DGalNAcAN-(α1→2)	Rundlöf et al.(1998)
O36	→4)-DMan-(α1→3)-LFuc-(α1→3)-DGlcNAc-(β1→ | LRha-(α1→2)-LFuc-(α1→3)	Perepelov et al.(2014b)
O37	→4)-DGalNAc-(β1→3)-DGal-(α1→4)-DGal-(β1→3)-DGalNAc-(β1→ | DGro-(1-P-3)	Shashkov et al.(2016a)
O38	DGlcNAc-(α1→4) | →6)-DGlcNAc-(β1→3)-DGalNAc-(β1→4)-DGlcA-(β1→3)-DGalNAc-(β1→	Knirel et al.(2016a)
O39	→3)-DQui4NBuR3OH-(β1→2)-DMan-(α1→4)-LRha-(α1→3)-DGlcNAc-(α1→ | DGal-(α1→3)	Perepelov et al.(2014c)
O40	→2)-DGal-(β1→4)-DMan-(β1→4)-DGal-(α1→3)-DGlcNAc-(β1→	Zhao et al.(2007a)

Table 1. Continued

O41	DGlcA-(β1→4) | →3)-DGal-(α1→3)-LFuc-(α1→4)-DGlcNAc-(β1→3)-LFuc-(α1→3)-DGlcNAc-(β1→	Zhu et al.(2012)
O42	→2)-DGro-(1-*P*-4)-DGlcNAc-(β1→3)-DGal*f*2Ac-(β1→3)-DGlcNAc-(α1→ | DGlc-(α1→3)	Fontana et al.(2015a)
O43	→4)-DMan2Ac-(β1→3)-DMan-(α1→3)-LFuc-(α1→3)-DGalNAc-(β1→ | LFuc-(α1→3)	Perepelov et al.(2015a)
O44	DGlc-(α1→4) | →6)-DMan-(α1→2)-DMan-(α1→2)-DMan-(β1→3)-DGlcNAc-(α1→	Staaf et al.(1995)
O45	→2)-DGlc-(β1→3)6dLTal2Ac-(α1→3)-DFucNAc-(α1→	Jann et al.(1995)
O45 rel	→2)-DGlc-(β1→3)6dLTal2Ac-(α1→3)-DGlcNAc-(β1→	Jann et al.(1995)
O46	→4)-DGlcA6LThr3Ac-(β1→6)-DGal-(β1→6)-DGlc-(β1→3)-DGalNAc6Ac-(β1→	Perepelov et al.(2015b)
O48	→4)-DGalNAc-(β1→3)-DGal-(α1→4)-LRha-(α1→3)-DGlcNAc-(β1→ | DGalA2Ac/3Ac-(α1→3)	Maclean et al.(2006)
O49	→2)-DQui4NBu*S*3OH-(α1→4)-DGalNAc-(β1→4)-LRha-(α1→3)-DGlcNAc6Ac-(β1→	Perepelov et al.(2008a)
O50	→3)-LRha-(α1→2)-LRha-(α1→3)-LRha-(β1→4)-DGlcNAc-(β1→	Yang et al.(2018)
O51	→3)-LRha-(α1→2)-LRha-(α1→4)-DGlc-(α1→3)-DGalNAc-(β1→ | DGlcNAc-(β1→2)	Perepelov et al.(2011c)
O52	→3)-DFuc*f*2Ac-(β1→3)6d*Dman*Hep-(β1→	Feng et al.(2004)
O53	→3)-LRha-(α1→4)-DGlcA-(β1→3)-LRha-(β1→4)-DGlcNAc-(β1→ | DGlc-(α1→2)	Knirel et al.(2016a)
O54	→4)-DGalA-(α1→2)-LRha-(α1→2)-DRib*f*-(β1→4)-DGal-(β1→3)-DGlcNAc-(β1→	Naumenko et al.(2018a)
O55	→6)-DGlcNAc-(β1→3)-DGal-(α1→3)-DGalNAc-(β1→ | Col-(α1→2)-DGal-(β1→3)	Lindberg et al.(1981)
O56	→7)-Neu5Ac-(α2→3)-DGlc-(β1→3)-DGlcNAc-(β1→ | DGal-(α1→2)	Gamian et al.(1994) Torgov et al.(1995)
O57	DGlc-(β1→4) | →3)-DGal-(α1→3)-LFucNAc-(α1→3)-DGlcNAc-(α1→ | DGalA2Ac/3Ac-(α1→2)	Naumenko et al.(2018b)
O58	→4)-DMan-(β1→4)-DMan-(α1→3)-DGlcNAc-(β1→ | LRha2Ac3*R*Lac-(α1→3)	Perepelov et al.(2008b)
O59	→6)-DMan-(α1→3)-DMan-(α1→3)-DMan-(β1→3)-DGlcNAc-(α1→ | DGalA-(α1→3)	Fontana et al.(2014)
O60[iv]	→2)-LRha-(α1→2)-LRha-(α1→4)-DGal-(β1→ | DGal-(α1→3)	Perepelov et al.(2018b)
O61	→4)-8eLeg5Ac7Ac8Ac-(α2→4)-DGlc-(β1→3)-DGalNAc-(α1→	Perepelov et al.(2009a)
O62	→5)-DGal*f*-(β1→5)-DGal*f*-(β1→3)-DGal*f*-(β1→ | | DGlc-(α1→2) DGlc-(α1→2)	Hou et al.(2017)
O63	→2)-DQui3N*allo*ThrAc-(β1→2)-DRib*f*-(β1→4)-DGal-(β1→3)-DGlcNAc-(α1-	Ståhle et al.(2019)
O64	DGal-(β1→6) | →3)-DManNAc-(α1→3)-DGlcA-(β1→3)-DGal-(β1→3)-DGlcNAc-(β1→	Perry et al.(1993)
O65	→2)-DQui3NAc-(β1→4)-DGalA6N-(α1→4)-DGalNAc-(α1→4)-DGalA-(β1→3)-DGlcNAc-(α1→	Perry et al.(1999)
O66	→2)-DGlc3Ac-(β1→3)6dLTal-(α1→3)-DGlcNAc-(α1→2)-DMan-(β1→3)-DGlcNAc-(α1→	Jann et al.(1995)
O68	LRha-(α1→3) | →6)-DMan-(α1→2)-DMan-(α1→2)-DMan-(α1→2)-DMan-(β1→3)-DGlcNAc-(α1→ | DGlc-(α1→3)	Jiang et al.(2014)
O69	→3)-DGlcNAc4*R*Lac6Ac-(β1→2)-LRha-(α1→2)-LRha-(α1→2)-DGal-(α1→3)-DGlcNAc-(α1→	Knirel et al.(2017)

Table 1. Continued

O70	DGlc-(α1→4) | →4)-DQui3NAc-(β1→3)-DFuc-(α1→3)-DGal-(β1→3)-DGalNAc-(α1→	MacLean et al.(2010a)
O71	→4)-DQui3NAc-(α1→3)-LRha4Ac-(α1→4)-DGal2Ac/3Ac-(β1→3)-DGalNAc-(α1→	Hu et al.(2010) MacLean et al.(2010b)
O73	DGlc-(α1→4) | →6)-DMan-(α1→2)-DMan-(α1→2)-DMan-(β1→3)-DGlcNAc-(α1→ | DGlc-(α1→3)	Wang et al.(2007)
O73 strain 73−1	→4)-DMan-(α1→2)-DMan-(α1→2)-DMan-(β1→3)-DGalNAc-(α1→ | DGlc-(α1→3)	Weintraub et al.(1993)
O74	→6)-DGlcNAc-(α1→4)-DGalA-(β1→3)-DGlcNAc-(β1→ | DFuc3NAc-(β1→3)	Jonsson et al.(2009)
O75	DMan-(β1→4) | →3)-DGal-(α1→4)-LRha-(α1→3)-DGlcNAc-(β1→	Erbing et al.(1978)
O76	→4)-DGlcA-(β1→4)-DGalNAc3Ac-(β1→4)-DGalNAc-(α1→3)-DGalNAc-(β1→	Perepelov et al.(2013a)
O77	→6)-DMan-(α1→2)-DMan-(α1→2)-DMan-(β1→3)-DGlcNAc-(α1→	Yildirim et al.(2001)
O78	→4)-DGlcNAc-(β1→4)-DMan-(β1→4)-DMan-(α1→3)-DGlcNAc-(β1→	Jansson et al.(1987b)
O79	→2)-DGal-(β1→4)-DMan6Ac-(β1→3)-DMan-(β1→4)-DGlcA-(β1→3)-DGlcNAc-(α1→ | LRha-(α1→3)	Liu et al.(2008) Knirel et al.(2016a)
O80	LFuc-(α1→2) LFuc-(α1→4) | | →6)-DGlc-(β1→4)-DGalNAc-(β1→4)-DGlcA-(β1→3)DGlcNAc(β1→	Senchenkova et al.(2016b)
O81	DGro-(1-P-3) DGal-(α1→6) | | →6)-DGalNAc-(β1→4)-DGalNAc-(β1→4)-DGlc-(β1→3)-DGalNAc-(β1→	Zdorovenko et al.(2018)
O82	DGroA-(2 −P-6) | →4)-DGlc-(α1→4)-DGal-(β1→4)-DGlc-(β1→3)-DGlcNAc-(β1→	Vilchez et al.(2009)
O83	→6)-DGlc-(α1→4)-DGlcA-(β1→6)-DGal-(β1→4)-DGal-(β1→4)-DGlcNAc-(β1→	Jann et al.(1994c)
O84	6dDTal2Ac3Ac/2Ac4Ac/3Ac4Ac-(α1→3) | →2)-LFuc-(β1→P-4)-DGlc-(α1→3)-LFuc-(α1→3)-DGlcNAc-(β1→	Knirel et al.(2016b)
O85	DGalf-(α1→4) | →2)-DGal-(α1→3)-DManNAc-(β1→6)-DGalf-(β1→3)-DGlcNAc-(β1→	Perepelov et al.(2011d)
O86	DGal-(α1→3) | →4)-LFuc-(α1→2)-DGal-(β1→3)-DGalNAc-(α1→3)-DGalNAc-(β1→	Andersson et al.(1989)
O86 *wzy*	→2)-DGal-(α1→3)-DGal-(β1→3)-DGalNAc-(α1→3)-DGalNAc-(α1→ | LFuc-(α1→2)	Guo et al.(2005) Yi et al.(2006)
O87	→3)-DGalf-(β1→3)-DGal-(α1→3)-DGalA4Ac-(α1→3)-DGalNAc-(β1→	Zdorovenko et al.(2015)
O88	→4)-DMan-(α1→3)-DMan-(α1→3)-DGlcNAc-(β1→ | 6dLTal-(α1→3)	Torgov et al.(1996)
O89[ii]		
O90	→4)-LFuc2Ac/3Ac-(α1→2)-DGal-(β1→3)-DGalNAc-(α1→3)-DGalNAc-(β1→	Ratnayake et al.(1994a)
O91	→4)-DQui3NBuR3OH-(α1→4)-DGal-(β1→4)-DGlcNAc-(β1→4)-DGlcA6NGly-(β1→3)-DGlcNAc-(β1→	Kjellberg et al.(1999)
O92	→1)-DFruf-(β2→3)-DFruf-(β2→	Senchenkova et al.(2017)
O93	Trisaccharide repeating unit	Unpublished, Widmalm and co-workers.
O95	→3)-DFuc-(α1→3)-DXulf-(β2→	Perepelov et al.(2018c)
O96	DGalf-(β1→4) | →2)-DGlcA-(α1→4)-DGlcNAc-(β1→5)-DGalf-(β1→3)-DGlcNAc-(β1→	Guo et al.(2016)
O97	DXulf-(β2→2) | →3)-LRha-(α1→3)-LRha-(β1→ | DXulf-(β2→2)	Perry et al.(1987)

Table 1. Continued

O98	→3)-LQuiNAc-(α1→4)-DGalNAcA-(α1→3)-LQuiNAc-(α1→3)-DGlcNAc-(β1→	Marsden et al.(1994)
O99	→3)-DRha-(α1→2)-DRha-(α1→2)-DRha-(α1→3)-DRha-(α1→ 　　　　　　｜　　　　　　　　　　　　　　　　　｜ DGlc-(α1→2)　　　　　　　　　　　　DGlc-(α1→2)	Perepelov et al.(2009b)
O100	DGro-(1-*P*-3) 　　　｜ 　→2)-LRha-(α1→3)-LRha-(α1→3)-DGal-(α1→3)-DGlcNAc-(β1→	Shashkov et al.(2016a)
O101	→6)-DGlcNAc-(α1→4)-DGalNAc-(α1→, →4)-DGalNAc-(β1→4)-DGalNAc-(α1→	Staaf et al.(1997) Shashkov et al.(2017)
O102	→4)-DGal-(α1→6)-DGlc-(β1→3)-DGalNAc-(β1→ 　　　　　　　　　　　　　　　　｜ LRha-(α1→4)-DGalNAc-(β1→3)	Perepelov et al.(2012a)
O103	→2)-DGlc-(β1→2)-DFuc3NBu*R*3OH-(β1→6)-DGlcNAc-(α1→4)-DGalNAc-(α1→3)-DGlcNAc-(β1→	Vinogradov et al.(2010)
O104	→4)-DGal-(α1→4)-Neu5Ac7Ac9Ac-(α2→3)-DGal-(β1→3)-DGalNAc-(β1→	Gamian et al.(1992) Kogan et al.(1992)
O105	→3)-LRha-(α1→3)-LRha-(β1→4)-LRha-(α1→3)-DGlcNAc-(β1→ 　　　　　　　　　　　　　　　　　｜ DRib*f*-(β1→4)-DGlcA-(α1→2)	Knirel et al.(2016a)
O106	DGlc-(α1→6) 　　　｜ →6)-DMan-(α1→2)-DMan-(α1→2)-DMan-(β1→3)-DGlcNAc-(α1→ 　　　　　　　　　　　　　　｜ 　　　　　　　　　DGlc-(α1→4)	Shashkov et al.(2018)
O107	→4)-DGalNAc-(β1→3)-LRha-(α1→4)-DGlcNAc-(α1→4)-DGal-(β1→3)-DGalNAc-(α1→	Wang et al.(2009)
O108	→4)-8eLeg5Ac7Ac-(α2→6)-DGal-(α1→3)-LFucNAc-(α1→3)-DGlcNAc-(α1→	Perepelov et al.(2010b)
O109	→4)-DGlc-(α1→3)-6dLTal2Ac-(α1→3)-DGlcNAc-(β1→ 　　　　　　　　｜ LRhaNAc3NAc4Ac-(β1→3)	Perepelov et al.(2011e)
O110	→4)-DGalA6D*allo*Thr-(α1→4)-DGal-(α1→3)-DGal-(α1→3)-DGlcNAc-(β1→	Arbatsky et al.(1999) Perepelov et al.,(2013b)
O111	Col-(α1→6) 　　　｜ 　→4)-DGlc-(α1→4)-DGal-(α1→3)-DGlcNAc-(β1→ 　　　　｜ Col-(α1→3)	Eklund et al.(1978)
O112ab	→4)-DGalNAc-(α1→4)-DGlc-(α1→4)-LIdoA-(α1→3)-DGalNAc-(β1→ 　　　　　　｜ DGlcNAc-(β1→3)	Perepelov et al.(2008c)
O112ac	→4)-DGalNAc-(α1→4)-DGlc-(α1→4)-DGal-(β1→3)-DGalNAc-(β1→ 　　　　　　｜ DGlcNAc46*S*Pyr-(β1→3)	Perepelov et al.(2008d)
O113	→4)-DGalNAc-(α1→4)-DGalA-(α1→3)-DGal-(α1→3)-DGlcNAc-(β1→ 　　　　　　｜ DGal-(β1→3)	Parolis et al.(1995)
O114	→4)-DQui3NLSerAc-(β1→3)-DRib*f*-(β1→4)-DGal-(β1→3)-DGlcNAc-(α1→	Dmitriev et al.(1983)
O115	LRha-(β1→4) 　　　｜ 　→3)-DGlcNAc-(β1→4)-DGalA-(α1→3)-DMan-(α1→3)-DGlcNAc-(β1→	Fontana et al.(2013a)
O116	→2)-DQui4NAc-(β1→6)-DGlcNAc-(α1→4)-DGalNAc-(α1→4)-DGalA-(α1→3)-DGlcNAc-(β1→	Leslie et al.(1999)
O117	→4)-DGalNAc-(β1→3)-LRha-(α1→4)-DGlc-(α1→4)-DGal-(β1→3)-DGalNAc-(α1→	Wang et al.(2009)
O118	→3)-DRibitol-(5-*P*-6)-DGal-(α1→3)-LFucNAm-(α1→3)-DGlcNAc-(β1→	Liu et al.(2010)
O119	→2)-DMan-(β1→3)-DGal-(α1→4)-LRha-(α1→3)-DGlcNAc-(α1→ 　　　　　　｜ LRhaNAc3NFo-(β1→3)	Anderson et al.(1992) Perepelov et al.(2011e)
O120	LRha-(α1→4) 　　　｜ 　→3)-DGlcA-(α1→2)-LRha-(α1→2)-LRha-(α1→2)-DGal-(α1→3)-DGalNAc-(β1→	Perepelov et al.(2012b)
O121	→3)-DQui4NGlyAc-(β1→4)-DGalNAcA6N3Ac-(α1→4)-DGalNAcA-(α1→3)-DGlcNAc-(α1→	Parolis et al.(1997)
O123	→3)-DQui4NDAlaBu*S*3OH-(β1→6)-DGlcNAc-(α1→3)-LQuiNAc-(α1→3)-DGlcNAc6Ac-(α1→	Perepelov et al.(2010c)
O124	DGlc4*R*Lac-(β1→6)-DGlc-(α1→4) 　　　　　　　　　　　　　｜ 　　　　　　→3)-DGal-(β1→6)-DGal*f*-(β1→3)-DGalNAc-(β1→	Jonsson et al.(2006)
O125ab	→4)-DGalNAc-(β1→2)-DMan-(α1→3)-LFuc-(α1→3)-DGalNAc-(α1→ 　　　　　｜　　　　　　　　｜ DGal-(β1→3)　　　　DGlc-(α1→3)	Kjellberg et al.(1996)

Table 1. Continued

O126	→2)-DMan-(β1→3)-DGal-(β1→3)-DGlcNAc-(α1→3)-DGlcNAc-(β1→
	                │
	         LFuc-(α1→2)
		Larsson et al.(2004)

O127	→2)-LFuc3Ac/4Ac-(α1→2)-DGal-(β1→3)-DGalNAc-(α1→3)-DGalNAc-(α1→	Widmalm et al.(1993) Perepelov et al.(2010d)
O128ab[v]	→6)-DGal-(β1→3)-DGalNAc-(β1→4)-DGal-(α1→3)-DGalNAc-(β1→ │ LFuc-(α1→2)	Sengupta et al.(1995)
O128ab strain 87/D2	→6)-DGal3Ac/4Ac-(β1→3)-DGalNAc-(β1→4)-DGalNAc-(α1→3)-DGalNAc-(β1→ │ LFuc-(α1→2)	Ali et al.(2008)
O128ac[v]	→6)-DGal-(β1→3)-DGalNAc-(β1→4)-DGal-(α1→3)-DGalNAc-(β1→ │ LFuc-(α1→2)	Ali et al.(2008)

O129	→2)-LRha3Ac4Ac-(α1→2)-LRha-(α1→3)-LRha-(α1→3)-DGlcNAc-(β1→
	           │
	    DGlc-(α1→3)
		Perepelov et al.(2010a)

O130	→4)-DGal-(α1→6)-DGlc-(β1→3)-DGalNAc-(β1→ │ Gro-(2-P-4)-DGalNAc-(β1→3)	Perepelov et al.(2007a)
O131	→8)-Neu5Ac-(α2→6)-DGal-(β1→6)-DGal-(β1→3)-DGalNAc-(β1→	Perepelov et al.(2016a)
O132	→2)-DGal*f*-(α1→3)-LRha2Ac-(α1→4)-DGlc-(α1→2)-LRha-(α1→3)-DGlcNAc-(β1→	Shashkov et al.(2016b)

O133	→3)-LRha-(α1→2)-LRha-(α1→2)-LRha-(α1→4)-DGalA-(α1→3)-DGlcNAc-(α1→
	           │
	      DGlc-(α1→3)
		Shashkov et al.(2016c)

O134	→4)-DGlcA6LThr-(β1→6)-DGal-(β1→6)-DGlc-(β1→3)-DGalNAc-(β1→	Perepelov et al.(2015b)
O135	DGlc-(α1→6) │ →2)-LRha-(α1→2)-LRha-(α1→3)-LRha2Ac-(α1→3)-DGlcNAc-(β1→	Perepelov et al.(2010a)
O136	→4)-Pse5Ac7Ac-(β2→4)-DGal-(β1→4)-DGlcNAc-(β1→	Staaf et al.(1999b)
O137	DGlc-(β1→4) │ →6)-DGlcNAc-(α1→4)-DGlcA-(β1→4)-DGlcNAc-(β1→	Perepelov et al.(2016b)
O138	→2)-LRha-(α1→3)-LRha-(α1→4)-DGalNAcA-(α1→3)-DGlcNAc-(β1→	Linnerborg et al.(1997a)

O139	→3)-LRha-(α1→4)-DGalA-(α1→2)-LRha-(α1→3)-LRha-(α1→2)-LRha-(α1→3)-DGlcNAc-(α1→
	                   │
	          DGlc-(β1→3)
		Marie et al.(1998)

O140	DGal*f*-(α1→2)-LRha-(α1→4) │ →3)-DGal-(β1→4)-DGlc-(α1→4)-DGlcA-(β1→3)-DGalNAc-(β1→	Perepelov et al.(2015c)
O141	LRha-(α1→3) │ →4)-DMan-(α1→3)-DMan6Ac-(α1→3)-DGlcNAc-(β1→ │ DGlcA-(β1→2)	Färnbäck et al.(1998)

O142	→2)-LRha-(α1→6)-DGalNAc-(α1→4)-DGalNAc-(α1→3)-DGalNAc-(α1→
	            │
	      DGlcNAc-(β1→3)
		Landersjö et al.(1997)

O143	Gro-(2-N-6) │ →2)-DGalA3Ac/4Ac-(β1→3)-DGalNAc-(α1→4)-DGlcA-(β1→3)-DGlcNAc-(β1→	Landersjö et al.(1996)
O144[ii]		
O145	→4)-Neu5Ac-(α2→3)-LFucNAm-(α1→3)-DGlcNAc-(β1→	Feng et al.(2005)
O146	→2)-LRha-(α1→2)-LRha-(α1→4)-DGal-(β1→3)-DGalNAc-(β1→ │ DGlc3SLac-(α1→3)	Knirel et al.(2017)
O147	→2)-LRha-(α1→2)-LRha-(α1→4)-DGalA-(β1→3)-DGalNAc-(β1→	Blakeman et al.(1998)
O148	→3)-LRha-(α1→3)-LRha-(α1→2)-DGlc-(α1→3)-DGlcNAc-(α1→	Feng et al.(2007)
O149	→3)-DGlcNAc46SPyr-(β1→3)-LRha-(β1→4)-DGlcNAc-(β1→	Adeyeye et al.(1988)

O150	→3)-DGlcNAc4SLac-(β1→2)-LRha-(α1→2)-LRha-(α1→3)-LRha-(α1→3)-DGlcNAc-(β1→
	                 │
	          DGlc-(β1→2)
		Perepelov et al.(2007b)

O151	DGlcNAc-(β1→4) │ →2)-DRibitol-(5-P-6)-DGal-(α1→3)-LFucNAm-(α1→3)-DGlcNAc-(β1→	Liu et al.(2010)
O152	LRha-(β1→4) │ →3)-DGlcNAc-(α1→P-6)-DGlc-(α1→2)-DGlc-(β1→3)-DGlcNAc-(β1→	Olsson et al.(2005)

Table 1. Continued

O153	→2)-DRib*f*-(β1→4)-DGal-(β1→4)-DGlcNAc-(α1→4)-DGal-(β1→3)-DGlcNAc-(α1→	Ratnayake et al.(1994b)
O154	→2)-LRha-(α1→3)-LRha-(α1→3)-LRha-(α1→3)-DGalNAc-(α1→ | DManNAc-(α1→3)	Perepelov et al.(2013c)
O155	DGal-(α1→6) | →3)-DGal-(α1→3)-DGal-(β1→3)-DGalNAc-(β1→ | DGalA-(β1→4)	Fontana et al.(2013b)
O156	→4)-LFuc-(α1→6)-DGal-(α1→3)-LFuc-(α1→3)-DGlcNAc-(β1→ | DGal2Ac46*R*Pyr-(α1→3)	Duan et al.(2016)
O157	→2)-DRha4NAc-(α1→3)-LFuc-(α1→4)-DGlc-(β1→3)-DGalNAc-(α1→	Perry et al.(1986)
O158	DGlc-(β1→6) | →4)-DManNAc-(β1→4)-DGalA-(α1→3)-DGlcNAc-(β1→ | LRha3Ac/4Ac-(α1→3)	Perepelov et al.(2011f)
O159	LFuc-(α1→4) | →3)-DGlcNAc-(β1→4)-DGalA-(α1→3)-LFuc-(α1→3)-DGlcNAc-(β1→	Linnerborg et al.(1999a)
O160	DGlc-(β1→6) | →4)-DGlcNAc-(β1→3)-DGal-(α1→*P*-6)-DGal-(β1→3)-DGalNAc-(β1→	Perepelov et al.(2015d)
O161	→8)-Leg5Ac7D*Ala*-(α2→4)-DGlcA-(β1→3)-DGlcNAc-(β1→	Li et al.(2010)
O162	4dD*ara*Hex-(β1→3) 4dD*ara*Hex-(β1→3) | | →6)-DGlcNAc-(α1→4)-DGalNAc-(α1→, →4)-DGalNAc-(β1→4)-DGalNAc-(α1→	Shashkov et al.(2017)
O163	→2)-DMan-(β1→4)-DGlcA-(β1→3)-LQuiNAc-(α1→3)-DGlcNAc-(α1→	Perepelov et al.(2015e)
O164	DGlc-(β1→6)-DGlc-(α1→4) | →3)-DGal-(β1→6)-DGal*f*-(β1→3)-DGalNAc-(β1→	Linnerborg et al.(1999b)
O165	→8)-Pse5Bu*R*3OH7Ac-(α2→6)-DGal-(β1→4)-DGlc-(β1→3)-DGlcNAc-(α1→	Senchenkova et al.(2016c)
O166	→3)-DGal-(α1→6)-DGal-(α1→4)-DGalNAc-(α1→3)-DGalNAc-(β1→ | DGlc-(β1→3)	Ali et al.(2007)
O167	DGal*f*-(α1→4) | →2)-DGalA6N*L*Ala-(β1→3)-DGlcNAc-(α1→2)-DGal*f*-(β1→5)-DGal*f*-(β1→3)-DGlcNAc-(β1→	Linnerborg et al.(1997b)
O168	→4)-DGlcNAc6Ac-(α1→4)-DGlcA-(α1→3)-LFuc-(α1→3)-DGlcNAc-(β1→ | LFuc-(α1→3)	Perepelov et al.(2007c)
O169	DGlc-(β1→6) | →3)-DGal-(α1→6)-DMan-(α1→2)-DMan-(α1→3)-DGalNAc-(β1→ | DGlcA-(β1→4)	Perepelov et al.(2015f)
O170	→4)-DGlcNAc-(β1→4)-DGlcA-(β1→3)-DGal*f*-(β1→3)-DGlcNAc-(β1→ | DGal*f*-(β1→3)	Shashkov et al.(2015)
O171	→4)-Neu5Ac7Ac9Ac-(α2→6)-DGal-(β1→6)-DGlc-(β1→3)-DGal-(β1→3)-DGalNAc-(β1→	Ali et al.(2006)
O172	→3)-LFucNAc-(α1→4)-DGlc6Ac-(α1→*P*-4)-DGlc-(α1→3)-LFucNAc-(α1→3)-DGlcNAc-(α1→	Landersjö et al.(2001)
O173	LFuc-(α1→4) | →3)-DGlc-(α1→*P*-6)-DGlc-(α1→2)-DGlc-(β1→3)-DGlcNAc-(β1→	Linnerborg et al.(1999c)
O174ab	→4)-DGlcA-(β1→3)-DGal-(β1→3)-DGalNAc-(β1→ | DGlcNAc-(β1→2)	Fontana et al.(2012)
O174ac	→4)-DGlcA-(β1→3)-DGal-(β1→3)-DGlcNAc-(β1→ | DGlcNAc-(β1→2)	Duda et al.(2013)
O175	→2)-DGlc-(α1→4)-DGlcA-(α1→3)-DMan-(α1→2)-DMan-(α1→3)-DGalNAc-(β1→	Svensson et al.(2011a)
O176	→4)-DMan-(α1→2)-DMan-(α1→2)-DMan-(β1→3)-DGalNAc-(α1→	Olsson et al.(2008)
O177	→2)-LRha-(α1→3)-LFucNAc-(α1→3)-LFucNAc-(α1→3)-DGlcNAc-(β1→	Svensson et al.(2011b)
O178	→4)-DGlcNAc-(α1→4)-DGalA-(α1→2)-DRib*f*-(β1→4)-DGal-(β1→3)-DGalNAc-(β1→	Ali et al.(2005)

Table 1. Continued

O179[vi]	→4)-DGlcA-(α1→3)-DMan-(β1→4)-DMan-(β1→3)-DGlcNAc-(β1→ | DGlc-(β1→2)	Fontana et al. 29th ICS Lisbon 2018, Poster no BS2
O180	→4)-DManNAc3NAcA-(β1→2)-LRha-(α1→3)-LRha-(β1→4)-DGlcNAc-(α1→	Wang et al.(2012b)
O181	→4)-DGalNAc6Ac-(α1→6)-DGlc-(α1→P-4)-LQuiNAc-(α1→3)-DGlcNAc-(β1→ | LQuiNAc-(α1→3)	Fontana et al.(2015b)
O182	→4)-DGlcNAc-(β1→4)-DGlcA-(β1→3)-DGlcNAc-(β1→ | LRhaNAc3NAc-(β1→3)	Senchenkova et al.(2016d)
O183	DRibf-(β1→4)-DGlcA-(β1→4) | →3)-DGal-(α1→6)-DMan-(α1→2)-DMan-(α1→3)-DGalNAc-(β1→	Senchenkova et al.(2016d)
O184	→2)-DFuc3NBuR3OH-(α1→3)-DMan6Ac-(α1→3)-LFuc-(α1→3)-DGalNAc-(α1→	Senchenkova et al.(2016d)
O185	→4)-DGalNAc-(β1→4)-DGlcA-(β1→3)-DRibf-(β1→4)-DGal-(β1→3)-DGalNAc-(α1→	Senchenkova et al.(2016d)
O186	→3)-DQui4NDAlaBuS3OH-(β1→6)-DGlc-(α1→3)-LQuiNAc-(α1→3)-DGlcNAc6Ac-(α1→	Senchenkova et al.(2016d)
O187	→2)-DFuc3NAc-(β1→3)-DGlcNAc-(α1→4)-DGlcA-(α1→3)-DMan-(α1→2)-DMan-(α1→3)-DGlcNAc-(β1→	Senchenkova et al.(2016d)

[i]Acyl = Ac or BuR3OH. [ii] Lacking O antigen. [iii] Rough strain, no LPS containing O−antigen polysaccharide could be isolated. Unusual locations of the O antigen gene cluster. [iv] O antigen structure does not match the gene cluster. [v] Contains O−acetyl groups. [vi] Contains O−acetyl group.

General features of *E. coli* O antigen gene clusters

Escherichia coli O antigen gene clusters have between 5 and 18 genes and most are located between the *galF* and *gnd* genes. However, in O8 and O9/O9a, the O antigen gene clusters are located between the *gnd* and *hisI* genes (Kido *et al.* 1995; Amor and Whitfield 1997; Sugiyama *et al.* 1998) which is atypical in *E. coli*, but normal in *Klebsiella*. The O62 O antigen gene cluster is far from *galF* and *gnd*, and is proposed to originate from *E. coli* O68 by the acquisition of a novel O antigen gene cluster from *E. aerogene,* followed by an IS event resulting in repression of O68 O antigen synthesis (Hou *et al.* 2017). Three O serogroups (O34, O89 and O144) are rough (lacking a typical O antigen). However, in each case, an O antigen gene cluster was reported and has a full set of genes (Iguchi *et al.* 2015a) but is presumably not expressed. The O14 type strain has the O antigen gene cluster deleted apparently by homologous recombination between the *manB* genes of the colanic acid and O antigen gene clusters, thereby removing the genes between them (Jensen and Reeves 2004) but the LPS includes an O antigen that has the structure of the ECA. The same situation was observed in other O14 strains that have the R1, R4 or K-12 outer core (Jensen and Reeves 2004; Kuhn, Meier-Dieter and Mayer 1988). It appears that the R1, R4 and K-12 WaaL ligases will add ECA repeat units to the outer core if the O antigen is not present. There is no O antigen gene cluster between *galF* and *gnd* in O57, although this serogroup is not rough (Naumenko *et al.* 2018b). For O20 and O60, their O-unit structures do not fit the genetic region flanked by *galF* and *gnd*. We therefore anticipate that the functional gene clusters for O20, O57 and O60 are elsewhere on the chromosome. In this review, we discuss the O antigen gene clusters of 178 serogroups, excluding O14, O20, O34, O57, O60, O89 and O144.

The GC% contents of the gene clusters are lower than that of the whole genome. Among the three classes of genes, the GC% for O unit processing genes and glycosyltransferase genes is lower than for nucleotide sugar synthesis genes. Most (93%) of the *E. coli* O antigens use the Wzx/Wzy-dependent process (gene clusters contain *wzx* and *wzy* genes) for the synthesis and translocation of O antigens and a few, including O8, O9/O9a, O52, O62, O92, O95, O97, O99, O101 and O162 use the ABC transporter pathway (gene clusters contain *wzm* and *wzt* genes). Many O antigens of the latter group are linear homopolysaccharides such as O8, O9/O9a and O92 or have a homopolysaccharide backbone with two monosaccharide side-branches per O unit (O62, O97 and O99). The others contain a short heterodisaccharide O unit (O52 and O95) or O-units of two types (O101 and O162) with a single monosaccharide side-branch in O162.

Genes for the biosynthesis of several monosaccharides including UDP-D-Glc, UDP-D-Gal, UDP-D-GlcNAc, UDP-D-GlcA, UDP-D-GalA, NDP-D-Ribf and NDP-D-Xulf are not found in the O antigen gene clusters: these sugars are also present in other *E. coli* structures, and the genes are elsewhere in the chromosome. The genes for biosynthesis of the sugars that are present in the gene clusters are described below. For most of the Wzx/Wzy-dependent O antigens the number of glycosyltransferase genes in the gene cluster is one less than the number of residues in the O unit, as expected because the initial GlcNAc or GalNAc is added by WecA. For the ABC transporter-dependent O antigens, the initial GlcNAc is present only once as a primer, followed by an adaptor region and then the O antigen (O8/O9/O9a). In some cases, the presence of fewer GTs than expected can be accounted for by the presence of D-Glc side-branch residues, as these are commonly added when the O antigen is in the periplasm by the Gtr process, involving three genes (*gtrA/B/C*). The genes are commonly clustered in a prophage genome. GtrA and GtrB are highly conserved among serogroups, and GtrC appers to be unique to each serogroup and is referred to as the serogroup-specific glucosyl transferase. For *S. flexneri*, a model has been proposed showing how the O antigen glucosylation occurs. Overall, GtrB first carries UDP-Glc to the bactoprenol carrier to yield Und-P-Glc precursor in the cytoplasm, then GtrA is thought to flip Und-P-Glc from cytoplasm to the periplasm. In the final step, GtrC attaches Und-P-Glc specifically to the appropriate sugar of the O-unit and recycles the lipid carrier to the

Table 2. Topology of the O antigen repeating units.

Topology	Abundance (%)
2-residue backbone	4
	3
	1
	<1
3-residue backbone	26
	7
	10
	3
	4
	2
	<1
4-residue backbone	54
	20
	28
	2
	4
	<1
5-residue backbone	14
	11
	3
6-residue backbone	1
	<1
	<1
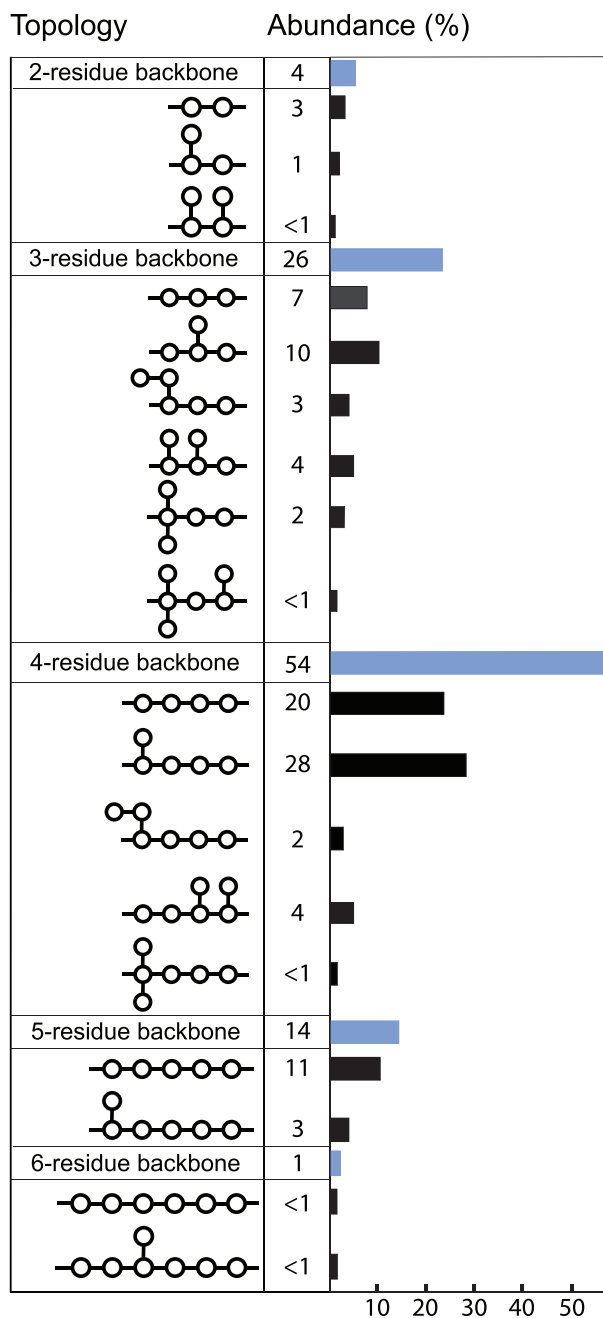
cytoplasm (Allison and Verma 2000; Korres and Verma 2004; Korres *et al.* 2005). Only 10 of the 20 *E. coli* O units with a D-Glc side-branch have a potential D-Glc glycosyltransferase gene in the O antigen gene cluster and the others are either known or proposed to be attached by the Gtr process involving genes elsewhere in the chromosome. In addition, O units with two D-Man residues present in the O unit have a glycosyltransferase that can transfer both, a situation found in the O9/O9a (Kido and Kobayashi 2000), O17/O44/O73/O77 (Wang *et al.* 2007a) and O176 (Shashkov *et al.* 2018) gene clusters. Third, in O units with more than one L-Rha residue, one glycosyltransferase is probably responsible for the formation of different linkages between

two L-Rha residues, a situation found in the O1, O2, O13, O19, O35, O50, O99, O129, O135, O139 and O150 gene clusters (Perepelov *et al.* 2007b; Perepelov *et al.* 2009b; Perepelov *et al.* 2010a; Perepelov *et al.* 2011b; Rundlöf, Weintraub and Widmalm 1998; Yang *et al.* 2018). However, several other O antigen gene clusters also have more or fewer GTs than expected with no obvious explanation: serogroups O5, O108, O111, O120, O147, O153 and O178 have one less glycosyltransferase than expected, while the O52, O143 and O167 O antigen gene clusters have one extra glycosyltransferase.

For the ABC transporter-dependent pathway capping of the non-reducing terminal residue of the polysaccharide is important for chain-termination (Mann, Kimber and Whitfield 2019); the *E. coli* gene clusters of O52, O60, O89, O97, O101 and O162 have a putative methyltransferase gene, which is proposed to be responsible for the termination of the O polysaccharide (Jansson *et al.* 1985; Senchenkova *et al.* 1996; Vinogradov *et al.* 2002; Feng *et al.* 2004), as there is not any methyl group in the O-unit structure. In the LPS of *Pseudomonas syringae* pv. phaseolicola GSPB 1552 the backbone homopolysaccharide consisting of D-rhamnopyranosyl residues is partially O-methylated at O3 of one of the sugar residues in the repeating unit and in addition also at the terminal sugar residue of the polymer (Zdorovenko *et al.* 2001). Whether the O-methylation at the terminal residue is important for chain-termination in the same way as for the *E. coli* O8 O antigen remains to be elucidated. Besides the chemical degradation approach complemented by methylation analysis and NMR spectroscopy used in the study one should also be able to differentiate internal and terminal O-methylation patterns by ^{13}C NMR relaxation studies that can reveal differential dynamics between terminal and internal residues. (Lycknert and Widmalm 2004).

Most of the ORFs in the O antigen gene clusters could be assigned functions based on similarity to genes in available databases (Fig. S1, Supporting Information). Gene names are used as previously assigned, or given according to the bacterial polysaccharide gene nomenclature (BPGN) system; see also the Bacterial Surface Polysaccharide database (BSPdb.nankai.edu.cn).

A total of 68 anomalies are found in 178 serogroups, which include 17 for unusual sugar biosynthesis gene location, five for unusual location gene or transcriptional direction of some other genes, 23 for presence of mobile elements, nine for presence of gene remnant(s) and 14 for presence of non-coding regions. These anomalies usually indicate that the gene cluster involved has undergone a relatively recent genetic event (Liu *et al.* 2008). The proportion of gene clusters with anomalies is similar to that in *Salmonella* (32%), but less than in *Shigella* (52%), indicating that *Shigella* is undergoing more rapid change in O antigens than *Salmonella* and typical *E. coli*. For details of these anomalies and O antigen structures of *Shigella* (Table S1), (Knirel *et al.* 2016a, Knirel *et al.* 2018) see supplementary information (SI).

Sugar biosynthetic pathways

The proposed or characterized biosynthetic pathways of 38 different sugars found in *E. coli* O antigens are shown in Scheme 1. D-GlcNAc, D-Gal, D-Glc, D-GlcA, D-GalA, D-Rib*f* and D-Xul*f* are also found in other structures in *E. coli* and the biosynthesis genes are at conserved loci elsewhere in the chromosome, but those responsible for synthesis of the other sugars are located in the O antigen gene clusters. The precursors of the sugar residues in the O unit are all transferred from nucleotide diphospho (NDP)

Table 3. Sugar composition of the *E. coli* O antigens.

Component	Abbreviation
colitose (3,6-dideoxy-L-*xylo*-hexose)	Col
4-deoxy-D-*arabino*-hexose	4d-D-*ara*Hex
6-deoxy-D-*manno*-heptose	6d-D-*man*Hep
D-fructofuranose	D-Fru*f*
3-acetamido-3-deoxy-D-fucose	D-Fuc3NAc
3-[(R)-3-hydroxybutanoylamino]-3,6-dideoxy-D-galactose	D-Fuc3NBu*R*3OH
4-acetamido-4,6-dideoxy-D-galactose	D-Fuc4NAc
D-fucofuranose	D-Fuc*f*
2-acetamido-2-deoxy-D-fucose	D-FucNAc
D-galactose	D-Gal
D-galacturonic acid	D-GalA
D-galacturonamide	D-GalAN
D-galactofuranose	D-Gal*f*
2-acetamido-2-deoxy-D-galactose	D-GalNAc
2-acetamido-2-deoxy-D-galacturonic acid	D-GalNAcA
2-acetamido-2-deoxy-D-galacturonamide	D-GalNAcAN
D-glucose	D-Glc
D-glucuronic acid	D-GlcA
2-acetamido-2-deoxy-D-glucose	D-GlcNAc
D-mannose	D-Man
2-acetamido-2-deoxy-D-mannose	D-ManNAc
2,3-diacetamido-2,3-dideoxy-D-mannuronic acid	D-ManNAc3NAcA
3-acetamido-3,6-dideoxy-D-glucose	D-Qui3NAc
3-[(R)-3-hydroxybutanoylamino]-3,6-dideoxy-D-glucose	D-Qui3NBu*R*3OH
3-[(2R,3R)-2-acetamido-3-hydroxybutanoylamino]-3,6-dideoxy-D-glucose	D-Qui3ND*allo*ThrAc
4-acetamido-4,6-dideoxy-D-glucose	D-Qui4NAc
4-[(R)-3-hydroxybutanoylamino]-4,6-dideoxy-D-glucose	D-Qui4NBu*R*3OH
4-[(S)-3-hydroxybutanoylamino]-4,6-dideoxy-D-glucose	D-Qui4NBu*S*3OH
4-[(S)-3-hydroxybutanoyl-D-alanylamino]-4,6-dideoxy-D-glucose	D-Qui4NDAlaBu*S*3OH
D-rhamnose (6-deoxy-D-mannose)	D-Rha
4-acetamido-4,6-dideoxy-D-mannose	D-Rha4NAc
D-ribofuranose	D-Rib*f*
D-*threo*-pent-2-ulose	D-Xul*f*
6-deoxy-L-talose	6d-L-Tal
L-fucose (6-deoxy-L-galactose)	L-Fuc
2-acetamido-2-deoxy-L-fucose	L-FucNAc
2-acetimidoylamino-2,6-dideoxy-L-galactose	L-FucNAm
L-iduronic acid	L-IdoA
2-acetamido-2-deoxy-L-quinovose (2-acetamido-2,6-dideoxy-L-glucose)	L-QuiNAc
L-rhamnose (6-deoxy-L-mannose)	L-Rha
2-acetamido-2-deoxy-L-rhamnose (2-acetamido-2,6-dideoxy-L-mannose)	L-RhaNAc
2,3-diacetamido-2,3,6-trideoxy-L-mannose	L-RhaNAc3NAc
2-acetamido-2,3,6-trideoxy-3-formamido-L-mannose	L-RhaNAc3NFo
5-acetamido-7-(D-alanylamino)-3,5,7,9-tetradeoxy-D-*glycero*-D-*galacto*-non-2-ulosonic acid	Leg5Ac7DAla
5,7-diacetamido-3,5,7,9-tetradeoxy-L-*glycero*-D-*galacto*-non-2-ulosonic (di-*N*-acetyl-8-epilegionaminic) acid	8eLeg5Ac7Ac
N-acetylneuraminic acid	Neu5Ac
5,7-diacetamido-3,5,7,9-tetradeoxy-L-*glycero*-L-*manno*-non-2-ulosonic (di-*N*-acetylpseudaminic) acid	Pse5Ac7Ac
7-acetamido-3,5,7,9-tetradeoxy-5-[(R)-3-hydroxybutanoylamino]-L-*glycero*-L-*manno*-non-2-ulosonic acid	Pse5Bu*R*3OH7Ac

derivatives such as dTDP, GDP, UDP and CDP, or a monophospho derivative CMP.

(i) dTDP sugar biosynthesis

Seven dTDP sugars are found in *E. coli* O antigens, including dTDP derivatives of L-Rha, D-Qui3NAc, D-Qui4NAc, 6d-L-Tal, D-Fuc3NAc, D-Fuc4NAc, D-Fuc*f*. All but the dTDP-D-Fuc4NAc pathway have been determined experimentally.

(ii) UDP sugar biosynthesis

Twenty UDP sugars are found in *E. coli* O antigens. Some proposed pathways including those for UDP-L-IdoA, UDP-L-FucNAm, UDP-D-FucNAc, UDP-D-GalNAcAN, UDP-L-RhaNAc3NAc, UDP-D-GalA6N and UDP-L-RhaNAc3NFo are not based on experimental data but are the only likely pathway

for the enzymes encoded in the gene cluster, and those for UDP-D-GalNAcAN and UDP-D-GalA6N are proposed in this study for the first time.

(iii) GDP sugar biosynthesis

Five GDP sugars including GDP derivatives of D-Man, L-Fuc, D-Rha, Col and D-Rha4NAc are found in *E. coli* O antigens. All of these GDP-sugar biosynthetic pathways have been identified, and all involve the enzymes encoded by *manABC* and *gmd* genes. However, manA is not located in the O antigen gene cluster (Samuel and Reeves 2003).

(iv) NDP-D-Rib*f* and NDP-D-Xul*f* sugar and ribitol-5-phosphate biosynthesis

RmlC → RmlD → dTDP-L-Rha
RmlC → T11 → dTDP-6d-L-Tal
Glc-1-P → RmlA → RmlB
FdtA → FdtB → FdtC → dTDP-D-Fuc3NAc
QdtA → QdtB → QdtC → dTDP-D-Qui3NAc
FdfA → FdfB → dTDP-D-Fuc4NAc
VioA → VioB → dTDP-D-Qui4NAc
FcfA → dTDP-D-Fuc → FcfB → dTDP-D-Fuc*f*

Fru-6-P → ManA[a] → ManB → ManC → GDP-Man → Gmd
Fcl → GDP-L-Fuc
ColA → ColB → GDP-L-Col
Rmd → GDP-D-Rha
PerA → PerB → GDP-D-Rha4NAc

Glc-1-P → GalU[a] → UDP-D-Glc
GalE → UDP-D-Gal → Glf → UDP-D-Gal*f*
Ugd → UDP-D-GlcA
Gla → UDP-D-GalA → WceT → UDP-D-GalA6N
Gae → UDP-L-IdoA

GlcNAc-1-P → GlmU[a] → UDP-D-GlcNAc
MnaA → UDP-D-ManNAc
Gne → UDP-D-GalNAc → GnaA → UDP-D-GalNAcA → WbpS → UDP-D-GalNAcAN
FnlA → FnlB → FnlC → UDP-L-FucNAc → WbuX,WbuY,WbuZ[b] → UDP-L-FucNAm
QnlA → UDP-L-RhaNAc → QnlB → UDP-L-QuiNAc
WbhS → WbhP → UDP-D-FucNAc
GnaA → MndA → MndB → MndC → MndD → UDP-D-ManNAc3NAcA
WeiP/WekE,WeiQ/WekF,WeiS/WekG
WeiO → UDP-L-RhaNAc3NAc
WekD → UDP-L-RhaNAc3NFo
Elg1 → Elg2 → Elg5 → Elg3 → Elg6 → Elg4 → Elg7 → CMP-8eLeg5Ac7Ac
Lea1 → Lea2 → Lea3 → Lea5 → Lea6 → Lea4 → Lea7 → CMP-Leg5Ac7Ala
Psb1 → Psb2 → Psb4 → Psb6 → Psb3 → CMP-Pse5Ac7Ac
NnaA → NnaB,NnaD → NnaC → CMP-Neu5Ac

ribulose-5-P → Rib → Ribitol-5-p → Orf[a] → CDP-D-Rib

Sedoheptulose-7-P → GmhA → HddA → GmhB → HddC → DmhA → DmhB → GDP-6d-D-*man*Hep

Scheme 1. Biosynthesis pathways for the sugars in *E. coli* O antigens. RmlA, glucose-1-phosphate thymidylyltransferase (Zuccotti *et al* 2001); RmlB, dTDP-D-glucose 4,6-dehydratase (Allard *et al.* 2001); RmlC, dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase (Giraud *et al.* 1999a); RmlD, dTDP-6-deoxy-L-mannose-dehydrogenase (Giraud *et al.* 1999b); ManA, phosphomannose isomerase; ManB, phosphomannomutase; ManC, mannose-1-phosphate guanylyltransferase (Samuel and Reeves 2003); Psb1, C6 dehydratase/C5 epimerase; Psb2, aminotransferase; Psb4, nucleotidase; Psb6, condensase; Psb3, cytidylyltransferase; FnlA, 4,6-dehydratase, 3- and 5-epimerase; FnlB, reductase; FnlC, C-2 epimerase (Kneidinger *et al.* 2003b); GalU, UTP-glucose-1-phosphate uridylyltransferase (Bonofiglio, Garcia and Mollerach 2005); GlmU, UDP-N-acetyl-glucosamine pyrophosphorylase (Mengin-Lecreulx and van Heijenoort 1993); GalE, UDP-glucose-4-epimerase (Samuel and Reeves 2003); Glf, UDP-galactopyranose mutase (Nassau *et al.* 1996); Gna, UDP-GalNAcA synthetase (Zhao *et al.* 2000); Gne, UDP-N-acetylglucosamine-4-epimerase (Bengoechea *et al.* 2002); Gae, C5-epimerase; Gla, UDP-galacturonatenase (Munoz *et al.* 1999); Ugd, UDP-glucose 6-dehydrogenase (Stevenson *et al.* 1996); QnlA, dTDP-4-dehydrorhamnose reductase; QnlB, C-2 epimerase (Kneidinger *et al.* 2003); Gmd, GDP-mannose-4,6-dehydratase (Somoza *et al.* 2000; Kneidinger *et al.* 2001); Fcl, GDP-L-fucose synthetase (Rosano *et al.* 2000); VioA, aminotransferase (Wang *et al.* 2007); VioB, N-acetyltransferase (Wang *et al.* 2007); ColA, GDP-4-keto-6-deoxy-D-mannose 3-dehydrase (Alam, Beyer and Liu 2004); ColB, GDP-colitose synthase (Alam, Beyer and Liu 2004); PerA, GDP-perosamine synthetase (Zhao *et al.* 2007b; Albermann and Beuttler 2008); PerB, GDP-perosamine N-acetyltransferase (Albermann and Beuttler 2008); Rib, ribulose 5-phosphate reductase/CDP-ritbitol pyrophosphorylase (Follens *et al.* 1999); FdtA, dTDP-6-deoxy-hex-4-ulose isomerase; FdtB, dTDP-6-deoxy-D-xylo-hex-3-ulose aminase; FdtC, dTDP-D-Fuc3N acetylase (Pfoestl *et al.* 2003); QdtA, dTDP-4-oxo-6-deoxy-D-glucose3,4-oxoisomerase; QdtB, dTDP-3-oxo-6-deoxy-D-glucose aminase (Pfostl *et al.* 2008); QdtC, dTDP-D-Qui3N acetylase (Pfostl *et al.* 2008); MnaA, UDP-N-acetylglucosamine-2-epimerase (Campbell *et al.* 2000); FcfA, C-4 reductase; FcfB, mutase; GmhA, Glyceromannoheptose-7-P isomerase; HddA, D-Heptose 7-phosphate kinase; GmhB, D-Heptose 1,7-biphosphate phosphatase; HddC, D-Heptose 1-phosphate guanosyltransferase; DmhA, NAD-dependentepimerase/dehydratase; DmhB, NAD-dependent epimerase/dehydratase; FdfA, aminotransferase; FdfB, acetyltransferase; Rmd, oxidoreductase; WbpS, aminotransferase; WbuX, aminotransferase; wbhS, UDP-D-GlcNAc 4,6-dehydratase; wbhP, NAD-dependent epimerase/dehydratase; GnaA, UDP-GlcNAc 6-dehydrogenase; MndA, oxidoreductase; MndB, acetyltransferase; MndC, aminotransferase; MndD, UDP-N-acetylglucosamine 2-epimerase; WeiP/WekE, aminotransferase; WeiQ/WekF, aminotransferase; WeiS/WekG, dehydratase/epimerase; WeiO, acetyltransferase; WekD, formyltransferase; NnaA, UDP-N-acetylglucosamine-2-epimerase; NnaB, N-acetylneuraminic acid synthetase; NnaC, CMP-N-acetylneuraminic acid synthetase; NnaD, N-acetylneuraminic acid synthetase; Elg1, UDP-N-acetylglucosamine 4,6-dehydratase; Elg2, aminotransferase; Elg3, UDP-N-acetylglucosamine 2-epimerase; Elg4, N-acetyl-neuraminic acid synthetase; Elg5, acetyltransferase; Elg6, sugar-phosphate nucleotide transferase; Elg7, cytidylyl-transferase; Lea1, C6 dehydratase; Lea2, aminotransferase; Lea3, alanyltransferase; Lea4, N-acetylneuraminic acid synthetase; Lea5, UDP-N-acetylglucosamine 2-epimerase; Lea6, sugar-phosphate nucleotide transferase; Lea7, CMP-N-acetylneuraminic acid synthetase; TarI, cytidylyltransferase. [a]The enzyme is encoded by the gene which is not located in the O antigen gene cluster. [b]The enzyme could not be assigned functionally.

D-Rib*f* is present in nine serogroups (O5, O20, O54, O105, O114, O153, O178, O183 and O185) and Xul*f* is present in O95 and O97; however, the existence of a nucleotide-activated Rib*f* has not been demonstrated and genes responsible for its nucleotide precursor synthesis are not found in either O antigen gene cluster. The *rib* gene, responsible for the synthesis of ribitol-5-phosphate from ribulose-5-phosphate is present in O118 and O151 O antigen gene clusters, which contain ribitol-5-phosphate in their O antigens.

(v) Biosynthesis of CMP-Neu5Ac and derivatives

Neu5Ac and derivatives such as Neu5Ac,7Ac,9Ac are present in five O antigens (O24, O56, O104, O145 and O171), and *nnaABCD* genes responsible for CMP-Neu5Ac biosynthesis are present in all five gene clusters.

(vi) CMP-8eLeg5Ac7Ac, CMP-Leg5Ac7Ala and CMP-Pse5Ac7Ac biosynthesis

8eLeg5Ac7Ac, Leg5Ac7Ala and Pse5Ac7Ac are present in the O108, O161 and O136 O antigens, respectively. Genes for biosynthesis of the first two were proposed in our previous studies (Li *et al.* 2010; Perepelov *et al.* 2010b), and is proposed in this study for the third.

(vii) Biosynthesis of GDP-6d-D-*man*Hep

6d-D-*man*Hep is a heptose present only in the O52 O antigen. Genes including *dmhAB*, *hddAC* and *gmhAB* in the O52 O antigen gene cluster are involved in the synthesis of its GDP derivative. However, the function of *dmhA* and *dmhB* has not been confirmed biochemically (Valvano, Messner and Kosma 2002; Feng *et al.* 2004)

Classification of GTs

GTs can be divided into groups in various ways, as in the frequently employed Pfam (Protein Families) and Cazy (Carbohydrate-Active enZYmes) databases. However, they provide limited guidance for allocating a GT to a specific linkage, and only rarely is it possible to do the assignment with confidence from sequence comparisons alone (Stevenson, Dieckelmann and Reeves 2008). Biochemical analysis of GTs are critical to determine their specific functions in O antigen assembly, but this is still difficult as the dinucleotide-sugar precursor often is not readily available and the acceptor is usually an incomplete O unit, almost never readily available (Stevenson, Dieckelmann and Reeves 2008; Czuchry, Szarek and Brockhausen 2018). Only a few of the GTs have been identified using biochemical methods or genetic methods, including WbbD in O7 (Riley *et al.* 2005), WbdA in O8 and O9/O9a (Greenfield *et al.* 2012a) (Kido and Kobayashi 2000), WbdD in O9/O9a (Clarke *et al.* 2011), WbgO in O55 (Liu *et al.* 2009), WfaP in O56 (Brockhausen *et al.* 2008), WcmABCD in O86 (Yi *et al.* 2005), WclY in O107 and O117 (Wang *et al.* 2012a), WbdM and WbdN in O111 (Stevenson, Dieckelmann and Reeves 2008), WbsJ in O128 (Li *et al.* 2008), WbdN in O157 (Gao *et al.* 2012), WfgD in O152 (Brockhausen *et al.* 2008), WfcD in O141 (Chen *et al.* 2017) and WbwABC in O104 (Wang *et al.* 2014; Czuchry *et al.* 2015; Czuchry, Szarek and Brockhausen 2018).

There is also limited analysis of the substrate range of the GTs found in the gene clusters. However, it is common to find clearly related GTs responsible for identical or similar reactions. It is therefore very useful to identify the GTs that fit into such a group. The GTs in *E. coli* O antigen gene clusters were analyzed using OrthoMCL (Li *et al.* 2003) with a cut-off of $1e^{-50}$ to assemble all the GTs into homology groups (HGs); 621 GTs from 178 O serogroups were annotated, and among them, 491 could be classified into 106 HGs (Table S2, Supporting Information). The functions of 155 of them can be predicted based on correlations between presence of GTs with a similar protein sequence and a shared or similar structural element in the corresponding O antigens.

GTs belonging to the same HG often have or are proposed to have identical or very similar functions. For example, the 18 GTs in HG5 share 51% to 99% identity in pair-wise comparisons. Of these, 13 WbuB are responsible for a L-Fuc*p*NAc-(α1→3)-D-Glc*p*NAc linkage, and two WbuB (O118, O145) are responsible for the similar L-Fuc*p*NAm-(α1→3)-D-Glc*p*NAc linkage. The remaining WbwW in O98, WbuB in O181 and WbwH in O123 are responsible for the transfer of L-Qui*p*NAc or L-Qui*p*NAcyl to form a linkage to D-Glc*p*NAc. All 18 form an α-(1→3)-linkage between sugars related to D-GlcNAc. Although the above groups of related GTs within an HG perform identical or almost identical functions, they were given different names. We therefore recommend that the genes within each such HG be given a single name based on established function and/or publication priority.

We noted above that having most of the O antigen gene clusters at a single locus facilitates transfer of the gene clusters between strains by homologous recombination in the flanking genes. The capacity for serotype substitution has the advantage that in one step a strain can disable an immune response to these immunodominant O antigens, and also gain a replacement O antigen to avoid the problems faced by a rough organism. It has also often been observed that GT genes are commonly present in inverse order of their function order, with the gene for the last sugar to be added being the first GT gene in the cluster, and so on. We used the large number of *E. coli* GTs with a putative function for a quantitative analysis, and found that 69 (39%) of the 178 suitable gene clusters have their GTs in reverse order to function. Kenyon *et al.* observed that the gene order also generally keeps the strain-specific genes in any pairwise combinations in a single block, which maintains divergence-level symmetry (Kenyon, Cuneen and Reeves 2017). A detailed description is given in supplementary information including Fig. S2 (Supporting Information).

Formation and expansion of O antigen variety in *E. coli*

Most of the O antigen gene clusters in *E. coli* and related species are at the *galF-gnd* locus, and transfer of the gene clusters between strains can occur by homologous recombination in the flanking genes, which will substitute one O antigen for another. IS elements appear to play an important role in mediating intraspecies and interspecies O antigen gene transfer that is not based on homology, and also plays a role in the diversification of O antigen forms by gene inactivation. Temperate phages play an important role by carrying genes responsible for O antigen modifications that are expressed when the phage is present in the bacterial genome. In addition, there are several cases of amino acid substitutions in key genes resulting in new O antigen forms. Analysis of the *E. coli* O antigen gene clusters shows that such genetic events are present in several groups of closely related O antigen gene clusters (Fig. 1).

(i) *E. coli* O8, O9/O9a. The O8, O9 and O9a O antigens are a family of related linear mannopyranose homopolymers, which have become prototypes for O antigen synthesis by the ABC transporter pathway. The O antigens of *E. coli* O9a and O8 are identical to those of *K. pneumoniae* O3 and O5 O-PSs, respectively. The genetic loci encoding the corresponding O antigens' biosynthesis enzymes are highly conserved, and their close evolutionary relationships reflect lateral gene transfer between the two species (Sugiyama *et al.* 1998; Reeves and Wang 2002). WbdC has been shown to possess identical activities in O8 and O9a, and this is the case for WbdB. WbdC is a GDP-Man:Glc*p*NAc-PP-Und-(α1→3)-mannosyltransferase and WbdB is an α-(1→3) mannosyltransferase that installs, in succession, two Man residues onto the disaccharide WbdC product α-Man*p*-(1→3)-β-Glc*p*NAc-PP-Und, forming the adaptor region. However, WbdAs from O8 and O9a share only 16% identity and differ in size, and experimental evidence indicates that WbdA is the serogroup-defining mannosyltransferase. WbdA (O8) has three predicted glycosyltransferase modules and polymerizes a trisaccharide repeat unit containing single α-(1→3)-, α-(1→2)-, and β-(1→2)-linked mannose residues. In contrast, WbdA (O9a) has two predicted

Figure 1. O antigen gene clusters and structures that are identical or closely related within *E. coli* (A) KO5 and O8; KO3, O9 and O9a; O157, O24 and O56; O169 and O183; (B) O86, O127 and O90; O107 and O117; O124 and O164; O118 and O151; (C) O2 and O50; O101 and O162; O17, O44, O73, O77 and O106. For gene key, see Supplementary Fig. S1.

O2 — *rmlB rmlD rmlA fdtA fdtB* (fdtC) *wzx wekP rmlC wzy wekQ wbuV* ISAs1 *wekS*
DNA identity %: 100 100 100 100 99 100 100 100 99 100 100
Protein identity %: 100 100 100 100 99 100 100 100 99 100 100

WekQ   WekQ   WbuV
→3)-LRha-(α1→2)-LRha-(α1→3)-LRha-(β1→4)-DGlcNAc-(β1→
|WekP
DFuc3NAc-(α1→2)

O50 — *rmlB rmlD rmlA fdtA* (fdtC) *fdtB wzx wekP rmlC wzy wekQ wbuV* ISAs1 ISAs1 ISAs1 ISAs1

WekQ   WekQ   WbuV
→3)-LRha-(α1→2)-LRha-(α1→3)-LRha-(β1→4)-DGlcNAc-(β1→

O101 — *gne werR wzm wzt wekV werS werW werT gnd* IS4 *wbxC werU wbxD wbxE werV wexK ugd*
DNA identity %: 100 99 99 99 99 100 100 99 100 / 100 100 100 100 / 100 100
Protein identity %: 100 99 100 100 99 100 100 99 100 / 100 100 100 100 / 100 100

→6)-DGlcNAc-(α1→4)-DGalNAc-(α1→;
→4)-DGalNAc-(β1→4)-DGalNAc-(α1→

O162 — *gne werR wzm wzt wekV werS wbxB werT gnd wbxC werU wbxD wbxE werV wexK ugd*

4dDaraHex-(β1→3)
|
→6)-DGlcNAc-(α1→4)-DGalNAc-(α1→
4dDaraHex-(β1→3)
|
→4)-DGalNAc-(β1→4)-DGalNAc-(α1→

O17 — *wbaC wbaD manC manB wzy wzx*

DGlc-(α1→6)
WbaC   WbaC   WbaD |
→6)-DMan-(α1→2)-DMan-(α1→2)-DMan-(β1→3)-DGlcNAc-(α1→

O44 — *wbaC wbaD manC manB wzy wzx*
DNA identity %: 99 100 99 99 99 99
Protein identity %: 99 100 99 99 99 99

DGlc-(α1→4)
WbaC   WbaC   WbaD
→6)-DMan-(α1→2)-DMan-(α1→2)-DMan-(β1→3)-DGlcNAc-(α1→

O73 — *wbaC wbaD manC manB wzy wzx*
DNA identity %: 100 100 99 99 99 99
Protein identity %: 100 100 99 99 99 99

DGlc-(α1→4)
WbaC | WbaC   WbaD
→6)-DMan-(α1→2)-DMan-(α1→2)-DMan-(β1→3)-DGlcNAc-(α1→
|
DGlc-(α1→3)

O77 — *wbaC wbaD manC manB wzy wzx*
DNA identity %: 99 100 99 99 99 99
Protein identity %: 99 100 99 99 99 99

WbaC   WbaC   WbaD
→6)-DMan-(α1→2)-DMan-(α1→2)-DMan-(β1→3)-DGlcNAc-(α1→

O106 — *wbaC wbaD manC manB wzy wzx*

DGlc-(α1→6)
WbaC   WbaC   WbaD |
→6)-DMan-(α1→2)-DMan-(α1→2)-DMan-(β1→3)-DGlcNAc-(α1→
|
DGlc-(α1→4)

Figure 1. Continued.

glycosyltransferase modules and polymerizes a tetrasaccharide repeat unit containing two α-(1→2)- and two α-(1→3)-linked mannose residues (Greenfield *et al.* 2012b). Functional analysis of chimeric genes and site-directed mutagenesis of WbdA in the O9 isolate showed that a single amino acid substitution (R55C) is responsible for O9a having one less mannose residue in an (α1→2)-linkage (Kido and Kobayashi 2000). Similarly, the glycosyltransferase WejI in *E. coli* O99 and WbrX in *E. coli* O52, also employing the ABC transporter pathway for their O antigen synthesis, are predicted bifunctional enzymes (Feng *et al.* 2004; Perepelov *et al.* 2009b), but further investigation is needed. In addition, WbdD (O9/O9a) is a bifunctional kinase-methyl transferase that adds a phospho-methyl group to the O9/O9a non-reducing terminus (Clarke *et al.* 2011), whereas WbdD (O8) adds only a methyl group in O8 (Clarke, Cuthbertson and Whitfield 2004), both halting polymerization and regulating the O antigen chain length.

(ii) *E. coli* O24 and O56. The O24 and O56 O antigens have similar tetrasaccharide repeat units but have D-Glc/D-Gal and D-GlcNAc/D-GalNAc substitutions. Comparison of the two O antigen gene clusters showed that the O24 O antigen gene cluster arose from that of O56 through inactivation of two genes and acquisition of two new glycosyltransferase genes from *E. coli* O157 and O152, respectively, which led to these sugar substitutions (Cheng *et al.* 2006). Several kinds of IS elements are found to play important roles in obtaining genes from other strains and inactivation of genes present in the gene cluster, but this is the only case where donor strains were identified.

(iii) *E. coli* O169 and O183. The O169 and O183 O units have identical backbones, but O169 has D-Glc and D-GlcA side-branch residues separately attached to the D-Gal residue in the backbone, while O183 has a two-sugar D-Ribf-(β1→4)-D-GlcpA side-branch attached to the same residue. The seven genes downstream of *galF* share > 97% identity level, which is consistent with their common backbone and the side-chain D-GlcA residue. The *wbaM* gene in O183 is assigned as a putative phosphoribosyltransferase gene, and its function in the transfer of D-Ribf to D-GlcpA has been confirmed (Senchenkova *et al.* 2005). A glycosyltransferase gene, which occurs at the 3'-end of the O antigen gene cluster of O169 but is absent in O183, may be responsible for side-chain glucosylation of O169.

(iv) *E. coli* O86, O90 and O127. The O127 O unit has the structure →2)-L-Fucp-(α1→2)-D-Galp-(β1→3)-D-GalpNAc-(α1→3)-D-GalpNAc-(α1→. The O90 O unit differs in having a (β1→4)-linkage instead of the (α1→2)-linkage between O units, and also in the acetylation position on L-Fuc. The gene clusters are almost 100% identical. However, it is noteworthy that the putative acetyl transferase WbiO of O127 shared 100% identity level to the 3'-end region of its O90 homolog, WerU, and this may result in the different acetylation of L-Fuc of the two serogroups. In addition, it is interesting that the *wzy* gene of O127 possesses three more amino acid residues than that of O90, possibly accounting for the Wzy linkage-difference upon polymerization between the two serogroups. The O86 structure has the Wzy (β1→4)-linkage of O90, and contains an added D-Gal side-branch instead of two acetyl

modifications on the ʟ-Fuc residue. The O86 and O127/O90 gene clusters are more than 96% identical for the first seven genes, but only 47–88% identical for the five downstream genes. It is proposed that one of the O86 and O90/O127 O antigen gene clusters evolved from the other by transfer of the downstream part of the O antigen gene cluster by homologous recombination (Feng *et al.* 2005).

(v) *E. coli* O107 and O117. The 5-sugar O units differ only in the substitution of a ᴅ-GlcNAc for ᴅ-Glc. The two gene clusters have the same genes and, and are 98.6% identical. The difference was shown by gene replacement experiments to be due to specificity of the two GT genes (*wekI*ₒ₁₀₇ and *wekI*ₒ₁₁₇), the products of which differ by only three amino acid substitutions (Wang *et al.* 2012a).

(vi) *E. coli* O124 and O164. The O124 O unit has the main chain structure →3)-ᴅ-Gal*p*-(*β*1→6)-ᴅ-Gal*f*-(*β*1→3)-ᴅ-Gal*p*NAc-(*β*→, with a ᴅ-Glc*p*LA-(*β*1→6)-ᴅ-Glc*p* side chain (*α*1→4)-linked to the ᴅ-Gal*p* residue. The O164 O unit has a very similar structure, with the only difference being the presence of a ᴅ-Glc*p* residue in place of glucolactic acid. The O124 and O164 genes are near identical with > 99.4% identity. The difference is due to a frame shift mutation in the *wfeP* gene in O164 responsible for the synthesis of UDP-ᴅ-Glc*p*LA. (Liu *et al.* 2008).

(vii) *E. coli* O118 and O151. The O118 and O151 O antigens have the same O-unit backbone, but differ in the polymerization linkage, which is (*β*1→3)-linked in O118 but (*β*1→2) in O151. The gene cluster sequences are nearly identical with > 99% overall identity, but the Wzy have a single amino acid substitution, which must be responsible for the O antigen difference (Liu *et al.* 2010). Another possibility is that O118 or O151 could acquire a polymerase inhibitor gene (which inhibits the Wzy in the O antigen gene cluster) and a second copy of a *wzy* gene outside the O antigen gene cluster. A similar case has been reported in *Pseudomonas aeruginosa* (Newton *et al.* 2001; Kaluzny, Abeyrathne and Lam 2007). In addition, the O151 antigen is modified by a single side-branch GlcNAc residue. It is most likely that a glycosyltransferase gene for transfer of GlcNAc in O151 is located elsewhere in the genome.

(viii) *E. coli* O2 and O50. The O2 and O50 O units have the same main chain but O2 has a ᴅ-Fuc*p*3NAc side-branch on the last sugar. The gene clusters are > 99% identical. The genetic basis for the lack of ᴅ-Fuc*p*3NAc in O50 is evidently the point mutation in the aminotransferase gene *fdtB* of the ᴅ-Fuc*p*3NAc synthesis pathway resulting in a single amino acid change from histidine to arginine (Yang *et al.* 2018).

(ix) *E. coli* O101 and O162. The O101 and O162 O units both have two structural forms. The O101 O unit contains →6)-ᴅ-Glc*p*NAc-(*α*1→4)-ᴅ-Gal*p*NAc-(*α*1→ (form 1) or →4)-ᴅ-Gal*p*NAc-(*α*1→4)-ᴅ-Gal*p*NAc-(*β*1→ (form 2); O162 has an additional sugar, 4d-ᴅ-*ara*Hex attached to the Glc*p*NAc of form 1, and to the second Gal*p*NAc of form 2. The O101 and O162 gene clusters are nearly identical, with an overall DNA identity of 99.8%, which is consistent with the common backbone structures of these two serogroups. There are four GT genes in this ABC transporter gene cluster that presumably cover both forms. We screened the additional six genes between *gnd* and *ugd* that normally reside outside the O antigen cluster, and found that a very high DNA identity level are shared between O101 and O162, except that the first gene downstream of *gnd* in O101 is interrupted by an IS element. The six genes were assigned as oxidoreductase, epimerase,

xylose isomerase, hypothetical protein and two GTs, respectively. To our knowledge, the biosynthetic pathway of 4d-ᴅ-*ara*Hex has not yet been identified. Whether the six genes are responsible for the 4d-ᴅ-*ara*Hex synthesis and transfer in O162, and that the inserted IS element accounts for the repression of the homologous genes in O101, thus leading to the failure of 4d-ᴅ-*ara*Hex synthesis, need further investigation.

(x) *E. coli* O17, O44, O73, O77 and O106. The DNA sequences of the *E. coli* O17, O44, O73, O77 and O106 gene clusters are > 99% identical. The structures share the same →6)-ᴅ-Man*p*-(*α*1→2)-ᴅ-Man*p*-(*α*1→2)-ᴅ-Man*p*-(*β*1→3)-ᴅ-Glc*p*NAc-(*α*1→ backbone but differ in having only a linear backbone in O77, whereas ᴅ-Glc*p* side-branches with different linkages are present in the others. The O antigen gene clusters are responsible for the synthesis of the backbones, and all but O77 are proposed to have the appropriate *gtr* genes required for addition of their ᴅ-Glc*p* side-branch (Wang *et al.* 2007). This was confirmed for O44 by finding an appropriate prophage with the *gtr* genes in the genome. These O antigens are referred as the O77-group, which has diversified by gain of different ᴅ-Glc*p* side-branches (Wang *et al.* 2007), as observed on a larger scale in *Shigella flexneri* (Jakhetia and Verma 2015).

CONCLUSIONS

Serotyping based on O antigen variability has become the primary method for classifying *E. coli* strains since the 1940s and standardized procedures have been developed. *E. coli* is divided into O and H (flagellar) antigen-defined serotypes and some have proven to be closely associated with human disease, with high morbidity and mortality. For example, the Shiga toxin-producing *E. coli* (STEC) O157:H7 serotype strain is well known for causing life-threatening complications, such as bloody diarrhea (hemolytic colitis) and hemolytic-uremic syndrome (HUS) (Qin *et al.* 2015; Li *et al.* 2017). On the other hand, the most common serogroups of non-O157 STEC (O26, O45, O103, O111, O121 and O145) usually cause foodborne illnesses, with infections being less severe than those caused by O157:H7, but do vary among serogroups (Gould *et al.* 2013). O antigen serotyping provides important pathotype information and is still the prevailing standard for detection of *E. coli* and some other major pathogens and is considered essential for outbreak investigation, risk management, epidemiological studies and treatment options by clinical microbiologists, infectious disease specialists and infection control agencies. Conventional serotyping is a delicate, laborious, time-consuming, and expensive process. In addition, the consistent production of fully validated antisera is difficult. During the last few decades a large number of molecular and PCR-based assays have been developed to target *E. coli* sero-specific genes resulting in reliable and cost-effective analysis procedures (Liu and Fratamico 2006; Lin *et al.* 2011; Iguchi *et al.* 2015b).

The decline of the cost of DNA sequencing, and the ongoing progress in developing user-friendly tools, has led to several whole-genome sequencing (WGS)-based *in silico* serotyping approaches (Ingle *et al.* 2016), which offer better resolution and are superior to conventional methods for pathogen epidemiological investigation and tracing of *E. coli* (Joensen *et al.* 2015), and some of other major pathogens (Ibrahim and Morin 2018; Zhang *et al.* 2015; Thrane *et al.* 2016; Wu *et al.* 2019). However, it is critical for these approaches to have a database containing the relevant details of sero-specific genes, which in turn requires a

thorough understanding of O antigen structures in conjunction with the genetic basis for their diversity and variation.

We herein presented 197 O antigens (including 21 subgroups) representing all but five (O14, O34, O89, O144 and the unpublished O93 data) of 185 *E. coli* O serogroups, where 178 serogroups gave an excellent correlation between the O antigen structure and the corresponding gene cluster. Anomalies for the remaining seven serogroups are summarized in the following way. For O14 lacking a typical O antigen, the O antigen gene cluster is deleted due to a recombination event, and the ECA is attached to the LPS core, suggesting a convenient mechanism for the evolution of *E. coli* whose typical O antigen becomes redundant (Jensen and Reeves 2004). The O34, O89 and O144 type strains are rough, but their O antigen gene clusters seem to be complete (Iguchi *et al.* 2015a). However, it would be difficult to identify catalytic capability using a simple sequence survey, as one or few point mutations may lead to the inactivation of an enzyme, thus resulting in the failure of O antigen synthesis. Notably, the O antigen structures of O20, O57 and O60 have been determined, but they do not fit well with the genetic loci between *galF* and *gnd*. We propose that the functioning O antigen gene clusters of them must be located elsewhere at the chromosome, yet to be identified.

Most *E. coli* structures and gene clusters show no obvious relationship to any others. However, the members within each of the 10 groups summarized herein share almost identical or closely related O antigen structures and gene clusters. This indicates that the members within each group are evolutionarily close, and may have evolved from a common ancestor to subsequently diversify further from one to another. The diversification within different groups can be mediated by recombinational replacement of genes (O24/O56, O169/O183, O86/O90/O127), point mutation leading to the inactivation or functional change of certain genes (O107/O117, O124/O164, O118/O151, O2/O50), IS element insertion resulting in the inactivation of certain gene (O101/O162), and prophages carrying *gtr* genes for the presence of D-Glc side-branch residues (O17/O44/O73/O77/O106).

Most *Shigella* serogroups fall into three clusters within *E. coli* (Pupo, Lan and Reeves 2000), the only exception being *Shigella boydii* 13 that is now recognized as an *Escherichia albertii* serogroup (Hyma *et al.* 2005). Furthermore, there are 22 pairs of O antigens and associated gene clusters identical or closely related between *E. coli* and *Shigella* (Liu *et al.* 2008; Knirel *et al.* 2016a, 2018). Compared to that of *Shigella* (17 anomalies out of 33 O antigen gene clusters), the proportion of anomalies (66 of 178) is lower in *E. coli*. The genetic basis for the O antigen synthesis of nearly all *E. coli* O antigens and documented correlations between O antigens and their gene clusters provide a greatly improved basis for understanding of the evolution of *E. coli*, and for development of sero-specific assays with higher discrimination as well as further improvements of vaccines for protection against pathogenic *E. coli* strains.

SUPPLEMENTARY DATA

Supplementary data are available at FEMSRE online.

FUNDING

REFERENCES

Achtman M, Pluschke G. Clonal analysis of descent and virulence among selected Escherichia coli. *Annu Rev Microbiol* 1986;**40**:185–210.

Adeyeye A, Jansson P-E, Lindberg B *et al*. Structural studies of the Escherichia coli O-149 O-antigen polysaccharide. *Carbohydr Res* 1988;**176**:231–6.

Al-Dabbagh B, Olatunji S, Crouvoisier M *et al*. Catalytic mechanism of MraY and WecA, two paralogues of the polyprenyl-phosphate N-acetylhexosamine 1-phosphate transferase superfamily. *Biochimie* 2016;**127**:249–57.

Alam J, Beyer N, Liu HW. 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* 2004;**43**:16450–60.

Albermann C, Beuttler H. Identification of the GDP-N-acetyl-D-perosamine producing enzymes from Escherichia coli O157:H7. *FEBS Lett* 2008;**582**:479–84.

Alexander DC, Valvano MA. Role of the rfe gene in the biosynthesis of the Escherichia coli O7-specific lipopolysaccharide and other O-specific polysaccharides containing N-acetylglucosamine. *J Bacteriol* 1994;**176**:7079–84.

Ali T, Urbina F, Weintraub A *et al*. Structural studies of the O-antigenic polysaccharides from the enteroaggregative Escherichia coli strain 522/C1 and the international type strain from Escherichia coli O178. *Carbohydr Res* 2005;**340**: 2010–4.

Ali T, Weintraub A, Widmalm G. Structural determination of the O-antigenic polysaccharide from Escherichia coli O166. *Carbohydr Res* 2007;**342**:274–8.

Ali T, Weintraub A, Widmalm G. Structural determination of the O-antigenic polysaccharide from the Shiga toxin-producing Escherichia coli O171. *Carbohydr Res* 2006;**341**:1878–83.

Ali T, Weintraub A, Widmalm G. Structural studies of the O-antigenic polysaccharides from the enteroaggregative Escherichia coli strain 87/D2 and international type strains from E. coli O128. *Carbohydr Res* 2008;**343**: 695–702.

Allard ST, Giraud MF, Whitfield C *et al*. The crystal structure of dTDP-D-Glucose 4,6-dehydratase (RmlB) from Salmonella enterica serovar Typhimurium, the second enzyme in the dTDP-L-rhamnose pathway. *J Mol Biol* 2001;**307**:283–95.

Allison GE, Verma NK. Serotype-converting bacteriophages and O-antigen modification in Shigella flexneri. *Trends Microbiol* 2000;**8**:17–23.

Amor PA, Whitfield C. Molecular and functional analysis of genes required for expression of group IB K antigens in Escherichia coli: characterization of the his-region containing gene clusters for multiple cell-surface polysaccharides. *Mol Microbiol* 1997;**26**:145–61.

Anderson AN, Richards JC, Perry MB. Structure of the O-antigen of Escherichia coli 0119 lipopolysaccharide. *Carbohydr Res* 1992;**237**:249–62.

Anderson MS, Eveland SS, Price NP. Conserved cytoplasmic motifs that distinguish sub-groups of the polyprenol phosphate:N-acetylhexosamine-1-phosphate transferase family. *FEMS Microbiol Lett* 2000;**191**:169–75.

Andersson M, Carlin N, Leontein K *et al.* Structural studies of the O-antigenic polysaccharide of Escherichia coli O86, which possesses blood-group B activity. *Carbohydr Res* 1989;**185**: 211–23.

Arbatsky NP, Shashkov AS, Toukach FV *et al.* Structure of the O-specific polysaccharide of a serologically separate strain Proteus penneri 2 from a new proposed serogroup O66. *Eur J Biochem* 1999;**261**:392–7.

Bao Y, Zhang H, Huang X *et al.* O-specific polysaccharide confers lysozyme resistance to extraintestinal pathogenic Escherichia coli. *Virulence* 2018;**9**:666–80.

Bartelt M, Shashkov AS, Kochanowski H *et al.* Structure of the O-specific polysaccharide of the O22-antigen (LPS) from Escherichia coli O22:K13. *Carbohydr Res* 1994;**254**:203–12.

Bartelt M, Shashkov AS, Kochanowski H *et al.* Structure of the O-specific polysaccharide of the O23 antigen (LPS) from Escherichia coli O23:K?:H16. *Carbohydr Res* 1993;**248**:233–40.

Bastin DA, Stevenson G, Brown PK *et al.* Repeat unit polysaccharides of bacteria: a model for polymerization resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length. *Mol Microbiol* 1993;**7**: 725–34.

Batley M, Redmond JW, Packer NH *et al.* The relationship between the structures of the O polysaccharides from Escherichia coli O17 and O16. *Carbohydr Res* 1997;**303**:313–8.

Baumann H, Jansson P-E, Kenne L *et al.* Structural studies of the Escherichia coli O1A-polysaccharide, using the computer program CASPER. *Carbohydr Res* 1991;**211**:183–90.

Bengoechea JA, Najdenski H, Skurnik M. Lipopolysaccharide O antigen status of Yersinia enterocolitica O:8 is essential for virulence and absence of O antigen affects the expression of other Yersinia virulence factors. *Mol Microbiol* 2004;**52**:451–69.

Bengoechea JA, Pinta E, Salminen T *et al.* Functional characterization of Gne (UDP-N-acetylglucosamine-4-epimerase), Wzz (chain length determinant), and Wzy (O-antigen polymerase) of Yersinia enterocolitica serotype O:8. *J Bacteriol* 2002;**184**:4277–87.

Bielaszewska M, Mellmann A, Zhang W *et al.* Characterisation of the Escherichia coli strain associated with an outbreak of haemolyticuraemic syndrome in Germany, 2011: a microbiological study. *Lancet Infect Dis* 2011;**11**:671–6.

Bi Y, Mann E, Whitfield C *et al.* Architecture of a channel-forming O-antigen polysaccharide ABC transporter. *Nature* 2018;**553**:361–5.

Blakeman KH, Weintraub A, Widmalm G. Structural determination of the O-antigenic polysaccharide from the enterotoxigenic Escherichia coli O147. *Eur J Biochem* 1998;**251**:534–7.

Bonofiglio L, Garcia E, Mollerach M. Biochemical characterization of the pneumococcal glucose 1-phosphate uridylyltransferase (GalU) essential for capsule biosynthesis. *Curr Microbiol* 2005;**51**:217–21.

Brockhausen I, Hu B, Liu B *et al.* Characterization of two β-1,3-glucosyltransferases from Escherichia coli serotypes O56 and O152. *J Bacteriol* 2008;**190**:4922–32.

Bronner D, Clarke BR, Whitfield C. Identification of an ATP-binding cassette transport system required for translocation of lipopolysaccharide O-antigen side-chains across the cytoplasmic membrane of Klebsiella pneumoniae serotype O1. *Mol Microbiol* 1994;**14**:505–19.

Bäckhed F, Roswall J, Peng Y *et al.* Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* 2015;**17**:690–703.

Caboni M, Pédron T, Rossi O *et al.* An O antigen capsule modulates bacterial pathogenesis in Shigella sonnei. *PLoS Pathog* 2015;**11**:e1004749.

Campbell RE, Mosimann SC, Tanner ME *et al.* The structure of UDP-N-acetylglucosamine 2-epimerase reveals homology to phosphoglycosyl transferases. *Biochemistry* 2000;**39**: 14993–5001.

Chen C, Hou X, Utkina N *et al.* Identification and biochemical characterization of a novel α-1,3-mannosyltransferase WfcD from Escherichia coli O141. *Carbohydr Res* 2017;**443–444**:78–86.

Cheng J, Wang Q, Wang W *et al.* Characterization of E. coli O24 and O56 O antigen gene clusters reveals a complex evolutionary history of the O24 gene cluster. *Curr Microbiol* 2006;**53**: 470–6.

Clarke BR, Cuthbertson L, Whitfield C. Nonreducing terminal modifications determine the chain length of polymannose O-antigens of Escherichia coli and couple chain termination to polymer export via an ATP-binding cassette transporter. *J Biol Chem* 2004;**279**: 35709–18.

Clarke BR, Greenfield LK, Bouwman C *et al.* Coordination of polymerization, chain termination, and export in assembly of the Escherichia coli lipopolysaccharide O9a antigen in an ATP-binding cassette transporter-dependent pathway. *J Biol Chem* 2009;**284**:30662–72.

Clarke BR, Richards MR, Greenfield LK *et al.* In vitro reconstitution of the chain termination reaction in biosynthesis of the Escherichia coli O9a O-polysaccharide. The chain-length regulator, WbdD, catalyzes the addition of methyl phosphate to the non-reducing terminus of the growing glycan. *J Biol Chem* 2011;**286**:41391–401.

Collins RF, Kargas V, Clarke BR *et al.* Full-length, oligomeric structure of Wzz determined by cryoelectron microscopy reveals insights into membrane-bound states. *Structure* 2017;**25**:806–15.

Cunneen MM, Liu B, Wang L *et al.* Biosynthesis of UDP-GlcNAc, UndPP-GlcNAc and UDP-GlcNAcA involves three easily distinguished 4-epimerase enzymes, Gne, Gnu and GnaB. *PLoS One* 2013;**8**:e67646.

Curvall M, Lindberg B, Lönngren J *et al.* Structural studies on the Klebsiella O group 3 lipopolysaccharide. *Acta Chem Scand* 1973;**27**:2645–9.

Cuthbertson L, Kimber MS, Whitfield C. Substrate binding by a bacterial ABC transporter involved in polysaccharide export. *Proc Natl Acad Sci USA* 2007;**104**:19529–34.

Czuchry D, Desormeaux P, Stuart M *et al.* Identification and Biochemical Characterization of the Novel α2,3-Sialyltransferase WbwA from Pathogenic Escherichia coli Serotype O104. *J Bacteriol* 2015;**197**:3760–8.

Czuchry D, Szarek WA, Brockhausen I. Identification and biochemical characterization of WbwB, a novel UDP-Gal: Neu5Ac-R α1,4-galactosyltransferase from the intestinal pathogen Escherichia coli serotype O104. *Glycoconj J* 2018;**35**:65–76.

Daniels C, Vindurampulle C, Morona R. Overexpression and topology of the Shigella flexneri O-antigen polymerase (Rfc/Wzy). *Mol Microbiol* 1998;**28**:1211–22.

Dean CR, Franklund CV, Retief JD *et al.* Characterization of the serotype O11 O-antigen locus of Pseudomonas aeruginosa PA103. *J Bacteriol* 1999;**181**:4275–84.

DebRoy C, Fratamico PM, Yan X *et al.* Comparison of O-antigen gene clusters of all O-serogroups of Escherichia coli and proposal for adopting a new nomenclature for O-typing. *PLoS One* 2016;**11**:e0147434.

Dicks LMT, Geldenhuys J, Mikkelsen LS *et al.* Our gut microbiota: a long walk to homeostasis. *Benef Microbes* 2018;**9**: 3–20.

Dmitriev BA, Lvov VL, Tochtamysheva NV *et al*. Cell-wall lipopolysaccharide of Escherichia coli O114:H2. Structure of the polysaccharide chain. *Eur J Biochem* 1983;**134**:517–21.

Duan Z, Senchenkova SN, Guo X *et al*. Structure and gene cluster of the O-antigen of Escherichia coli O156 containing a pyruvic acid acetal. *Carbohydr Res* 2016;**430**:24–28.

Duda KA, Fruth A, Holst O. Structural studies of the O-antigenic polysaccharide of the bovine mastitis isolate Escherichia coli serotype O174. *Carbohydr Res* 2013;**373**:18–21.

Eichler J, Imperiali B. Stereochemical divergence of polyprenol phosphate glycosyltransferases. *Trends Biochem Sci* 2018;**43**:10–17

Eklund K, Garegg PJ, Kenne L *et al*. *Structural studies on the Escherichia coli O111 lipopolysaccharide*. Abstr. 9th Intern. Symp. Carbohydr. Chem., London, UK 1978; p. 493.

Erbing C, Kenne L, Lindberg B *et al*. Structure of the O-specific side-chains of the Escherichia coli O 75 lipopolysaccharide: a revision. *Carbohydr Res* 1978;**60**:400–3.

Feng L, Perepelov AV, Zhao G *et al*. Structural and genetic evidence that the Escherichia coli O148 O antigen is the precursor of the Shigella dysenteriae type 1 O antigen and identification of a glucosyltransferase gene. *Microbiology* 2007;**153**:139–47.

Feng L, Senchenkova SN, Tao J *et al*. Structural and genetic characterization of enterohaemorrhagic Escherichia coli O145 O antigen and development of an O145 serogroup-specific PCR assay. *J Bacteriol* 2005;**187**:758–64.

Feng L, Senchenkova SN, Yang J *et al*. Synthesis of the heteropolysaccharide O antigen of Escherichia coli O52 requires an ABC transporter: structural and genetic evidence. *J Bacteriol* 2004;**186**:4510–9.

Follens A, Veiga-da-Cunha M, Merckx R *et al*. acs1 of Haemophilus influenzae type a capsulation locus region II encodes a bifunctional ribulose 5-phosphate reductase-CDP-ribitol pyrophosphorylase. *J Bacteriol* 1999;**181**:2001–7.

Fontana C, Lundborg M, Weintraub A *et al*. Rapid structural elucidation of polysaccharides employing predicted functions of glycosyltransferases and NMR data: application to the O-antigen of Escherichia coli O59. *Glycobiology* 2014;**24**:450–7.

Fontana C, Lundborg M, Weintraub A *et al*. Structural studies and biosynthetic aspects of the O-antigen polysaccharide from Escherichia coli O174. *Carbohydr Res* 2012;**354**:102–5.

Fontana C, Ramstrom K, Weintraub A *et al*. Structural studies of the O-antigen polysaccharide from Escherichia coli O115 and biosynthetic aspects thereof. *Glycobiology* 2013a;**23**:354–62.

Fontana C, Weintraub A, Widmalm G. Facile structural elucidation of glycans using NMR spectroscopy data and the program CASPER: Application to the O-antigen polysaccharide of Escherichia coli O155. *ChemPlusChem* 2013b;**78**:1327–9.

Fontana C, Weintraub A, Widmalm G. Structural elucidation of the O-antigen polysaccharide from Escherichia coli O181. *ChemistryOpen* 2015b;**4**:47–55.

Fontana C, Weintraub A, Widmalm G. Structural studies and biosynthetic aspects of the O-antigen polysaccharide from Escherichia coli O42. *Carbohydr Res* 2015a;**403**:174–81.

Fontana C, Weintraub A, Widmalm G. Structural studies and biosynthetic aspects of the O-antigen polysaccharide from shigatoxin-producing E.coli O179, 29th International Carbohydrate Symposium, Lisbon, Portugal, *Poster no BS2*. 2018.

Fundin J, Weintraub A, Xu J-G *et al*. NMR analysis of the O-antigen polysaccharide from Escherichia coli strain F171. *Magn Reson Chem* 2003;**41**:202–5.

Färnbäck M, Weintraub A, Widmalm G. Structural determination of the O-antigenic polysaccharide from Escherichia coli O141. *Eur J Biochem* 1998;**254**:168–71.

Gamian A, Kenne L, Mieszala M *et al*. Structure of the Escherichia coli O24 and O56 O-specific sialic-acid-containing polysaccharides and linkage of these structures to the core region in lipopolysaccharides. *Eur J Biochem* 1994;**225**:1211–20.

Gamian A, Romanowska E, Ulrich J *et al*. The structure of the sialic acid-containing Escherichia coli O104 O-specific polysaccharide and its linkage to the core region in lipopolysaccharide. *Carbohydr Res* 1992;**236**:195–208.

Gao Y, Liu B, Strum S *et al*. Biochemical characterization of WbdN, a β-1,3-glucosyltransferase involved in O-antigen synthesis in enterohemorrhagic Escherichia coli O157. *Glycobiology* 2012;**22**:1092–102.

Giraud MF, Gordon FM, Whitfield C *et al*. Purification, crystallization and preliminary structural studies of dTDP-6-deoxy-D-xylo-4-hexulose 3,5-epimerase (RmlC), the third enzyme of the dTDP-L-rhamnose synthesis pathway, from Salmonella enterica serovar typhimurium. *Acta Crystallogr D Biol Crystallogr* 1999a;**55**:706–8.

Giraud MF, McMiken HJ, Leonard GA *et al*. Overexpression, purification, crystallization and preliminary structural study of dTDP-6-deoxy-L-lyxo-4-hexulose reductase (RmlD), the fourth enzyme of the dTDP-L-rhamnose synthesis pathway, from Salmonella enterica serovar Typhimurium. *Acta Crystallogr D Biol Crystallogr* 1999b;**55**:2043–6.

Gould LH, Mody RK, Ong KL *et al*. Emerging Infections Program Foodnet Working Group. Increased recognition of non-O157 Shiga toxin-producing Escherichia coli infections in the United States during 2000–2010: epidemiologic features and comparison with E. coli O157 infections. *Foodborne Pathog Dis* 2013;**10**:453–60.

Greenfield LK, Richards MR, Li J *et al*. Biosynthesis of the polymannose lipopolysaccharide O-antigens from Escherichia coli serotypes O8 and O9a requires a unique combination of single- and multiple-active site mannosyltransferases. *J Biol Chem* 2012b;**287**:35078–91.

Greenfield LK, Richards MR, Vinogradov E *et al*. Domain Organization of the Polymerizing Mannosyltransferases Involved in Synthesis of the Escherichia coli O8 and O9a Lipopolysaccharide O-antigens. *J Biol Chem* 2012a;**287**:38135–49.

Guan W, Cai L, Fang J *et al*. Enzymatic synthesis of UDP-GlcNAc/UDP-GalNAc analogs using N-acetylglucosamine 1-phosphate Uridyltransferase (GlmU). *Chem Commun* 2009;6976–8.

Guo H, Yi W, Shao J *et al*. Molecular analysis of the O-antigen gene cluster of Escherichia coli O86:B7 and characterization of the chain length determinant gene (wzz). *Appl Environ Microbiol* 2005;**71**:7995–8001.

Guo X, Senchenkova SN, Shashkov AS *et al*. Structure and gene cluster of the O-antigen of Escherichia coli O96. *Carbohydr Res* 2016;**420**:1–5.

Gupta DS, Shashkov AS, Jann B *et al*. Structures of the O1B and O1C lipopolysaccharide antigens of Escherichia coli. *J Bacteriol* 1992;**174**:7963–70.

Hakulinen JK, Hering J, Brändén G *et al*. MraY-antibiotic complex reveals details of tunicamycin mode of action. *Nat Chem Biol* 2017;**13**:265–67.

Han W, Wu B, Li L *et al*. Defining function of lipopolysaccharide O-antigen ligase WaaL using chemoenzymatically synthesized substrates. *J Biol Chem* 2011;**287**:5357–65.

Hong Y, Reeves PR. Diversity of O-antigen repeat-unit structures can account for the substantial sequence variation in Wzx translocases. *J Bacteriol* 2014;**196**:1713–22.

Hou X, Perepelov AV, Guo X *et al*. A gene cluster at an unusual chromosomal location responsible for the novel O-antigen synthesis in Escherichia coli O62 by the ABC transporter-dependent pathway. *Glycobiology* 2017;**27**:669–76.

Hu B, Perepelov AV, Liu B *et al*. Structural and genetic evidence for the close relationship between Escherichia coli O71 and Salmonella enterica O28 O-antigens. *FEMS Immunol Med Microbiol* 2010;**59**:161–9.

Hyma KE Lacher DW, Nelson AM *et al*. Evolutionary genetics of a new pathogenic Escherichia species: Escherichia albertii and related Shigella boydii strains. *J Bacteriol* 2005;**187**:619–28.

Ibrahim GM, Morin PM, Salmonella Serotyping Using Whole Genome Sequencing. *Front Microbiol* 2018;**9**:2993.

Ibrahim GM, Morin PM. Salmonella serotyping using whole genome sequencing. *Front Microbiol* 2018;**9**:2993.

Iguchi A, Iyoda S, Kikuchi T *et al*. A complete view of the genetic diversity of the Escherichia coli O-antigen biosynthesis gene cluster. *DNA Res* 2015a;**22**:101–7.

Iguchi A, Iyoda S, Seto K *et al*. Escherichia coli O-genotyping PCR: a comprehensive and practical platform for molecular O serogrouping. *J Clin Microbiol* 2015b;**53**:2427–32.

Ingle DJ, Valcanis M, Kuzevski A *et al*. In silico serotyping of E. coli form short read identifies limited novel O-loci but extensive diversity of O:H serotype combinations within and between pathogenic lineages. *Microbial Genomics* 2016;**2**:64.

Islam ST, Lam JS. Wzx flippase-mediated membrane translocation of sugar polymer precursors in bacteria. *Environ Microbiol* 2013;**15**:1001–15.

Jakhetia R, Verma NK, Identification and Molecular Characterisation of a Novel Mu-Like Bacteriophage, SfMu, of Shigella flexneri. *PLoS ONE* 2015;**10**:e0124053.

Jann B, Shashkov A, Torgov V *et al*. NMR investigation of the 6-deoxy-L-talose-containing O45, O45-related (O45rel), and O66 polysaccharides of Escherichia coli. *Carbohydr Res* 1995;**278**:155–65.

Jann B, Shashkov AS, Gupta DS *et al*. The O18-antigens (lipopolysaccharides) of Escherichia coli. Structural characterization of the O18A specific, O18A1-specific, O18B-specific and O18B1-specific polysaccharides. *Eur J Biochem* 1992b;**210**:241–8.

Jann B, Shashkov AS, Gupta DS *et al*. The O1 antigen of Escherichia coli: structural characterization of the O1A1-specific polysaccharide. *Carbohydr Polym* 1992a;**18**:51–57.

Jann B, Shashkov AS, Hahne M *et al*. Structure of the O83-specific polysaccharide of Escherichia coli O83:K24:H31. *Carbohydr Res* 1994c;**261**:215–22.

Jann B, Shashkov AS, Kochanowski H *et al*. Structural comparison of the O4-specific polysaccharides from E. coli O4:K6 and E. coli O4:K52. *Carbohydr Res* 1993;**248**:241–51.

Jann B, Shashkov AS, Kochanowski H *et al*. Structural comparison of the O6 specific polysaccharides from E. coli O6:K2:H1, E. coli O6:K13:H1, and E. coli O6:K54:H10. *Carbohydr Res* 1994a;**263**:217–25.

Jann B, Shashkov AS, Kochanowski H *et al*. Structure of the O16 polysaccharide from Escherichia coli O16:K1: an NMR investigation. *Carbohydr Res* 1994b;**264**:305–11.

Jansson P-E, Kenne L, Widmalm G. Structure of the O-antigen polysaccharide from Escherichia coli O18ac: a revision using computer-assisted structural analysis with the program CASPER. *Carbohydr Res* 1989;**193**:322–5.

Jansson P-E, Lennholm H, Lindberg B *et al*. Structural studies of the O-specific side-chains of the Escherichia coli O2 lipopolysaccharide. *Carbohydr Res* 1987a;**161**:273–9.

Jansson P-E, Lindberg B, Lönngren J *et al*. Structural studies of the Escherichia coli O-antigen 6. *Carbohydr Res* 1984b;**131**:277–83.

Jansson P-E, Lindberg B, Ogunlesi M *et al*. Structural studies of the O-antigen polysaccharide of Escherichia coli O4. *Carbohydr Res* 1984a;**134**:283–91.

Jansson P-E, Lindberg B, Widmalm G *et al*. Structural studies of the Escherichia coli O78 O-antigen polysaccharide. *Carbohydr Res* 1987b;**165**:87–92.

Jansson PE, Lönngren J, Widmalm G *et al*. Structural studies of the O-antigen polysaccharides of Klebsiella O5 and Escherichia coli O8. *Carbohydr Res* 1985;**145**:59–66.

Jensen SO, Reeves PR. Deletion of the Escherichia coli O14:K7 O antigen gene cluster. *Can J Microbiol* 2004;**50**:299–302.

Jiang L, Filatov AV, Liu B *et al*. Structure and gene cluster of the O-antigen of Escherichia coli O68. *Carbohydr Res* 2014;**397**:27–30.

Joensen KG, Tetzschner AM, Iguchi A *et al*. Rapid and easy in silico serotyping of Escherichia coli isolates by use of whole-genome sequencing data. *J Clin Microbiol* 2015;**53**:2410–26.

Jonsson KH, Weintraub A, Widmalm G. Structural determination of the O-antigenic polysaccharide from Escherichia coli O74. *Carbohydr Res* 2009;**344**:1592–5.

Jonsson KH, Weintraub A, Widmalm G. Structural studies of the O-antigenic polysaccharides from Shigella dysenteriae type 3 and Escherichia coli O124, a reinvestigation. *Carbohydr Res* 2006;**341**:2986–9.

Kaluzny K, Abeyrathne PD, Lam JS. Coexistence of two distinct versions of O-antigen polymerase, Wzyα and Wzyβ, in P. aeruginosa serogroup O2 and their contributions to cell surface diversity. *J Bacteriol* 2007;**189**:4141–52.

Kaper JB, Nataro JP, Mobley HL. Pathogenic Escherichia coli. *Nat Rev Microbiol* 2004;**2**:123–40.

Kauffman F The serology of the coli group. *J Immunol* 1947;**57**:71–100.

Keenleyside WJ, Perry M, Maclean L *et al*. A plasmid-encoded rfbO:54 gene cluster is required for biosynthesis of the O:54 antigen in Salmonella enterica serovar Borreze. *Mol Microbiol* 1994;**11**:437–48.

Keenleyside WJ, Whitefield C. A novel pathway for O-polysaccharide biosynthesis in Salmonella enterica serovar Borreze. *J Biol Chem* 1996;**271**:28581–92.

Kenne L, Lindberg B, Lugowski C *et al*. Structural studies of the O-specific side-chains of the Escherichia coli O10 lipopolysaccharide. *Carbohydr Res* 1986;**151**:349–58.

Kenne L, Lindberg B, Madden JK *et al*. Structural studies of the Escherichia coli O-antigen 25. *Carbohydr Res* 1983;**122**:249–56.

Kenyon JJ, Cunneen MM, Reeves PR. Genetics and evolution of Yersinia pseudotuberculosis O-specific polysaccharide: a novel pattern of O-antigen diversity. *FEMS Microbiol Rev* 2017;**41**:200–17.

Kido N, Kobayashi H. A Single Amino Acid Substitution in a Mannosyltransferase, WbdA, Converts the Escherichia coli O9 Polysaccharide into O9a: Generation of a New O-Serotype Group. *J Bacteriol* 2000;**182**:2567–73.

Kido N, Torgov VI, Sugiyama T *et al*. Expression of the O9 polysaccharide of Escherichia coli: sequencing of the E. coli O9 rfb gene cluster, characterization of mannosyl transferases, and evidence for an ATP-binding cassette transport system. *J Bacteriol* 1995;**177**:2178–87.

Kjellberg A, Urbina F, Weintraub A *et al*. Structural analysis of the O-antigenic polysaccharide from the enteropathogenic Escherichia coli O125. *Eur J Biochem* 1996;**239**:532–8.

Kjellberg A, Weintraub A, Widmalm G. Structural determination and biosynthetic studies of the O-antigenic polysaccharide from the enterohemorrhagic Escherichia coli O91 using 13C-enrichment and NMR spectroscopy. *Biochemistry* 1999;**38**:12205–11.

Kneidinger B, Graninger M, Adam G *et al*. Identification of two GDP-6-deoxy-D-lyxo-4-hexulose reductases synthesizing GDP-D-rhamnose in Aneurinibacillus thermoaerophilus L420-91T. *J Biol Chem* 2001;**276**:5577–83.

Kneidinger B, Larocque S, Brisson J *et al*. 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* 2003;**371**:989–95.

Knirel YA, Guo X, Senchenkova SN *et al*. Structures and gene clusters of the O-specific polysaccharides of the lipopolysaccharides of Escherichia coli O69 and O146 containing glycolactilic acids: ether conjugates of D-GlcNAc and D-Glc with (R)- and (S)-lactic acid. *Glycoconj J* 2017;**34**:71–84.

Knirel YA, Qian C, Senchenkova SN *et al*. Structure of the β-L-fucopyranosyl phosphate-containing O-specific polysaccharide of Escherichia coli O84. *Int J Biol Macromol* 2016b;**88**:578–85.

Knirel YA, Qian C, Shashkov AS *et al*. Erratum to: Structual relationships between genetically closely related O-antigens of Escherichia coli and Shigella spp. *Biochemistry (Moscow)* 2018;**83**:1422–3.

Knirel YA, Qian C, Shashkov AS *et al*. Structural relationships between genetically closely related O-antigens of Escherichia coli and Shigella spp. *Biochemistry (Moscow)* 2016a;**81**:600–8.

Kogan G, Jann B, Jann K. Structure of the Escherichia coli 0104 polysaccharide and its identity with the capsular K9 polysaccharide. *FEMS Microbiol Lett* 1992;**91**:135–40.

Korres H, Mavris M, Morona R *et al*. Topological analysis of GtrA and GtrB proteins encoded by the serotype-converting cassette of Shigella flexneri. *Biochem Biophys Res Commun* 2005;**328**:1252–60.

Korres H, Verma NK. Topological analysis of glucosyltransferase GtrV of Shigella flexneri by a dual reporter system and identification of a unique reentrant loop. *J Biol Chem* 2004;**279**:22469–76.

Kos V, Cuthbertson L, Whitfield C. The Klebsiella pneumoniae O2a antigen defines a second mechanism for O antigen ATP-binding cassette transporters. *J Biol Chem* 2009;**284**:2947–56.

Kuhn H-M, Meier-Dieter U, Mayer H. ECA, the enterobacterial common antigen. *FEMS microbiology Reviews* 1988;**54**:195–222.

Kuk ACY, Mashalidis EH, Lee S-Y. Crystal structure of the MOP flippase MurJ in an inward-facing conformation. *Nat Struct Mol Biol* 2017;**24**:171–76.

L'vov VL, Shashkov AS, Dmitriev BA *et al*. Structural studies of the O-specific side chain of the lipopolysaccharide from Escherichia coli O:7. *Carbohydr Res* 1984;**126**:249–59.

Lam JS, Taylor VL, Islam ST *et al*. Genetic and functional diversity of Pseudomonas aeruginosa lipopolysaccharide. *Front Microbiol* 2011;**2**:118.

Landersjö C, Weintraub A, Widmalm G. Structural analysis of the O-antigenic polysaccharide from the enteropathogenic Escherichia coli O142. *Eur J Biochem* 1997;**244**:449–53.

Landersjö C, Weintraub A, Widmalm G. Structural analysis of the O-antigen polysaccharide from the Shiga toxin-producing Escherichia coli O172. *Eur J Biochem* 2001;**268**:2239–45.

Landersjö C, Weintraub A, Widmalm G. Structure determination of the O-antigen polysaccharide from the enterovasive Escherichia coli (EIEC) O143 by component analysis and NMR spectroscopy. *Carbohydr Res* 1996;**291**:209–16.

Larkin A, Imperiali B. Biosynthesis of UDP-GlcNAc(3NAc)A by WbpB, WbpE, and WbpD: enzymes in the Wbp pathway responsible for O-antigen assembly in Pseudomonas aeruginosa PAO1. *Biochemistry* 2009;**48**:5446–55.

Larsson EA, Urbina F, Yang Z *et al*. Structural and immunochemical relationship between the O-antigenic polysaccharides from the enteroaggregative Escherichia coli strain 396/C-1 and Escherichia coli O126. *Carbohydr Res* 2004;**339**:1491–6.

Lehrer J, Vigeant KA, Tatar LD *et al*. Functional characterization and membrane topology of Escherichia coli WecA, a sugar-phosphate transferase initiating the biosynthesis of enterobacterial common antigen and O-antigen lipopolysaccharide. *J Bacteriol* 2007;**189**:2618–28.

Leslie MR, Parolis H, Parolis LAS. The structure of the O-antigen of Escherichia coli O116:K+:H10. *Carbohydr Res* 1999;**321**:246–56.

Li B, Liu H, Wang W. Multiplex real-time PCR assay for detection of Escherichia coli O157:H7 and screening for non-O157 Shiga toxin-producing E coli. *BMC Microbiol* 2017;**17**:215.

Li L, Stoeckert CJ, Jr., Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 2003;**13**:2178–89.

Li M, Liu XW, Shao J *et al*. Characterization of a novel α1,2-fucosyltransferase of Escherichia coli O128:b12 and functional investigation of its common motif. *Biochemistry* 2008;**47**:378–87.

Lin A, Nguyen L, Lee T *et al*. Rapid O serogroup identification of the ten most clinically relevant STECs by Luminex microbead-based suspension array. *J Microbiol Methods* 2011;**87**:105–10.

Lindberg B, Lindh F, Lönngren J *et al*. Structural studies of the O-specific side-chain of the lipopolysaccharide from Escherichia coli O55. *Carbohydr Res* 1981;**97**:105–12.

Linnerborg M, Weintraub A, Widmalm G. Structural studies of the O-antigen polysaccharide from Escherichia coli O138. *Eur J Biochem* 1997a;**247**:567–71.

Linnerborg M, Weintraub A, Widmalm G. Structural studies of the O-antigen polysaccharide from the enteroinvasive Escherichia coli O164 cross-reacting with Shigella dysenteriae type 3. *Eur J Biochem* 1999b;**266**:460–6.

Linnerborg M, Weintraub A, Widmalm G. Structural studies of the O-antigen polysaccharide from the enteroinvasive Escherichia coli O173. *Carbohydr Res* 1999c;**320**:200–8.

Linnerborg M, Weintraub A, Widmalm G. Structural studies utilizing 13C-enrichment of the O-antigen polysaccharide from the enterotoxigenic Escherichia coli O159 cross-reacting with Shigella dysenteriae type 4. *Eur J Biochem* 1999a;**266**:246–51.

Linnerborg M, Wollin R, Widmalm G. Structural studies of the O-antigen polysaccharide from Escherichia coli O167. *Eur J Biochem* 1997b;**246**:565–73.

Linton KJ, Higgins CF. The Escherichia coli ATP-binding cassette (ABC) proteins. *Mol Microbiol* 1998;**28**:5–13.

Liston SD, Mann E, Whitfield C. Glycolipid substrates for ABC transporters required for the assembly of bacterial cell-envelopeand cell-surface glycoconjugates. *Biochim Biophys Acta* 2017;**1862**:1394–1403.

Liu B, Knirel YA, Feng L *et al*. Structural diversity in Salmonella O antigens and its genetic basis. *FEMS Microbiol Rev* 2014;**38**:56–89.

Liu B, Knirel YA, Feng L *et al.* Structure and genetics of Shigella O antigens. *FEMS Microbiol Rev* 2008;**32**:627–53; ibid. 2010; **436**:606.

Liu B, Perepelov AV, Guo D *et al.* Structural and genetic relationships between the O-antigens of Escherichia coli O118 and O151. *FEMS Immunol Med Microbiol* 2010;**60**:199–207.

Liu XW, Xia C, Li L *et al.* Characterization and synthetic application of a novel *β*-1,3-galactosyltransferase from Escherichia coli O55:H7. *Bioorg Med Chem* 2009;**17**:4910–5.

Liu Y, Fratamico P. Escherichia coli O antigen typing using DNA microarrays. *Mol Cell Probes* 2006;**20**:239–44.

Li X, Perepelov AV, Wang Q *et al.* Structural and genetic characterization of the O-antigen of Escherichia coli O161 containing a derivative of a higher acidic diamino sugar, legionaminic acid. *Carbohydr Res* 2010;**345**:1581–7.

Li Y, Perepelov AV, Guo D *et al.* Structural and genetic relationships of two pairs of closely related O-antigens of Escherichia coli and Salmonella enterica: E. coli O11/S. enterica O16 and E. coli O21/S. enterica O38. *FEMS Immunol Med Microbiol* 2011;**61**:258–68.

Lundborg M, Modhukur V, Widmalm G. Glycosyltransferase functions of E. coli O-antigens. *Glycobiology* 2010;**20**:366–8.

Lycknert K, Widmalm G. Dynamics of the Escherichia coli O91 O-antigen Polysaccharide in solution as studied by Carbon-13 NMR relaxation. *Biomacromolecules* 2004;**5**:1015–20.

Lüthje P, Brauner A. Virulence factors of uropathogenic E. coli and their interaction with the host. *Adv Microb Physiol* 2014;**65**:337–72.

MacLean LL, Perry MB. Characterization of the antigenic O-polysaccharide produced by Escherichia coli serotype O:70. *Carbohydr Res* 2010a;**345**:644–8.

MacLean LL, Perry MB. Structural characterization of the serotype O:5 O-polysaccharide antigen of the lipopolysaccharide of Escherichia coli O:5. *Biochem Cell Biol* 1997;**75**: 199–205.

MacLean LL, Vinogradov E, Perry MB. The structure of the antigenic O-polysaccharide in the lipopolysaccharide of enterohaemorrhagic Escherichia coli serotype O71:H12. *Biochem Cell Biol* 2010b;**88**:439–44.

MacLean LL, Webb AC, Perry MB. Structural elucidation of the O-antigenic polysaccharide from enterohemorrhagic (EHEC) Escherichia coli O48:H21. *Carbohydr Res* 2006;**341**:2543–9.

Manca MC, Weintraub A, Widmalm G. Structural studies of the Escherichia coli O26 O-antigen polysaccharide. *Carbohydr Res* 1996;**281**:155–60.

Mann E, Kimber MS, Whitfield C. Bioinformatics analysis of diversity in bacterial glycan chain-termination chemistry and organization of carbohydrate binding modules linked to ABC transporters. *Glycobiology* 2019;**29**:822–38.

Mann E, Mallette E, Clarke BR *et al.* The Klebsiella pneumoniae O12 ATP-binding Cassette (ABC) transporter recognizes the terminal residue of Its O-antigen polysaccharide substrate. *J Biol Chem* 2016;**291**:9748–61.

March C, Cano V, Moranta D *et al.* Role of bacterial surface structures on the interaction of Klebsiella pneumoniae with phagocytes. *PLoS ONE* 2013;**8**:e56847.

Marie C, Weintraub A, Widmalm G. Structural studies of the O-antigenic polysaccharide from Escherichia coli O139. *Eur J Biochem* 1998;**254**:378–81.

Marolda CL, Li B, Lung M *et al.* Membrane topology and identification of critical amino acid residues in the Wzx O-antigen translocase from Escherichia coli O157:H4. *J Bacteriol* 2010;**192**:6160–71.

Marolda CL, Vicarioli J, Valvano MA. Wzx proteins involved in biosynthesis of O antigen function in association with the first sugar of the O-specific lipopolysaccharide subunit. *Microbiology* 2004;**150**:4095–105.

Marsden BJ, Bundle DR, Perry MB. Serological and structural relationships between Escherichia coli O:98 and Yersinia enterocolitica O:11,23 and O:11,24 lipopolysaccharide O-antigens. *Biochem Cell Biol* 1994;**72**:163–8.

Masoud H, Perry MB. Structural characterization of the O-antigenic polysaccharide of Escherichia coli serotype O17 lipopolysaccharide. *Biochem Cell Biol* 1996;**74**:241–8.

McGrath BC, Osborn MJ. Localisation of the terminal steps of O-antigen synthesis in Salmonella typhimurium. *J Bacteriol* 1991;**173**:649–54.

Medina EC, Widmalm G, Weintraub A *et al.* Structural studies of the O-antigenic polysaccharides of Escherichia coli O3 and the enteroaggregative Escherichia coli strain 17-2. *Eur J Biochem* 1994;**224**:191–6.

Mengin-Lecreulx D, van Heijenoort J. Identification of the glmU gene encoding N-acetylglucosamine-1-phosphate uridyltransferase in Escherichia coli. *J Bacteriol* 1993;**175**:6150–7.

Merino S, Gonzalez V, Tomás JM. The first sugar of the repeat units is essential for the Wzy polymerase activity and elongation of the O-antigen lipopolysaccharide. *Future Microbiol* 2016;**11**:903–18.

Mulford CA, Osborn MJ. A intermediate step in translocation of lipopolysaccharide to outer membrane of Salmonella typhimurium. *Proc Natl Acad Sci USA* 1983;**80**:1159–63.

Muniesa M, Hammerl JA, Hertwig S *et al.* Shiga toxin-producing Escherichia coli O104:H4: a new challenge for microbiology. *Appl Environ Microbiol* 2012;**78**:4065–73.

Muñoz R, López R, de Frutos M *et al.* First molecular characterization of a uridine diphosphate galacturonate 4-epimerase: an enzyme required for capsular biosynthesis in Streptococcus pneumoniae type 1. *Mol Microbiol* 1999;**31**:703–13.

Nassau PM, Martin SL, Brown RE *et al.* Galactofuranose biosynthesis in Escherichia coli K-12: identification and cloning of UDP-galactopyranose mutase. *J Bacteriol* 1996;**178**: 1047–52.

Nataro JP, Kaper JB. Diarrheagenic Escherichia coli. *Clin Microbiol Rev* 1998;**11**:142–201.

Naumenko OI, Guo X, Senchenkova SN *et al.* Structure and gene cluster of the O-antigen of Escherichia coli O54. *Carbohydr Res* 2018a;**462**:34–38.

Naumenko OI, Song J, Senchenkova SN *et al.* Structural studies on the O-polysaccharide of Escherichia coli O57. *Carbohydr Res* 2018b;**465**:1–3.

Newton JG, Daniels C, Burrows LL *et al.* Three-component-mediated serotype conversion in Pseudomonas aeruginosa by bacteriophage D3. *Mol Microbiol* 2001;**39**:1237–47.

Okuda S, Sherman DJ, Silhavy TJ *et al.* Lipopolysaccharide transport and assembly at the outer membrane: the PEZ model. *Nat Rev Microbiol* 2016;**14**:337–45.

Olsson U, Lycknert K, Stenutz R *et al.* Structural analysis of the O-antigen polysaccharide from Escherichia coli O152. *Carbohydr Res* 2005;**340**:167–71.

Olsson U, Weintraub A, Widmalm G. Structural determination of the O-antigenic polysaccharide from the verocytotoxin-producing Escherichia coli O176. *Carbohydr Res* 2008;**343**: 805–9.

Parolis H, Parolis LAS, Olivieri G. Structural studies on the Shigella-like Escherichia coli O121 O-specific polysaccharide. *Carbohydr Res* 1997;**303**:319–25.

Parolis H, Parolis LAS. The structure of the O-specific polysaccharide from Escherichia coli O113 lipopolysaccharide. *Carbohydr Res* 1995;**267**:263–9.

Parolis LAS, Parolis H, Dutton GGS. Structural studies of the O-antigen polysaccharide of Escherichia coli O9a. *Carbohydr Res* 1986;**155**:272–6.

Paulsen IT, Beness AM, Saier MH, Jr. Computer-based analyses of the protein constituents of transport systems catalysing export of complex carbohydrates in bacteria. *Microbiology* 1997;**143**:2685–99.

Perepelov AV, Chen T, Senchenkova SN *et al*. Structure and genetics of the O-specific polysaccharide of the lipopolysaccharide of Escherichia coli O27. *CarbohydrRes* 2018a;**456**:1–4.

Perepelov AV, Filatov AV, Wang Q *et al*. Structure elucidation and gene cluster annotation of the O-antigen of Escherichia coli O39; application of anhydrous trifluoroacetic acid for selective cleavage of glycosidic linkages. *Carbohydr Res* 2014c;**388**:30–36.

Perepelov AV, Guo X, Filatov AV *et al*. Structure and gene cluster of the O-antigen of Escherichia coli O43. *Carbohydr Res* 2015a;**416**:32–36.

Perepelov AV, Guo X, Senchenkova SN *et al*. Structure and gene cluster of the O-antigen of Escherichia coli O137. *Carbohydr Res* 2016b;**422**:13–16.

Perepelov AV, Guo X, Senchenkova SN *et al*. Structure and genetics of biosynthesis of the phosphate-containing O-polysaccharide of Escherichia coli O160. *Carbohydr Res* 2015d;**417**:89–93.

Perepelov AV, Guo X, Shashkov AS *et al*. Structure elucidation and analysis of biosynthesis genes of the O-antigen of Escherichia coli O131 containing N-acetylneuraminic acid. *Carbohydr Res* 2016a;**436**:41–44.

Perepelov AV, Han W, Senchenkova SN *et al*. Structure of the O-polysaccharide of Escherichia coli O150 containing 2-acetamido-4-O-[(S)-1-carboxyethyl]-2-deoxy-D-glucose. *Carbohydr Res* 2007b;**342**:648–52.

Perepelov AV, Hao X, Senchenkova SN *et al*. Structure and gene cluster of the O-antigen of Escherichia coli O30. *Carbohydr Res* 2014a;**389**:196–8.

Perepelov AV, Li D, Liu B *et al*. Structural and genetic characterization of Escherichia coli O99 antigen. *FEMS Immunol Med Microbiol* 2009b;**57**:80–87.

Perepelov AV, Li D, Liu B *et al*. Structural and genetic characterization of the closely related O-antigens of Escherichia coli O85 and Salmonella enterica O17. *Innate Immun* 2011d;**17**:164–73.

Perepelov AV, Liu B, Senchenkova SN *et al*. Close relation of the O-polysaccharide structure of Escherichia coli O168 and revised structure of the O-polysaccharide of Shigella dysenteriae type 4. *Carbohydr Res* 2007c;**342**:2676–81.

Perepelov AV, Liu B, Senchenkova SN *et al*. O-antigen structure and gene clusters of Escherichia coli O51 and Salmonella enterica O57; another instance of identical O-antigens in the two species. *Carbohydr Res* 2011c;**346**:828–32.

Perepelov AV, Liu B, Senchenkova SN *et al*. Structure of the O-antigen and characterization of the O-antigen gene cluster of Escherichia coli O108 containing 5,7-diacetamido-3,5,7,9-tetradeoxy-L-glycero-D-galacto-non-2-ulosonic (8-epilegionaminic) acid. *Biochemistry (Moscow)* 2010b;**75**:19–24.

Perepelov AV, Liu B, Senchenkova SN *et al*. Structure of the O-polysaccharide of Escherichia coli O112ab containing L-iduronic acid. *Carbohydr Res* 2008c;**343**:571–5.

Perepelov AV, Liu B, Senchenkova SN *et al*. Structures of the O-polysaccharides of Salmonella enterica O59 and Escherichia coli O15. *Carbohydr Res* 2011a;**346**:381–3.

Perepelov AV, Liu B, Senchenkova SN *et al*. The O-antigen of Salmonella enterica O13 and its relation to the O-antigen of Escherichia coli O127. *Carbohydr Res* 2010d;**345**:1808–11.

Perepelov AV, Liu B, Senchenkova SN *et al*. The structure of the glycerol phosphate-containing O-specific polysaccharide from Escherichia coli O130. *Bioorg Khim* 2007a;**33**:57–60.

Perepelov AV, Liu B, Shevelev SD *et al*. Relatedness of the O-polysaccharide structures of Escherichia coli O123 and Salmonella enterica O58, both containing 4,6-dideoxy-4-{N-(S)-3-hydroxybutanoyl]-D-alanyl}amino-D-glucose; revision of the E. coli O123 O-polysaccharide structure. *Carbohydr Res* 2010c;**345**:825–9.

Perepelov AV, Naumenko OI, Senchenkova SN *et al*. Structure of the O-polysaccharide of Escherichia coli O60. *Russ Chem Bull Int Ed* 2018b;**67**:2131–4.

Perepelov AV, Ni Z, Wang Q *et al*. Structure and gene cluster of the O-antigen of Escherichia coli O109; chemical and genetic evidences of the presence of L-RhaN3N derivatives in the O-antigens of E. coli O109 and O119. *FEMS Immunol Med Microbiol* 2011e;**61**:47–53.

Perepelov AV, Senchenkova SN, Kalinchuk NA *et al*. Structure of O-polysaccharide of Escherichia coli O95: a disaccharide repeating unit containing D-fucose and D-threo-pent-2-ulose (xylulose). *Russ Chem Bull Int Ed Eng* 2018c;**67**:1931–3.

Perepelov AV, Senchenkova SN, Shashkov AS *et al*. Antigenic polysaccharides of bacteria. 41. Structures of the O-specific polysaccharides of Shigella dysenteriae types 4 and 5 revised by NMR spectroscopy. *Bioorg Khim* 2008b;**34**:460–7.

Perepelov AV, Shashkov AS, Guo X *et al*. Structure and genetics of the O-antigen of Escherichia coli O169 related to the O-antigen of Shigella boydii type 6. *Carbohydr Res* 2015f;**414**:46–50.

Perepelov AV, Shevelev SD, Liu B *et al*. Structures of the O-antigens of Escherichia coli O13, O129 and O135 related to the O-antigens of Shigella flexneri. *Carbohydr Res* 2010a;**345**:1594–9.

Perepelov AV, Wang Q, Filatov AV *et al*. Structures and gene clusters of the closely related O-antigens of Escherichia coli O46 and O134, both containing D-glucuronoyl-D-allothreonine. *Carbohydr Res* 2015b;**409**:20–24; ibid. 2016, **436**, 54.

Perepelov AV, Wang Q, Levina EA *et al*. Structure and gene cluster of the O-antigen of Escherichia coli O36. *Carbohydr Res* 2014b;**390**:46–49.

Perepelov AV, Wang Q, Liu B *et al*. Structure of the O-polysaccharide of Escherichia coli O61, another E. coli O-antigen that contains 5,7-diacetamido-3,5,7,9-tetradeoxy-L-glycero-D-galacto-non-2-ulosonic (di-N-acetyl-8-epilegionaminic) acid. *J Carbohydr Chem* 2009a;**28**:463–72.

Perepelov AV, Wang Q, Senchenkova SN *et al*. Structural and genetic studies of the O-antigen of Escherichia coli O163. *Carbohydr Res* 2015e;**404**:34–38.

Perepelov AV, Wang Q, Senchenkova SN *et al*. Structure and characterization of the gene cluster of the O-antigen of Escherichia coli O49 containing 4,6-dideoxy-4-[(S)-3-hydroxybutanoylamino]-D-glucose. *Biochemistry (Moscow)* 2008a;**73**:406–10.

Perepelov AV, Wang Q, Senchenkova SN *et al*. Structure and gene cluster of the O-antigen of Escherichia coli O102. *Carbohydr Res* 2012a;**361**:73–77.

Perepelov AV, Wang Q, Senchenkova SN *et al*. Structure and gene cluster of the O-antigen of Escherichia coli O110 containing

an amide of D-galacturonic acid with D-allothreonine. *Carbohydr Res* 2013b;**368**:57–60.

Perepelov AV, Wang Q, Senchenkova SN *et al*. Structure and gene cluster of the O-antigen of Escherichia coli O120. *Carbohydr Res* 2012b;**353**:106–10.

Perepelov AV, Wang Q, Senchenkova SN *et al*. Structure and gene cluster of the O-antigen of Escherichia coli O140. *Carbohydr Res* 2015c;**411**:33–36.

Perepelov AV, Wang Q, Senchenkova SN *et al*. Structure and gene cluster of the O-antigen of Escherichia coli O154. *Carbohydr Res* 2013c;**379**:51–54.

Perepelov AV, Wang Q, Senchenkova SN *et al*. Structure and gene cluster of the O-antigen of Escherichia coli O76. *Carbohydr Res* 2013a;**377**:14–17.

Perepelov AV, Wang Q, Senchenkova SN *et al*. Structure of a teichoic-acid like O-polysaccharide of Escherichia coli O29. *Carbohydr Res* 2006;**341**:2176–80.

Perepelov AV, Wang Q, Shashkov AS *et al*. Structure and genetics of the O-antigen of Escherichia coli O158. *Carbohydr Res* 2011f;**346**:2274–7.

Perepelov AV, Weintraub A, Liu B *et al*. The O-polysaccharide of Escherichia coli O112ac has the same structure as that of Shigella dysenteriae type 2 but is devoid of O-acetylation; a revision of the S. dysenteriae type 2 O-polysaccharide structure. *Carbohydr Res* 2008d;**343**:977–81.

Perepelov AV, Zhu H, Senchenkova SN *et al*. Structure and gene cluster of the O-antigen of Escherichia coli O19ab. *Carbohydr Res* 2011b;**346**:2812–5.

Perry MB, MacLean L, Griffith DW. Structure of the O-chain polysaccharide of the phenol-phase soluble lipopolysaccharide of Escherichia coli O:157:H7. *Biochem Cell Biol* 1986;**64**:21–8.

Perry MB, MacLean LL, Brisson J-R. The characterization of the O-antigen of Escherichia coli O64: K99 lipopolysaccharide. *Carbohydr Res* 1993;**248**:277–84.

Perry MB, MacLean LL. Structural characterization of the antigenic O-chain of the lipopolysaccharide of Escherichia coli serotype O65. *Carbohydr Res* 1999;**322**:57–66.

Perry MB, MacLean LL. Structure of the lipopolysaccharide O-chain of Yersinia enterocolitica O:5,27. *Biochem Cell Biol* 1987;**65**:1–7.

Pfoestl A, Hofinger A, Kosma P *et al*. Biosynthesis of dTDP-3-acetamido-3,6-dideoxy-α-D-galactose in Aneurinibacillus thermoaerophilus L420-91T. *J Biol Chem* 2003;**278**:26410–7.

Pfostl A, Zayni S, Hofinger A *et al*. Biosynthesis of dTDP-3-acetamido-3,6-dideoxy-α-D-glucose. *Biochem J* 2008;**410**:187–94.

Plainvert C, Bidet P, Peigne C *et al*. A new O-antigen gene cluster has a key role in the virulence of the Escherichia coli meningitis clone O45:K1:H7. *J Bacteriol* 2007;**189**:8528–36.

Pluschke G, Mayden J, Achtman M *et al*. Role of the capsule and the O-antigen in resistance of O18:K1 Escherichia coli to complement-mediated killing. *J Bacteriol* 1983;**42**:907–13.

Prehm P, Jann B, Jann K. The O9 antigen of Escherichia coli. Structure of the polysaccharide chain. *Eur J Biochem* 1976;**67**:53–56.

Pupo GM, Lan R, Reeves PR. Multiple independent origins of Shigella clones of Escherichia coli and convergent evolution of many of their characteristics. *Proc Natl Acad Sci USA* 2000;**97**:10567–72.

Qin X, Klein EJ, Galanakis E *et al*. Real-time PCR assay for detection and differentiation of shiga toxin-producing Escherichia coli from clinical samples. *J Clin Microbiol* 2015;**53**:2148–53.

Ratnayake S, Weintraub A, Widmalm G. Structural studies of the enterotoxigenic Escherichia coli (ETEC) O153 O-antigenic polysaccharide. *Carbohydr Res* 1994b;**265**:113–20.

Ratnayake S, Widmalm G, Weintraub A *et al*. Structural studies of the Escherichia coli O90 O-antigen polysaccharide. *Carbohydr Res* 1994a;**263**:209–15.

Raymond CK, Sims EH, Kas A *et al*. Genetic variation at the O-antigen biosynthetic locus in Pseudomonas aeruginosa. *J Bacteriol* 2002;**184**:3614–22.

Raynaud C, Meibom KL, Lety MA *et al*. Role of the wbt locus of Francisella tularensis in lipopolysaccharide O-antigen biogenesis and pathogenicity. *Infect Immun* 2007;**75**:536–41.

Reeves PR, Pacinelli E, Wang L. O antigen gene clusters of Yersinia pseudotuberculosis. *Adv Experim Med Biol* 2003;**529**:199–206.

Reeves PR, Wang L. Genomic organization of LPS-specific loci. *Curr Top Microbiol Immunol* 2002;**264**:109–35.

Reeves PR. Variation in O-antigens, niche-specific selection and bacterial populations. *FEMS Microbiol Lett* 1992;**100**:509–16.

Rick PD, Hubbard GL, Barr K. Role of the rfe gene in the synthesis of the O8 Antigen in Escherichia coli K-12. *J Bacteriol* 1994;**176**:2877–84.

Riley JG, Menggad M, Montoya-Peleaz PJ *et al*. The wbbD gene of E. coli strain VW187 (O7:K1) encodes a UDP-Gal: GlcNAc α-pyrophosphate-R β-1,3-galactosyltransferase involved in the biosynthesis of O7-specific lipopolysaccharide. *Glycobiology* 2005;**15**:605–13.

Rocchetta HL, Burrows LL, Pacan JC *et al*. Three rhamnosyltransferases responsible for assembly of the A-band D-rhamnan polysaccharide in Pseudomonas aeruginosa: a fourth transferase, WbpL, is required for the initiation of both A-band and B-band lipopolysaccharide synthesis. *Mol Microbiol* 1998;**28**:1103–19.

Rojas-Macias MA, Ståhle J, Lütteke T *et al*. Development of the ECODAB into a relational database for Escherichia coli O-antigens and other bacterial polysaccharides. *Glycobiology* 2015;**25**:341–7.

Rosano C, Bisso A, Izzo G *et al*. Probing the catalytic mechanism of GDP-4-keto-6-deoxy-D-mannose epimerase/reductase by kinetic and crystallographic characterization of site-specific mutants. *J Mol Biol* 2000;**303**:77–91.

Ruan X, Loyola DE, Marolda CL *et al*. The WaaL O-antigen lipopolysaccharide ligase has features in common with metal ion-independent inverting glycosyltransferases. *Glycobiology* 2012;**22**:288–99.

Rundlöf T, Weintraub A, Widmalm G. Structural determination of the O-antigenic polysaccharide from Escherichia coli O35 and cross-reactivity to Salmonella arizonae O62. *Eur J Biochem* 1998;**258**:139–43.

Rundlöf T, Weintraub A, Widmalm G. Structural studies of the enteroinvasive Escherichia coli (EIEC) O28 O-antigenic polysaccharide. *Carbohydr Res* 1996;**291**:127–39.

Rush JS, Alaimo C, Robbiani R *et al*. A novel epimerase that converts GlcNAc-P-P-undecaprenol to GalNAc-P-P-undecaprenol in Escherichia coli O157. *J Biol Chem* 2010;**285**:1671–80.

Russo TA, Johnson JR. Proposal for a new inclusive designation for extraintestinal pathogenic isolates of Escherichia coli: ExPEC. *J Infect Dis* 2000;**181**:1753–4.

Samuel G, Reeves PR. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr Res* 2003;**338**:2503–19.

Sarkar S, Ulett GC, Totsika M *et al*. Role of capsule and O antigen in the virulence of uropathogenic Escherichia coli. *PLoS ONE* 2014;**9**: e94786.

Scheutz F, Cheasty T, Woodward D *et al*. Designation of O174 and O175 to temporary O groups OX3 and OX7, and six new E. coli O groups that include Verocytotoxin-producing E. coli (VTEC): O176, O177, O178, O179, O180 and O181. *Apmis* 2004;**112**:569–84.

Senchenkova SN, Feng L, Yang J *et al*. Structural and genetic characterization of the Shigella boydii type 10 and type 6 O antigens. *J Bacteriol* 2005;**187**:2551–4.

Senchenkova SN, Guo X, Filatov AV *et al*. Structure and gene cluster of the O-antigen of Escherichia coli O80. *Carbohydr Res* 2016b;**432**:83–87.

Senchenkova SN, Guo X, Naumenko OI *et al*. Structure and genetics of the O-antigens of Escherichia coli O182-O187. *Carbohydr Res* 2016d;**435**:58–67.

Senchenkova SN, Hou W, Naumenko OI *et al*. Structure and genetics of a glycerol 2-phosphate-containing O-specific polysaccharide of Escherichia coli O33. *Carbohydr Res* 2018;**460**:47–50.

Senchenkova SN, Knirel YA, Shashkov AS *et al*. The O-specific polysaccharide chain of Campylobacter fetus serotype B lipopolysaccharide is a linear D-rhamnan terminated with 3-O-methyl-D-rhamnose (D-acofriose). *Eur J Biochem* 1996;**239**:434–8.

Senchenkova SN, Xu G, Qi Y *et al*. New fructan - the O-specific polysaccharide from Escherichia coli O92. *Russ Chem Bull Int Ed Eng* 2017;**66**:1304–7.

Senchenkova SN, Zhang Y, Perepelov AV *et al*. Structure and biosynthesis gene cluster of the O-antigen of Escherichia coli O12. *Biochemistry (Moscow)* 2016a;**81**:401–6.

Senchenkova SN, Zhang Y, Perepelov AV *et al*. Structure and gene cluster of the O-antigen of Escherichia coli O165 containing 5-N-acetyl-7-N-[(R)-3-hydroxybutanoyl]pseudaminic acid. *Glycobiology* 2016c;**26**:345–2.

Sengupta P, Bhattacharyya T, Shashkov AS *et al*. Structure of the O-specific side chain of the Escherichia coli O128 lipopolysaccharide. *Carbohydr Res* 1995;**277**:283–90.

Shashkov AS, Senchenkova SN, Naumenko OI *et al*. Escherichia coli O106, a new member of a group of enteric bacteria sharing an O-polysaccharide backbone structure. *Russ Chem Bull Int Ed Eng* 2018;**67**:1538–41.

Shashkov AS, Senchenkova SN, Perepelov AV *et al*. Structure of the O-polysaccharides of Escherichia coli O162 containing 4-deoxy-D-arabino-hexose and structurally related O-polysaccharides of E. coli O101. *Russ Chem Bull Int Ed Eng* 2017;**66**:732–4.

Shashkov AS, Wang T, Perepelov AV *et al*. Structure elucidation and biosynthesis gene cluster organization of the O-antigen of Escherichia coli O170. *Carbohydr Res* 2015;**417**:11–14.

Shashkov AS, Yang B, Senchenkova SN *et al*. Structures and genetics of biosynthesis of glycerol 1-phosphate-containing O-polysaccharides of Escherichia coli O28ab, O37, and O100. *Carbohydr Res* 2016a;**426**:26–32.

Shashkov AS, Zhang W, Perepelov AV *et al*. Structure of the O-polysaccharide of Escherichia coli O132. *Carbohydr Res* 2016b;**427**:44–47.

Shashkov AS, Zhang Y, Sun Q *et al*. Structure and gene cluster of the O-antigen of Escherichia coli O133. *Carbohydr Res* 2016c;**430**:82–84.

Simpson BW, May JM, Sherman DJ *et al*. Lipopolysaccharide transport to the cell surface: biosynthesis and extraction from the inner membrane. *Philos Trans R Soc Lond B Biol Sci* 2015;**370**:20150029.

Somoza JR, Menon S, Schmidt H *et al*. Structural and kinetic analysis of Escherichia coli GDP-mannose 4,6 dehydratase provides insights into the enzyme's catalytic mechanism and regulation by GDP-fucose. *Structure* 2000;**8**:123–35.

Staaf M, Urbina F, Weintraub A *et al*. Structural elucidation of the O-antigenic polysaccharide from Escherichia coli O21 and the enteroaggregative Escherichia coli strain 105. *Eur J Biochem* 1999a;**266**:241–5.

Staaf M, Urbina F, Weintraub A *et al*. Structure determination of the O-antigenic polysaccharide from the enterotoxigenic Escherichia coli (ETEC) O101. *Carbohydr Res* 1997;**297**:297–9.

Staaf M, Weintraub A, Widmalm G. Structure determination of the O-antigenic polysaccharide from the enteroinvasive Escherichia coli O136. *Eur J Biochem* 1999b;**263**:656–61.

Staaf M, Widmalm G, Weintraub A *et al*. Structural elucidation of the O-antigenic polysaccharide from Escherichia coli O44:H18. *Eur J Biochem* 1995;**233**:473–7.

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

Stevenson G, Andrianopoulos K, Hobbs M *et al*. Organization of the Escherichia coli K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid. *J Bacteriol* 1996;**178**:4885–93.

Stevenson G, Dieckelmann M, Reeves PR. Determination of glycosyltransferase specificities for the Escherichia coli O111 O antigen by a generic approach. *Appl Environ Microbiol* 2008;**74**:1294–8.

Stevenson G, Neal B, Liu D *et al*. Structure of the O antigen of Escherichia coli K-12 and the sequence of its rfb gene cluster. *J Bacteriol* 1994;**176**:4144–56.

Ståhle J, Fontana C, Weintraub A *et al*. Elucidation of the O-antigen structure of Escherichia coli O63, *Glycobiology* 2019;**29**:179–87.

Sugiyama T, Kido N, Kato Y *et al*. Generation of Escherichia coli O9a serotype, a subtype of E. coli O9, by transfer of the wb* gene cluster of Klebsiella O3 into E. coli via recombination. *J Bacteriol* 1998;**180**:2775–8.

Svensson MV, Weintraub A, Widmalm G. Structural elucidation of the O-antigenic polysaccharide from Escherichia coli O175. *Carbohydr Res* 2011a;**346**:449–53.

Svensson MV, Weintraub A, Widmalm G. Structural studies of the O-antigenic polysaccharide from Escherichia coli O177. *Carbohydr Res* 2011b;**346**:2300–3.

Thrane SW, Taylor VL, Lund O *et al*. Application of WGS data for O-specific antigen analysis and in silico serotyping of Pseudomonas aeruginosa isolates. *J Clin Microbiol* 2016;**54**:1782–8.

Torgov VI, Shashkov AS, Jann B *et al*. NMR reinvestigation of two N-acetylneuraminic acid-containing O-specific polysaccharides (O56 and O24) of Escherichia coli. *Carbohydr Res* 1995;**272**:73–90.

Torgov VI, Shashkov AS, Kochanowski H *et al*. NMR analysis of the structure of the O88 polysaccharide (O88 antigen) of Escherichia coli O88:K-:H25. *Carbohydr Res* 1996;**283**:223–7.

Urbina F, Nordmark EL, Yang Z *et al*. Structural elucidation of the O-antigenic polysaccharide from the enteroaggregative Escherichia coli strain 180/C3 and its immunochemical relationship with E. coli O5 and O65. *Carbohydr Res* 2005;**340**: 645–50.

Valvano MA, Messner P, Kosma P. Novel pathways for biosynthesis of nucleotide-activated glycero-manno-heptose precursors of bacterial glycoproteins and cell surface polysaccharides. *Microbiology* 2002;**148**:1979–89.

Valvano MA Export of O-specific lipopolysaccharide. *Front Biosci* 2003;**8**:s452–471.

Vasil'ev V, Zakharova I, Shaskov A. Structure of the O-specific polysaccharides of Escherichia coli Str. O20ab:K84:H34 (145) and O20ac:K61:H- according to 13C NMR spectroscopy. *Bioorg Khim* 1982;**8**:120–5.

Vasil'ev VN, Zakharova IY. Structure of the determinant group of the O-specific polysaccharide of Escherichia coli O20:K84:H34(145). *Bioorg Khim* 1976;**2**:199–206.

Vila J, Sáez-López E, Johnson JR et al. Escherichia coli: an old friend with new tidings. *FEMS Microbiol Rev* 2016;**40**:437–63.

Vilchez S, Lundborg M, Urbina F et al. Structural studies of the O-antigenic polysaccharides from the enteroaggregative Escherichia coli strain 94/D4 and the international type strain Escherichia coli O82. *Carbohydr Res* 2009;**344**:2528–32.

Vinogradov E, Frirdich E, MacLean LL et al. Structures of lipopolysaccharides from Klebsiella pneumoniae. Elucidation of the structure of the linkage region between core and polysaccharide O chain and identification of the residues at the non-reducing termini of the O chains. *J Biol Chem* 2002;**277**:25070–81.

Vinogradov E, MacLean LL, Perry MB. Structural determination of the O-antigenic polysaccharide of enteropathogenic Escherichia coli O103:H2. *Can J Microbiol* 2010;**56**:366–72.

Wang L, Huskic S, Cisterne A et al. The O-antigen gene cluster of Escherichia coli O55:H7 and identification of a new UDP-GlcNAc C4 epimerase gene. *J Bacteriol* 2002;**184**:2620–5.

Wang L, Wang Q, Reeves PR. The variation of O antigens in Gram-negative bacteria. *Subcell Biochem* 2010;**53**:123–52.

Wang Q, Ding P, Perepelov AV et al. Characterization of the dTDP-D-fucofuranose biosynthetic pathway in Escherichia coli O52. *Mol Microbiol* 2008;**70**:1358–67.

Wang Q, Perepelov AV, Beutin L et al. Structural and genetic characterization of the Escherichia coli O180 O antigen and identification of a UDP-GlcNAc 6-dehydrogenase. *Glycobiology* 2012b;**22**:1321–31.

Wang Q, Perepelov AV, Feng L et al. Genetic and structural analyses of Escherichia coli O107 and O117 O-antigens. *FEMS Immunol Med Microbiol* 2009;**55**:47–54.

Wang Q, Perepelov AV, Wen L et al. Identification of the two glycosyltransferase genes responsible for the difference between Escherichia coli O107 and O117 O-antigens. *Glycobiology* 2012a;**22**:281–7.

Wang S, Czuchry D, Liu B et al. Characterization of two UDP-Gal:GalNAc-diphosphate-lipid β-1,3-galactosyltransferases WbwC from Escherichia coli serotypes O104 and O5. *J Bacteriol* 2014;**196**:3122–33.

Wang W, Perepelov AV, Feng L et al. A group of Escherichia coli and Salmonella enterica O antigens sharing a common backbone structure. *Microbiology* 2007;**153**:2159–67.

Wang Y, Xu Y, Perepelov AV et al. Biochemical characterization of dTDP-D-Qui4N and dTDP-D-Qui4NAc biosynthetic pathways in Shigella dysenteriae type 7 and Escherichia coli O7. *J Bacteriol* 2007;**189**:8626–35.

Weintraub A, Leontein K, Widmalm G et al. Structural studies of the O-antigenic polysaccharide of an enteroaggregative Escherichia coli strain. *Eur J Biochem* 1993;**213**:859–64.

West NP, Sansonetti P, Mounier J et al. Optimization of virulence functions through glucosylation of Shigella LPS. *Science* 2005;**307**:1313–7.

Whitfield C, Trent MS. Biosynthesis and export of bacterial lipopolysaccharides. *Annu Rev Biochem* 2014;**83**:99–128.

Widmalm G, Leontein K. Structural studies of the Escherichia coli O127 O-antigen polysaccharide. *Carbohydr Res* 1993;**247**:255–62.

Wildschutte H, Lawrence JG. Differential Salmonella survival against communities of intestinal amoebae. *Microbiology* 2007;**153**:1781–9.

Wu Y, Lau HK, Lee T et al. Whole Genome Sequencing-Based in silico Serotyping Improves Accuracy for Shigella Identification. *Appl Environ Microbiol* 2019;**85**:e00165–19.

Yang B, Senchenkova SN, Naumenko OI et al. Structural and genetic relatedness of the O antigens of Escherichia coli O50 and O2. *Carbohydr Res* 2018;**15**:8–11.

Yao Z, Valvano MA. Genetic analysis of the O-specific lipopolysaccharide biosynthesis region (rfb) of Escherichia coli K-12 W3110: identification of genes that confer group 6-specificity to Shigella flexneri serotypes Y and 4a. *J Bacteriol* 1994;**176**:4133–43.

Yildirim H, Weintraub A, Widmalm G. Structural studies of the O-polysaccharide from the Escherichia coli O77 lipopolysaccharide. *Carbohydr Res* 2001;**333**:179–83.

Yi W, Bystricky P, Yao Q et al. Two different O-polysaccharides from Escherichia coli O86 are produced by different polymerization of the same O-repeating unit. *Carbohydr Res* 2006;**341**:100–8.

Yi W, Shao J, Zhu L et al. Escherichia coli O86 O-antigen biosynthetic gene cluster and stepwise enzymatic synthesis of human blood group B antigen tetrasaccharide. *J Am Chem Soc* 2005;**127**:2040–1.

Zaccheus MV, Broeker NK, Lundborg M et al. Structural studies of the O-antigen polysaccharide from Escherichia coli TD2158 having O18 serogroup specificity and aspects of its interaction with the tailspike endoglycosidase of the infecting bacteriophage HK620. *Carbohydr Res* 2012;**357**:118–25.

Zdorovenko EL, Golomidova AK, Liu B et al. Structure of the O-polysaccharide from Escherichia coli HS1/2 and type strain of E. coli O87. *Carbohydr Res* 2015;**412**:15–8; ibid. 2018;**464**:1.

Zdorovenko EL, Ovod VV, Zatonsky GV et al. Location of the O-methyl groups in the O polysaccharide of Pseudomonas syringae pv. *Phaselicola Carbohydr Res* 2001;**330**:505–10.

Zdorovenko EL, Wang Y, Shashkov AS et al. O-Antigens of Escherichia coli strains O81 and HS3-104 are structurally and genetically related, except O-antigen glucosylation in E. coli HS3-104. *Biochemistry (Moscow)* 2018;**83**:534–41.

Zhang S, Yin Y, Jones MB et al. Salmonella serotype determination utilizing high-throughput genome sequencing data. *J Clin Microbiol* 2015;**53**:1685–92.

Zhao G, Liu J, Liu X et al. Cloning and characterization of GDP-perosamine synthetase (Per) from Escherichia coli O157:H7 and synthesis of GDP-perosamine in vitro. *Biochem Biophys Res Commun* 2007b;**363**:525–30.

Zhao G, Perepelov AV, Senchenkova SN et al. Structural relation of the antigenic polysaccharides of Escherichia coli O40, Shigella dysenteriae type 9, and E. coli K47. *Carbohydr Res* 2007a;**342**:1275–9.

Zhao X, Creuzenet C, Belanger M et al. WbpO, a UDP-N-acetyl-D-galactosamine dehydrogenase from Pseudomonas aeruginosa serotype O6. *J Biol Chem* 2000;**275**:33252–9.

Zhu H, Perepelov AV, Senchenkova SN et al. Structure and gene cluster of the O-antigen of Escherichia coli O41. *Carbohydr Res* 2012;**349**:86–89.

Zuccotti S, Zanardi D, Rosano C et al. Kinetic and crystallographic analyses support a sequential-ordered bi bi catalytic mechanism for Escherichia coli glucose-1-phosphate thymidyltransferase. *J Mol Biol* 2001;**313**:831–43.