

RESEARCH ARTICLE

Genetic diversity of the O antigens of *Proteus* species and the development of a suspension array for molecular serotyping

Xiang Yu^{1,2,3,4}, Agnieszka Torzewska⁵, Xinjie Zhang^{1,2,3,4}, Zhiqiu Yin^{1,2,3,4}, Dominika Drzewiecka⁵, Hengchun Cao^{1,2,3,4}, Bin Liu^{1,2,3,4}, Yuriy A. Knirel⁶, Antoni Rozalski⁵, Lei Wang^{1,2,3,4*}

1 Key Laboratory of Molecular Microbiology and Technology of the Ministry of Education, TEDA College, Nankai University, Tianjin, P. R. China, **2** TEDA Institute of Biological Sciences and Biotechnology, Nankai University, Tianjin, P. R. China, **3** Tianjin Research Center for Functional Genomics and Biochips, TEDA College, Nankai University, Tianjin, P. R. China, **4** Tianjin Key Laboratory of Microbial Functional Genomics, TEDA College, Nankai University, Tianjin, P. R. China, **5** Department of Immunobiology of Bacteria, Department of General Microbiology Institute of Microbiology, Biotechnology and Immunology, Faculty of Biology and Environmental Protection, University of Lodz, Lodz, Poland, **6** N.D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Moscow, Russian Federation

* wanglei@nankai.edu.cn



OPEN ACCESS

Citation: Yu X, Torzewska A, Zhang X, Yin Z, Drzewiecka D, Cao H, et al. (2017) Genetic diversity of the O antigens of *Proteus* species and the development of a suspension array for molecular serotyping. PLoS ONE 12(8): e0183267. <https://doi.org/10.1371/journal.pone.0183267>

Editor: Mikael Skurnik, University of Helsinki, FINLAND

Received: March 24, 2017

Accepted: August 1, 2017

Published: August 17, 2017

Copyright: © 2017 Yu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by the National Natural Science Foundation of China (<http://www.nsf.gov.cn/>; No. 31530083 and No. 31371259 to LW, No. 81471904 and No. 31270003 to BL), and the International Science & Technology Cooperation Program of China (<http://www.cistc.gov.cn/>; No. 2013DFR30640 to LW). The funders had no role in study design, data collection and

Abstract

Proteus species are well-known opportunistic pathogens frequently associated with skin wound and urinary tract infections in humans and animals. O antigen diversity is important for bacteria to adapt to different hosts and environments, and has been used to identify serotypes of *Proteus* isolates. At present, 80 *Proteus* O-serotypes have been reported. Although the O antigen structures of most *Proteus* serotypes have been identified, the genetic features of these O antigens have not been well characterized. The O antigen gene clusters of *Proteus* species are located between the *cpxA* and *secB* genes. In this study, we identified 55 O antigen gene clusters of different *Proteus* serotypes. All clusters contain both the *wzx* and *wzy* genes and exhibit a high degree of heterogeneity. Potential functions of O antigen-related genes were proposed based on their similarity to genes in available databases. The O antigen gene clusters and structures were compared, and a number of glycosyltransferases were assigned to glycosidic linkages. In addition, an O serotype-specific suspension array was developed for detecting 31 *Proteus* serotypes frequently isolated from clinical specimens. To our knowledge, this is the first comprehensive report to describe the genetic features of *Proteus* O antigens and to develop a molecular technique to identify different *Proteus* serotypes.

Introduction

Proteus species are gram-negative bacterial opportunistic pathogens that belong to the Enterobacteriaceae family [1]. The genus *Proteus* is widely distributed in the natural environment and in the microflora of human and animal intestines [1]. Under favorable conditions, they

analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

most commonly cause skin wound and urinary tract infections (UTIs) in humans and animals [2–4], and roles in rheumatoid arthritis have also been reported [5]. Currently, this genus contains seven named species, *P. mirabilis*, *P. penneri*, *P. vulgaris*, *P. myxofaciens*, *P. hauseri*, *P. terrae*, and *P. cibarius* as well as three unnamed *Proteus* genomospecies, 4, 5, and 6 [6–9]. Among these, *P. mirabilis*, *P. penneri*, and *P. vulgaris* are the most common human pathogens, and isolates of *P. mirabilis* cause UTIs with the highest frequency [4]. *P. myxofaciens*, *P. terrae*, and *P. cibarius* have no pathogenicity for humans [3, 8, 10].

Proteus species express a series of virulence factors that are associated with infection processes and disease, such as fimbria, flagella, hemolysins, urease, proteases, amino acid deaminases, lipopolysaccharide (LPS) and capsular polysaccharides (CPSs) [3, 11–13]. LPS is an endotoxin that constitutes the outer cell membrane of the gram-negative bacteria and is its most variable component [14]. LPS is thought to play an important role in the process of the UTIs and to affect both bladder and kidney stone formation [15–17]. Furthermore, LPS confers protection against serum-mediated bactericidal activity to bacteria [18].

LPS consists of three parts: a lipid A anchor, a core oligosaccharide, and an O-specific polysaccharide (OPS) [19]. The OPS consists of oligosaccharide repeating units (O units) that usually contains 2 to 8 sugar residues [20]. The OPS is the most variable component of the LPS, which defines the serological specificity of gram-negative bacteria [19, 21]. OPS variation is predominantly determined by the types of sugars present as well as the order of sugar residues and the linkages between them [21, 22]. The OPS is essential for bacterial survival, and the loss of OPS causes many bacteria serum-sensitive or affects their virulence in another way [21, 23].

The OPS synthesis related genes usually form a gene cluster that is located at a fixed position on the chromosome [19, 21]. For example, in *Salmonella*, *Escherichia coli*, and *Shigella*, the O antigen gene clusters most commonly map between the *galF* and *gnd* genes [19, 21]. One or more of the genes involved in OPS synthesis may sometimes map outside the gene clusters [19, 21]. The O antigen gene clusters contain three classes of genes: nucleotide sugar biosynthesis pathway genes, glycosyltransferase (GT) genes, and O antigen processing genes [19, 21, 23]. Sugars commonly found in other polysaccharide structures or involved in metabolism, such as galactose (Gal), glucose (Glc), and *N*-acetylglucosamine (GlcNAc), are usually synthesized by enzymes encoded by genes outside the O antigen gene cluster [19, 21]. There are three different pathways to synthesize and translocate O antigen: the Wzx/Wzy pathway, which is most frequently utilized; the synthase pathway; and the ATP-binding cassette transporter pathway (ABC pathway) [19, 21, 24]. In the Wzx/Wzy pathway, the O units are synthesized by initial transfer of a sugar phosphate, then sequential transfer of the other sugars from their nucleotide sugars donor to the carrier undecaprenyl phosphate (UndP) [25]. The assembled O units are flipped across the cell membrane by Wzx and then polymerized by Wzy to form polysaccharide chains [25]. In *E. coli*, *Shigella*, and *Salmonella*, the chain-length determinant *Wzz* imposes a modal chain-length distribution on the OPS, loss of *wzz* results in the uncontrolled polymerization of O-units by Wzy-producing nonmodal chain-lengths ranging from short to long, the principle is still unknown [26]. The OPS is eventually ligated to the lipid A core to form LPS [27].

Detection of bacterial serotypes is critical for prevention and control of pathogens. However, traditional antiserum serotyping methods are laborious and cross-reactive, many molecular and chemical serotyping techniques have been developed. Such as real-time PCR assays and short sequencing assays based on serotype-specific genes [28–29], or chemometric analysis of attenuated total reflectance infrared spectra based on defined LPS structures [30]. Recent development of gene chip technology, including solid phase arrays and liquid bead-based suspension arrays, has given us a more sensitive and accurate method to identify bacterial serotypes [31]. The suspension array system is based on microspheres labeled with a unique dye

combination. The microspheres are coupled with specific probes for targets which are amplified from the samples using biotin-labeled primers. The fluorescent emission of the target analyte is measured by exciting the fluorescent reporter bound to the microspheres [31].

Currently, according to OPS diversity, *Proteus* species are divided into 80 different serotypes [32–34]. We previously reported 5 O antigen gene clusters (O3ab, O10, O23ac, O27, and O47) [35]. In this study, we identified 55 new O antigen gene clusters. Collectively, we have thus characterized 60 O antigen gene clusters from different *Proteus* serotypes (O1–O3, O5–O6, O8–O14, O16–O21, O23–O34, O36–O37, O40–O42, O44–O45, O47–O48, O50–O62, O65, O67, O69, and O71–O75). All these clusters are located between *cpxA* and *secB* and contain both *wzx* and *wzy* genes. They include genes associated with nucleotide sugar biosynthesis, sugar transfer, O antigen processing, and several other genes. In most cases, the O antigen gene clusters correspond to known O antigen structures. The *wzx* and *wzy* genes in these 60 *Proteus* strains were polytropic, which provided a basis for rapid molecular serotyping. Using the *wzx* and *wzy* genes, we developed a PCR-based suspension array to distinguish 31 different *Proteus* serotypes (O1–O3, O5–O6, O8–O14, O17–O21, O23–O24, O27, O29–O34, O36, O40, O42, O45, and O47), which are frequently isolated from clinical specimens. More serotypes can be added to this array in the future by designing new probes based on their *wzx* or *wzy* genes.

Materials and methods

Bacterial strains

All of the *Proteus* strains used in this study are shown in [S1 Table](#), which were provided by the Department of Immunobiology of Bacteria and the Department of General Microbiology at the Institute of Microbiology, Biotechnology and Immunology, Faculty of Biology and Environmental Protection, University of Lodz (Lodz, Poland).

Genomic DNA extraction and O antigen gene cluster amplification

The *Proteus* strains were grown in Luria Broth and then harvested by centrifugation [36]. Genomic DNA samples were isolated using a Bacteria Extraction Kit (CW BIO Co., Ltd, China). Primers wl_31262 (5′ –GAGTTATTACGHGAAACGGTAAAAGC–3′) and wl_31263 (5′ –GTTAACTTTGATGCGTTGTTTATGAACTA–3′) designed based on the *cpxA* and *secB* genes, respectively, were used to amplify the *Proteus* O antigen gene clusters [35]. The PCR program used was as follows: an initial denaturation at 95°C for 3 min, followed by 30 cycles of denaturation at 95°C for 45 s, annealing at 55°C for 45 s, and extension at 68°C for 15 min with a final extension at 68°C for 5 min [35].

O antigen gene cluster sequencing and analysis

The PCR products were fragmented with DNase I, then the fragments were cloned into pGEM-T Easy vector to construct a library as described previously [37]. Sequencing was performed using an ABI 3730 automated DNA sequencer (Applied Biosystems, Foster City, CA), with 12–20 fold coverage of the O antigen gene clusters. Sequencing data were assembled using the Staden package and were annotated by Artemis [38, 39]. Use TBLAST and PSI-BLAST to search available database, including the Pfam protein database and the GenBank database, and to identify potential functions of the O antigen synthesis related genes [40]. The potential transmembrane segments were identified using the TMHMM 2.0 program [40]. The GT genes were divided into homology groups (HGs) using the OrthoMCL program v2.0 [41] with a 50% protein sequence identity used as the cut-off. ClustalW v2.0 was used for sequences

alignment, and JC69 module and phyML v3.0 were used to construct maximum likelihood trees [42].

Development of PCR system

DNA from different samples was amplified using the Hot Start PCR Kit (Promega, Madison, WI). PCR primers designed based on specific *wzx/wzy* genes were used to generate PCR fragments of 100 to 495 bp (S2 Table) [43]. The reverse primer was biotinylated at the 5'-end that can be combined with microspheres coupling the dye streptavidin-R-phycoerythrin. The median fluorescence intensities (MFI) were detected using the Bio-Plex 100 suspension array system (Bio-Rad). A single multiplex PCR system was used to amplify the *wzx/wzy* genes as follows: an initial denaturation at 95°C for 5 min, followed by 30 cycles of denaturation at 94°C for 45 s, annealing at 50°C for 1 min, and extension at 72°C for 1 min, and the final extension was done at 72°C for 10 min [31]. The PCR products were then used directly in the hybridization reaction to couple beads.

Probes design and beads coupling

Serotype-specific probes were designed based on the *wzx/wzy* genes (S3 Table). BioEdit software 7.0 version was used for multiple sequence alignments. The carboxylated beads (Bio-Rad, Hercules, CA) were coupled to specific probes with an amino C-12 modification at the 3'-end (AuGCT, China).

Hybridization and staining

17 µl of the biotinylated amplicon was mixed with 33 µl of the bead mixture containing 2,500 beads in a 1.5× tetramethylammonium chloride (TMAC) solution (Sigma, St. Louis, MO). Then the mixture was denatured at 95°C for 5 min followed by hybridization at 55°C for 15 min. The hybridization product was collected using centrifugation at 8000 rpm and resuspended using 75 µl 1× TMAC solution containing 10 ng/mL streptavidin-R-phycoerythrin (Molecular Probes, Eugene, OR), then incubated at 55°C for 10 min.

Data acquisition and analysis

The fluorescence intensities of the beads were analyzed using a Bio-Plex 100 suspension array system (Bio-Rad). The MFIs were calculated from 100 replicate measurements using the digital signal processor and the Bio-Plex Manager software 4.1. A positive result was defined as an MFI > 150 and a signal/background (S/B = MFI/Blank) > 6.0 [31].

Results and discussion

In this study, we identified 55 new *Proteus* O antigen gene clusters. Combined with the 5 previously published gene clusters (O3ab, O10, O23ac, O27, and O47) [35], we have a total of 60 *Proteus* O antigen gene clusters characterized. All these clusters are located between *cpxA* and *secB* and contain both *wzx* and *wzy* genes, various GT genes, and nucleotide sugar synthesis genes (Fig 1). The GC content of these O antigen synthesis related genes ranged from 19.5 to 35.7%, which is lower than the rest of the *Proteus* genome (38.9%) [44], indicating that the O antigen gene clusters of *Proteus* species may originate from other bacteria [35].

Comparison of *Proteus* O antigen gene clusters and structures

Genes related to the biosynthesis of common sugar nucleotide precursors (such as UDP-GlcNAc, UDP-Glc, and UDP-Gal) are not located in the O antigen gene cluster [20, 22]. Genes

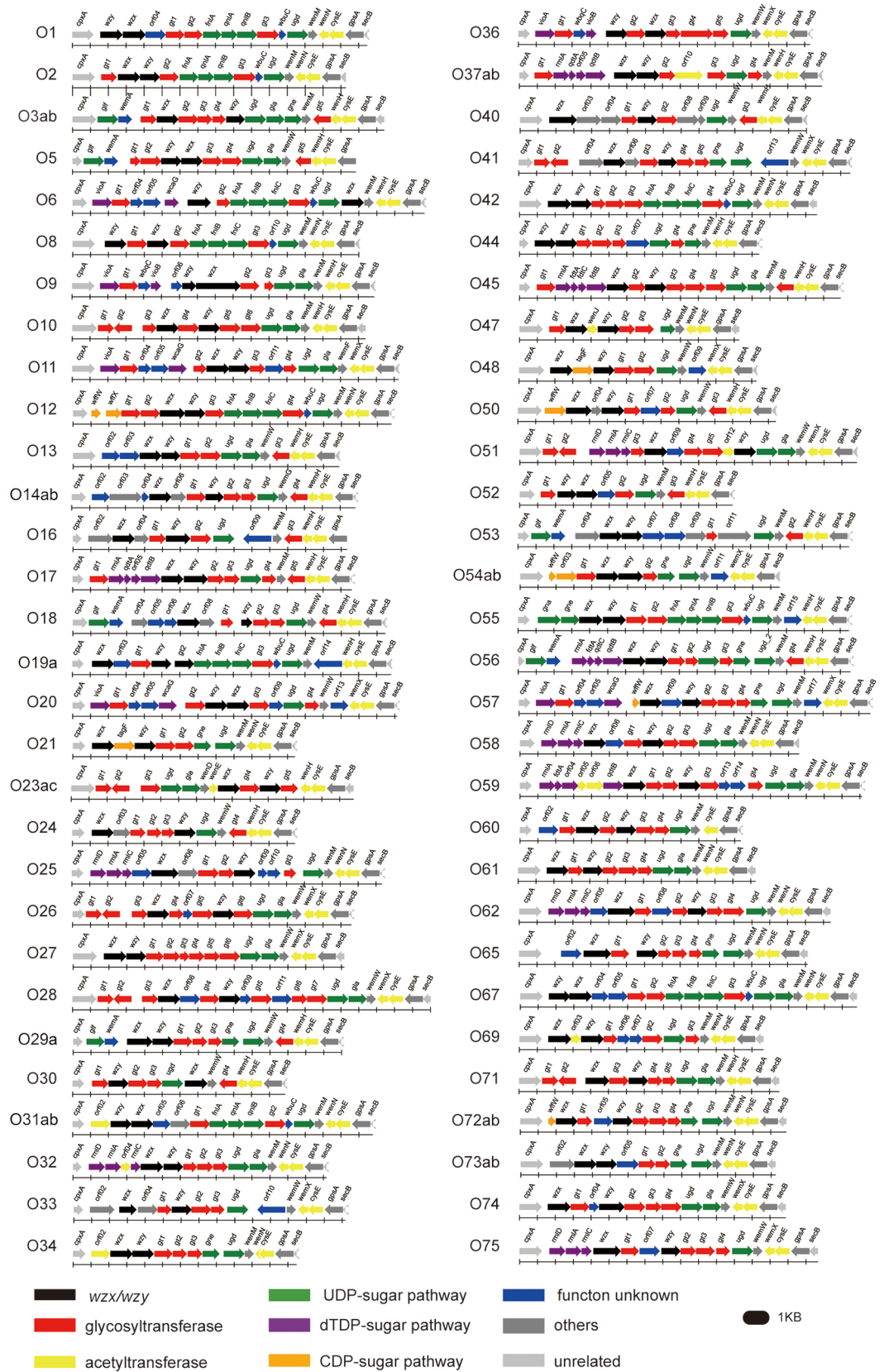


Fig 1. The O antigen gene clusters from the 60 *Proteus* serotypes. The sequences of the 60 *Proteus* O antigen gene clusters have been deposited to GenBank with accession numbers KY710685 to KY710739.

<https://doi.org/10.1371/journal.pone.0183267.g001>

related to the biosynthesis of rare monosaccharide precursors (such as UDP-QuiN, UDP--FucN, dTDP-Qui3N, dTDP-Fuc3N, dTDP-Rha) were typically located in the O antigen gene cluster [19, 21]. Based on these features, we compared the O antigen gene clusters and structures in *Proteus* (Figs 1 and 2, S4 Table) [32]. Of the 60 *Proteus* O antigen gene clusters, 57 (95% of the collection) were found to correspond to their known O antigen structures. Next, the genetic and structural consistency of rare monosaccharides will be described.

There are 4 *Proteus* O antigens whose structures include QuiNac (O1, O2, O31, and O55), and their gene clusters all contain *fnlA-qlnAB*. There are 6 *Proteus* O antigens whose structures include FucNac (O6, O8, O12, O19a, O42, and O67), and their gene clusters all contain *fnlABC*. UDP-GlcNAc is converted to UDP-2-acetamido-2,6-dideoxy-β-L-lyxo-hexos-4-ulose by *fnlA* [45], which can be further converted to UDP-L-QuiNac by *qlnAB* or to UDP-L-FucNac by *fnlBC* [45–47].

There are 2 *Proteus* O antigens whose structures include Fuc3Nac (O17 and O45). The O45 gene cluster contains *rmlA-fdtABC*, and the O17 gene cluster contains *rmlA-fdtAB*. There are 2 *Proteus* O antigens whose structures include Qui3Nac (O56 and O59), and their gene

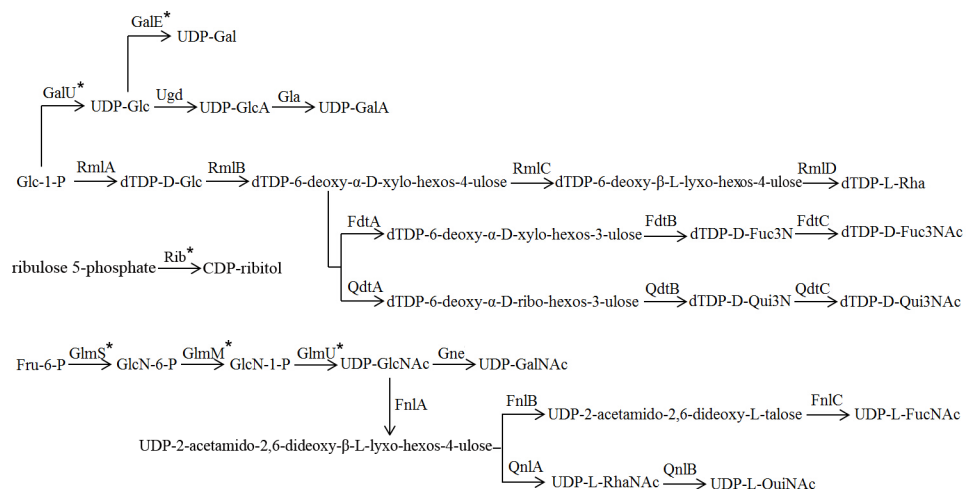


Fig 2. Biosynthesis pathways for the sugars in *Proteus* O antigens. GalU, UTP-glucose-1-phosphate uridylyltransferase; GalE, UDP-glucose-4-epimerase; Ugd, UDP-glucose 6-dehydrogenase; Gla, UDP-glucuronate 4-epimerase; GlmS, glutamine:fructose-6-phosphate transaminase; GlmM, phosphoglucosamine mutase; GlmU, UDP-N-acetyl-glucosamine pyrophosphorylase; Gne, UDP-N-acetylglucosamine-4-epimerase; RmlA, glucose-1-phosphate thymidyltransferase; RmlB, dTDP-D-glucose 4,6-dehydratase; RmlC, dTDP-6-deoxy-α-D-xylo-hexos-4-ulose 3,5-epimerase; RmlD, dTDP-6-deoxy-β-L-lyxo-hexos-4-ulose 4-reductase; FdtA, dTDP-6-deoxy-α-D-xylo-hexos-4-ulose 3,4-isomerase; FdtB, dTDP-6-deoxy-α-D-xylo-hexos-3-ulose aminase; FdtC, dTDP-D-Fuc3N acetylase; QdtA, dTDP-6-deoxy-α-D-xylo-hexos-4-ulose 3,4-isomerase; QdtB, dTDP-6-deoxy-α-D-ribo-hexos-3-ulose aminase; QdtC, dTDP-D-Qui3N acetylase; FnlA, UDP-D-GlcNAc 4,6-dehydratase, 3- and 5-epimerase; FnlB, UDP-2-acetamido-2,6-dideoxy-β-L-lyxo-hexos-4-ulose 4-reductase; FnlC, UDP-2-acetamido-2,6-dideoxy-L-talose 2-epimerase; QnlA, UDP-2-acetamido-2,6-dideoxy-β-L-lyxo-hexos-4-ulose 4-reductase; QnlB, UDP-L-RhaNac 2-epimerase; * indicates the genes located outside the O antigen gene cluster (all enzymes encoded by these genes can be found in the 68 *Proteus* genomes; the amino acid sequence identities of GalU to the homolog in *E. coli* K12 are 75.44–76.43%, the amino acid sequence identities of GalE to the homolog in *E. coli* K12 are 59.13–65.26%, the amino acid sequence identities of Rib to the homolog in *E. coli* K12 are 94.29–94.6%, the amino acid sequence identities of GlmS to the homolog in *E. coli* K12 are 78.89–82.79%, the amino acid sequence identities of GlmM to the homolog in *E. coli* K12 are 80.97–88.12%, the amino acid sequence identities of GlmU to the homolog in *E. coli* K12 are 73.94–79.8%, data not shown).

<https://doi.org/10.1371/journal.pone.0183267.g002>

clusters both contain *rmlA-qdtABC*. There are 7 *Proteus* O antigens whose structures include Rha (O25, O32, O51, O55, O58, O62, and O75), and all corresponding gene clusters except that of O55 contain *rmlACD*. Glc-1-P is converted to dTDP-6-deoxy- α -D-xylo-hexos-4-ulose by *rmlAB* [48, 49], which is further converted to dTDP-L-Rha by *rmlCD* [50, 51] or dTDP-D-Fuc3NAc by *fdtABC* [52] or dTDP-D-Qui3NAc by *qdtABC* [53]. However, we did not find the *rmlB* gene in any of these *Proteus* O antigen gene clusters, indicating that the *rmlB* gene may be located outside the clusters in these strains. Therefore, we downloaded 68 *Proteus* genomes (S5 Table), and identified the *RmlB* in all these genomes, which had amino acid sequence identities to the homolog in *E. coli* K12 from 71.95% to 74.76% (data not shown).

There are 4 *Proteus* O antigens whose structures include Ribf (O9, O25, O36, and O59), the only pentose, which is available from the NAD salvage pathway [54]. The gene responsible for the synthesis of UDP-Ribf is not always located in O antigen gene clusters [19, 21], and we did not find it in these strains too. There are 5 *Proteus* O antigens whose structures include Rib-ol (O16, O33, O41, O53, and O73ab), which is synthesized from ribulose 5-phosphate by *rib* [54]. However, we did not find the *rib* gene in any of these *Proteus* O antigen gene clusters, indicating that the *rib* gene may be located outside the clusters in these strains.

There are 39 *Proteus* O antigens whose structures do not contain any rare monosaccharides (O3ab, O5, O9-O11, O13-O14, O16, O18, O20-O24, O26-O30, O33-O34, O37ab, O40-O41, O44, O47-O50, O52, O54ab, O57, O60-O61, O65, O69, and O71-O74). Of these, there are 38 antigens whose O antigen gene clusters are relatively simple and do not contain any genes responsible for the synthesis of rare monosaccharides. The only exception is O37ab, which is discussed below.

There are also some other genes in the O antigen gene clusters. Most O antigens of *Proteus* contain uronic acids, and the *ugd* and *gla* genes were found in many *Proteus* O antigen gene clusters [3]. The *ugd* gene is involved in UDP-GlcA biosynthesis [55, 56], and it was present in the O antigen gene clusters of all 60 O antigens (O56 has two copies of *ugd* gene). The *gla* gene, which is involved in UDP-GalA biosynthesis [55, 56], was found in 15 of the *Proteus* O antigen gene clusters. Of these, 14 O antigens (except for O61) contain GalA. The *gne* gene, which is responsible for UDP-GalNAc synthesis, was found in 12 *Proteus* O antigen gene clusters, and all corresponding O antigens contain GalNAc [57]. The *glf* gene, which is involved in the synthesis of UDP-Galf, was found in 5 *Proteus* O antigen gene clusters, but none of these O antigens contain Galf, suggesting that *glf* may not be involved in *Proteus* O antigen biosynthesis [58].

There are 3 *Proteus* O antigens (5% of the collection) whose O antigen gene clusters did not correspond to their known O antigen structures (O37ab, O53, and O55). The O37ab gene cluster contains *rmlA-qdtAB*, but its O antigen does not contain Qui3NAc. O53 antigen contains FucNAc but no *fnlABC* genes were found in the O antigen gene cluster. Similarly, O55 contains Rha in the O antigen but no *rmlACD* genes in the O antigen gene cluster. We have rechecked these three strains by sequencing based on serotype-specific genes, and the possibility of mixing up of other strains can be excluded. The possible explanation is that the strains we used in sequencing are different from those used for structure analysis. We will identify the O antigen structures of the three strains we had, and sequence the O antigen gene clusters of other strains of these three serotypes in the future.

Glycosyltransferase genes

In *E. coli* and *Shigella*, the first sugar residue of the O antigen synthesis is GalNAc or GlcNAc, and the initial transferase (IT) encoded by *wecA* is responsible for initiating the O antigen synthesis, which is usually located outside the O antigen gene cluster [27, 59]. Almost all of the

Proteus OPS structures analyzed in this study (except for O53) contain GalNAc or GlcNAc [32]. The IT genes are usually conserved across different species, and we identified the WecA in all 68 *Proteus* genomes, which had amino acid sequence identities to the homolog in *E. coli* K12 from 73.3% to 76.74% (S6 Table). The identities between the *Proteus* WecA are 90.46%–100% (data not shown). Therefore, we propose that WecA initiates the synthesis of the OPS in most *Proteus* strains.

Glycosyltransferases sequentially transfer sugars to growing glycan chains until the O-units have been completely synthesized [19, 21]. Each studied *Proteus* O antigen gene cluster contained 2 to 7 putative GT genes, with a total of 216 GT genes identified in the 60 O antigen gene clusters. According to the similarity of the protein sequences, we have classified 78 of these GTs into 19 homology groups that contain at least 2 GTs (HG01–HG19), as shown in S7 Table. The GTs in the same HGs are considered to have similar functions. By comparing the structures of the different O antigens that contain GTs belonging to the same HG, we predicted the functions of some of these GTs, and some examples are discussed below.

For instance, the GTs of HG02 share 62–65% identity to *E. coli* WbuB, which is a known L-FucNAc transferase [60]. By comparing the O antigen structures whose corresponding gene clusters contain GTs belonging to HG02, we found that O6, O8, O12, O19, and O42 all contain an α -L-FucNAc-(1→3)-D-GlcNAc linkage in their structures. Therefore, we predicted that the GTs of HG02 have a similar function and are responsible for the formation of the α -L-FucNAc-(1→3)-D-GlcNAc linkage.

In the same manner, the GTs of HG15 share 55% identity to *Citrobacter europaeus* PglA, which is a known α -1,3-D-GalNAc transferase [35, 61]. By comparing the known O antigen structures whose corresponding gene clusters contain GTs belonging to HG15 (O16 and O48), we found that they contain an α -D-GalNAc-(1→3)-D-GlcNAc linkage. Therefore, we predicted that the GTs of HG15 have a similar function and are responsible for the formation of the α -D-GalNAc-(1→3)-D-GlcNAc linkage.

By comparing the known 6 *Proteus* O antigen structures whose corresponding gene clusters containing GTs belonging to HG05, we found that they contain only one common linkage, α -D-GalA-(1→3)-D-GlcNAc. Therefore, we suggest that GTs of HG05 are responsible for the formation of the α -D-GalA-(1→3)-D-GlcNAc linkage.

O antigen processing genes

All 60 studied *Proteus* O antigen gene clusters contained both *wzx* and *wzy* genes, but none of them contained *wzz* gene. We identified the Wzz in all 68 *Proteus* genomes, which had amino acid sequence identities from 86.16% to 100% to *E. coli* K12 Wzz (S8 Table). Therefore, the *wzz* gene is located outside the *Proteus* O antigen gene clusters. As expected, all Wzx contain 10 to 12 transmembrane segments, and all Wzy contain 9 to 12 transmembrane segments. We constructed the maximum likelihood phylogenetic trees using the *wzx* and *wzy* genes, individually, which show the high levels of diversity of these two genes from different strains (Fig 3). The maximum gene sequence identity of *wzx* is 87.9%, and the maximum gene sequence identity of *wzy* is 82.7%; the identities between either *wzx* or *wzy* are not more than 80%. The diversity of the *wzx* and *wzy* genes provided us a basis to develop molecular techniques to detect and identify different *Proteus* O serotypes.

Additional genes identified

A putative methyltransferase gene, a glycerol-3-phosphate dehydrogenase gene, and two serine acetyltransferase genes were found between *cpxA* and *secB* genes in all 60 serotypes. The methyltransferase gene shared the same transcriptional promoter on the leading strand with

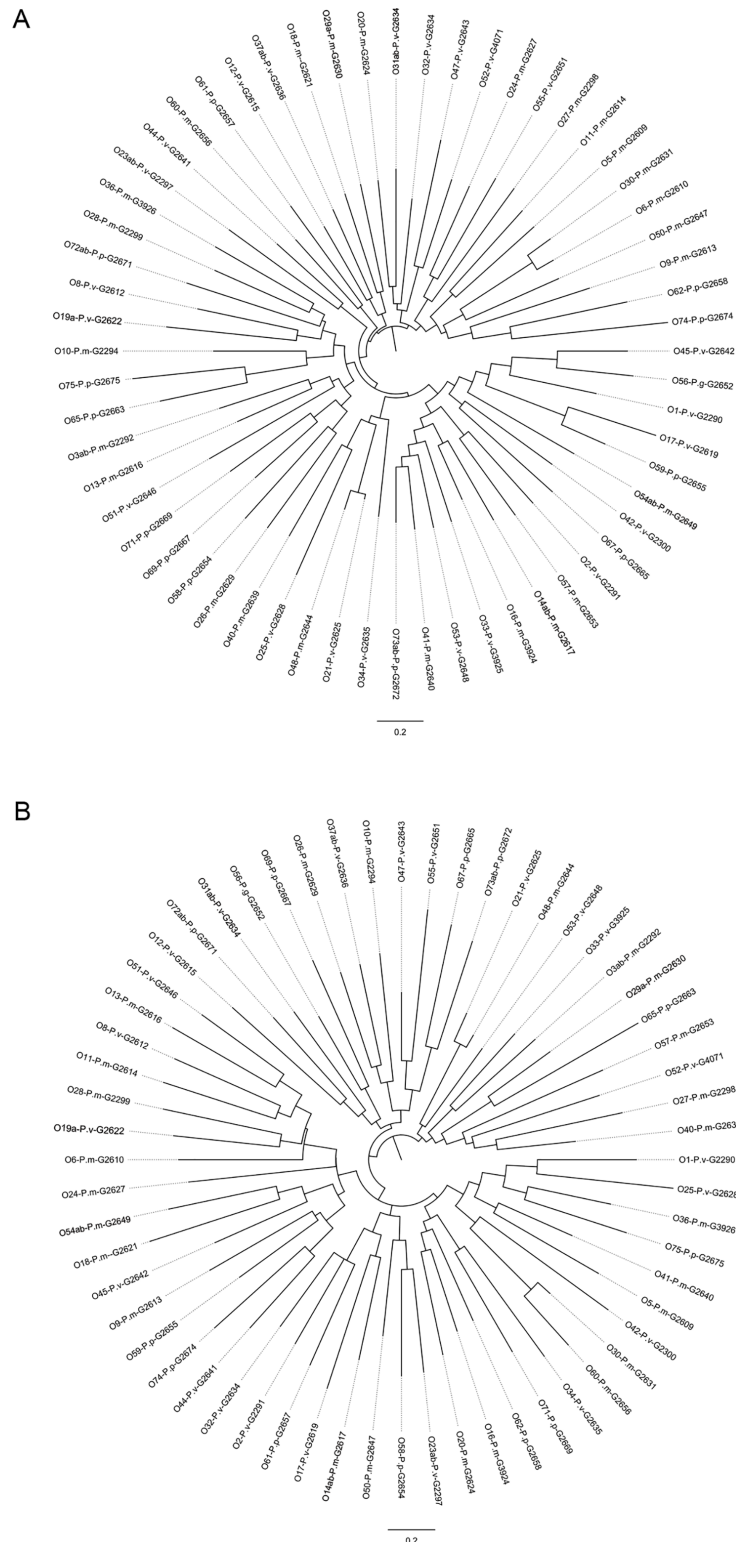


Fig 3. The phylogenetic trees for *wzx* and *wzy* genes from the 60 *Proteus* serotypes. The *wzx* (A) and *wzy* (B) trees were constructed using *wzx* and *wzy* genes. The sequences were aligned using ClustaIW v2.0, and the trees were constructed using the JC69 substitution model and the phyML v3.0.

<https://doi.org/10.1371/journal.pone.0183267.g003>

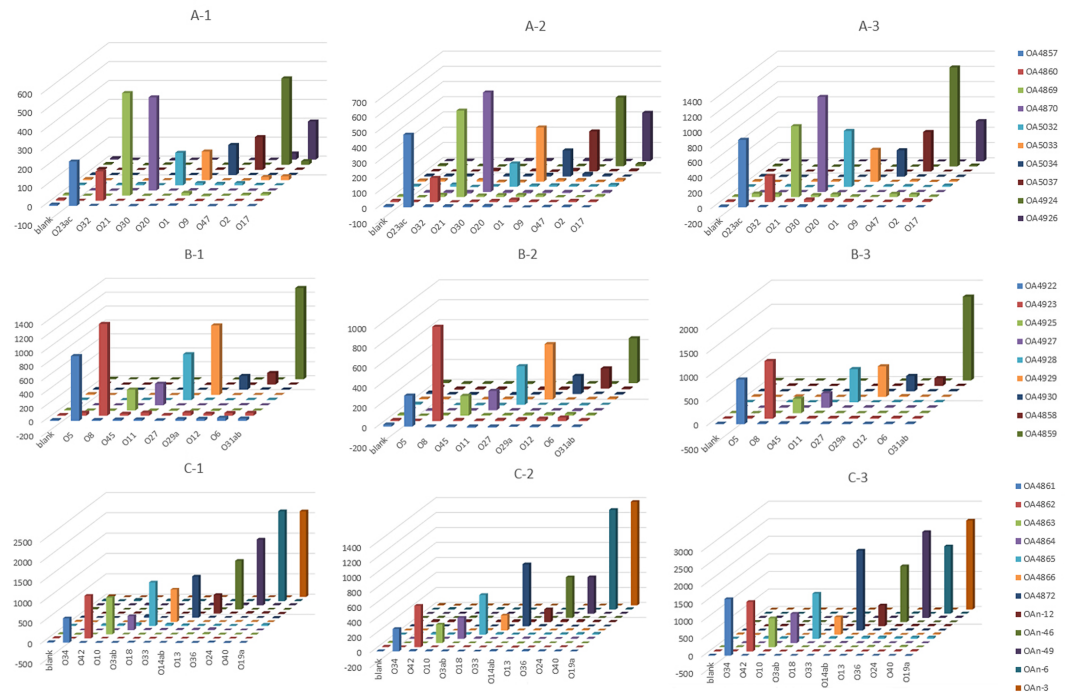


Fig 4. The hybridization results of the 31 *Proteus* strains. The suspension arrays were divided into 3 groups: (A) O1, O2, O9, O17, O20, O21, O23ac, O30, O32 and O47; (B) O5, O6, O8, O11, O12, O27, O29a, O31ab and O45; (C) O3ab, O10, O13, O14ab, O18, O19a, O24, O33, O34, O36, O40 and O42; no cross reactions were observed for any probe tested in this study, and the Blank was a negative control; the x-axis represents the PCR products of different serotypes, the y-axis represents the MFI values, and the z-axis represents the specific probes used for detection.

<https://doi.org/10.1371/journal.pone.0183267.g004>

the other O antigen synthesis genes. However, the other three genes use a different transcriptional promoter on the lagging strand, suggesting that these genes may not be associated with *Proteus* O antigen synthesis.

PCR-based suspension arrays for molecular detection of 31 different *Proteus* O serotypes

With the development of molecular techniques, many PCR-based molecular serotyping methods have been developed based on the O antigen specific genes for serological identification of many species, such as *E. coli*, *Salmonella* and *Yersinia pseudotuberculosis* [62]. According to the *Proteus* O antigen gene cluster analysis we performed in this study, the *wzx* and *wzy* genes were specific for different serotypes, indicating that the *wzx* and *wzy* genes could be used for molecular serotyping. At present, 37 *Proteus* serotypes have been reported to be frequently isolated from clinical specimens (O1-O15, O17-O21, O23-O24, O27-O34, O36, O38, O40, O42, and O45-O47) [35, 63, 64], and 31 of these 37 O antigens were analyzed in this study. A PCR-based suspension array was developed for molecular serotyping of all these 31 *Proteus* O serotypes using the *wzx* or *wzy* genes (Fig 4).

Primers were designed based on the *wzx* or *wzy* genes to amplify the PCR products, as described in the materials and methods. In most cases, we used *wzy* gene as target to amplify PCR products. If serotypes appeared to cross react due to the high sequence identities of their *wzy* genes, we tried to use *wzx* to obtain the PCR amplicons. Under optimal conditions, the multiplex PCR was performed to amplify the target amplicons varied from 100 to 495 bp.

Serotype-specific probes (19 to 30 bp) were designed based on target genes for each serotype, and the optimum hybridization temperatures were determined by detecting the hybridization efficiencies at different temperatures (from 45°C to 60°C). Consequently, the probe hybridization temperature was determined to be 56–59°C.

To distinguish all of the 31 different strains tested, suspension arrays were divided into 3 groups: (A) O1, O2, O9, O17, O20, O21, O23ac, O30, O32 and O47; (B) O5, O6, O8, O11, O12, O27, O29a, O31ab and O45; (C) O3ab, O10, O13, O14ab, O18, O19a, O24, O33, O34, O36, O40 and O42. The results of the suspension array can be repeated from three repeat detections. The MFIs for probes hybridized with their homologous DNA are > 151.5, and the S/Bs for probes hybridized with their homologous DNA are > 14.9. The MFIs for probes hybridized with their nonhomologous DNA are < 51, and S/Bs for probes hybridized with their nonhomologous DNA are < 5.0. The MFIs and S/Bs for probes hybridized with their homologous DNA were significantly higher than the MFIs and S/Bs obtained from probes hybridized with nonhomologous DNA. No cross reactions were observed for any probe tested in this study (Fig 4).

In order to determine the sensitivity of the suspension array, a ten-fold gradient dilution experiment was carried out (0 fg/μl, 1.0 fg/μl, 10.0 fg/μl, 100.0 fg/μl, 1.0 pg/μl, 10.0 pg/μl, 100.0 pg/μl, 1.0 ng/μl and 10.0 ng/μl of genomic DNA). Positive signals could be generated as low as 10–100 pg of genomic DNA.

In conclusion, the primers and probes designed in this study worked well for each strain, and no obvious nonspecific signals were observed. However, like any other molecular detection method, this suspension array has limitations because the probes must be designed based on known sequences. More serotypes can be distinguished using this method if new specific probes and primers are designed to complement our suspension assay. Overall, this *wzx/wzy*-based suspension array provides us a potential tool to identify different *Proteus* O serotypes.

Conclusions

OPS is an important component of gram-negative bacterial cell membranes with high variability within and between species. In this study, we identified 55 new O antigen gene clusters from different *Proteus* serotypes. Together with previously reported gene cluster data [35], we have analyzed a total of 60 *Proteus* O antigen gene clusters and have confirmed that the *Proteus* O antigen gene clusters are located between *cpxA* and *secB* genes, and the synthesis of *Proteus* O antigen is *Wzx/Wzy* pathway dependent. By comparison with their known O antigen structures, we found that most O antigen gene clusters correlated well with the corresponding O antigen structures (57 of 60, 95%). We also predicted the functions of some of the GTs by comparing the known O antigen structures whose corresponding gene clusters contain GTs belonging to the same HGs. The diversity of the *wzx* and *wzy* genes provides a basis for rapid molecular detection of different *Proteus* O serotypes. We therefore developed a suspension array to distinguish 31 different *Proteus* O serotypes using specific primers and probes designed based on the *wzx/wzy* genes. Our work comprehensively describes the O antigen gene clusters of *Proteus* species and provides a basis for future serological studies.

Supporting information

S1 Table. The strains used in this study (coming from the University of Lodz). ATCC—American Type Culture Collection, USA;
CCUG—Cultures Collection, University of Goeteborg, Sweden;
CDC—Center for Disease Control and Prevention, Atlanta, USA;
CNCTC—Czech National Collection of Type Cultures, Prague, Czech Republic;

NCTC–National Collection of Type Culture, London, UK.
(DOC)

S2 Table. The primers used in this study.
(DOC)

S3 Table. The probe used in this study.
(DOC)

S4 Table. The O antigen structures from the 60 *Proteus* serotypes.
(DOC)

S5 Table. The 68 *Proteus* genomes.
(DOC)

S6 Table. The WecA in the 68 *Proteus* genomes. The amino acid sequence identities were obtained in comparison with the *E. coli* K12 WecA.
(DOC)

S7 Table. The GTs and HGs.
(DOC)

S8 Table. The Wzz in the 68 *Proteus* genomes. The amino acid sequence identities were obtained in comparison with the *E. coli* K12 Wzz.
(DOC)

Acknowledgments

The authors thank the Department of Immunobiology of Bacteria and the Department of General Microbiology at the Institute of Microbiology, Biotechnology and Immunology, Faculty of Biology and Environmental Protection, University of Lodz for their help to obtain and handle the strains. We thank Dr. Lingyan Jiang and Dr. Min Wang for helpful suggestions.

Author Contributions

Conceptualization: Xiang Yu, Bin Liu, Lei Wang.

Data curation: Xiang Yu, Zhiqiu Yin, Hengchun Cao.

Formal analysis: Xiang Yu, Zhiqiu Yin, Hengchun Cao.

Funding acquisition: Bin Liu, Lei Wang.

Project administration: Lei Wang.

Validation: Agnieszka Torzewska, Xinjie Zhang, Dominika Drzewiecka, Yuriy A. Knirel, Antoni Rozalski.

Writing – original draft: Xiang Yu, Yuriy A. Knirel, Antoni Rozalski, Lei Wang.

Writing – review & editing: Xiang Yu, Lei Wang.

References

1. Drzewiecka D. Significance and Roles of *Proteus* spp. Bacteria in Natural Environments. *Microb Ecol.* 2016; 72(4):741–58. <https://doi.org/10.1007/s00248-015-0720-6> PMID: 26748500; PubMed Central PMCID: PMC5080321.
2. Kim BN, Kim NJ, Kim MN, Kim YS, Woo JH, Ryu J. Bacteraemia due to tribe Proteeae: a review of 132 cases during a decade (1991–2000). *Scand J Infect Dis.* 2003; 35(2):98–103. PMID: 12693558.

3. Rozalski A, Sidorczyk Z, Kotelko K. Potential virulence factors of *Proteus bacilli*. *Microbiol Mol Biol Rev*. 1997; 61(1):65–89. PMID: [9106365](#); PubMed Central PMCID: [PMCPMC232601](#).
4. Schaffer JN, Pearson MM. *Proteus mirabilis* and Urinary Tract Infections. *Microbiol Spectr*. 2015; 3(5). <https://doi.org/10.1128/microbiolspec.UTI-0017-2013> PMID: [26542036](#); PubMed Central PMCID: [PMCPMC4638163](#).
5. Wilson C, Thakore A, Isenberg D, Ebringer A. Correlation between anti-*Proteus* antibodies and isolation rates of *P. mirabilis* in rheumatoid arthritis. *Rheumatol Int*. 1997; 16(5):187–9. PMID: [9032817](#).
6. O'Hara CM, Brenner FW, Miller JM. Classification, identification, and clinical significance of *Proteus*, *Providencia*, and *Morganella*. *Clin Microbiol Rev*. 2000; 13(4):534–46. PMID: [11023955](#); PubMed Central PMCID: [PMCPMC88947](#).
7. O'Hara CM, Brenner FW, Steigerwalt AG, Hill BC, Holmes B, Grimont PA, et al. Classification of *Proteus vulgaris* biogroup 3 with recognition of *Proteus hauseri* sp. nov., nom. rev. and unnamed *Proteus* genomospecies 4, 5 and 6. *Int J Syst Evol Microbiol*. 2000; 50 Pt 5:1869–75. <https://doi.org/10.1099/00207713-50-5-1869> PMID: [11034498](#).
8. Behrendt U, Augustin J, Sproer C, Gelbrecht J, Schumann P, Ulrich A. Taxonomic characterisation of *Proteus terrae* sp. nov., a N₂O-producing, nitrate-ammonifying soil bacterium. *Antonie Van Leeuwenhoek*. 2015; 108(6):1457–68. <https://doi.org/10.1007/s10482-015-0601-5> PMID: [26437638](#).
9. Hyun DW, Jung MJ, Kim MS, Shin NR, Kim PS, Whon TW, et al. *Proteus cibarius* sp. nov., a swarming bacterium from Jeotgal, a traditional Korean fermented seafood, and emended description of the genus *Proteus*. *Int J Syst Evol Microbiol*. 2016; 66(6):2158–64. <https://doi.org/10.1099/ijsem.0.001002> PMID: [26944634](#).
10. Jacobsen SM, Stickler DJ, Mobley HL, Shirliff ME. Complicated catheter-associated urinary tract infections due to *Escherichia coli* and *Proteus mirabilis*. *Clin Microbiol Rev*. 2008; 21(1):26–59. <https://doi.org/10.1128/CMR.00019-07> PMID: [18202436](#); PubMed Central PMCID: [PMCPMC2223845](#).
11. Jacobsen SM, Lane MC, Harro JM, Shirliff ME, Mobley HL. The high-affinity phosphate transporter Pst is a virulence factor for *Proteus mirabilis* during complicated urinary tract infection. *FEMS Immunol Med Microbiol*. 2008; 52(2):180–93. <https://doi.org/10.1111/j.1574-695X.2007.00358.x> PMID: [18194341](#).
12. Mobley HL, Belas R. Swarming and pathogenicity of *Proteus mirabilis* in the urinary tract. *Trends Microbiol*. 1995; 3(7):280–4. PMID: [7551643](#).
13. Armbruster CE, Mobley HL. Merging mythology and morphology: the multifaceted lifestyle of *Proteus mirabilis*. *Nat Rev Microbiol*. 2012; 10(11):743–54. <https://doi.org/10.1038/nrmicro2890> PMID: [23042564](#); PubMed Central PMCID: [PMCPMC3621030](#).
14. Raetz CR, Whitfield C. Lipopolysaccharide endotoxins. *Annu Rev Biochem*. 2002; 71:635–700. <https://doi.org/10.1146/annurev.biochem.71.110601.135414> PMID: [12045108](#); PubMed Central PMCID: [PMCPMC2569852](#).
15. Beynon LM, Dumanski AJ, McLean RJ, MacLean LL, Richards JC, Perry MB. Capsule structure of *Proteus mirabilis* (ATCC 49565). *J Bacteriol*. 1992; 174(7):2172–7. PMID: [1551839](#); PubMed Central PMCID: [PMCPMC205835](#).
16. Perry MB, MacLean LL. The structure of the polysaccharide produced by *Proteus vulgaris* (ATCC 49990). *Carbohydr Res*. 1994; 253:257–63. PMID: [8156552](#).
17. Torzewska A, Staczek P, Rozalski A. Crystallization of urine mineral components may depend on the chemical nature of *Proteus* endotoxin polysaccharides. *J Med Microbiol*. 2003; 52(Pt 6):471–7. <https://doi.org/10.1099/jmm.0.05161-0> PMID: [12748265](#).
18. Kaca W, Arabski M, Fudala R, Holmstrom E, Sjolholm A, Weintraub A, et al. Human complement activation by smooth and rough *Proteus mirabilis* lipopolysaccharides. *Arch Immunol Ther Exp (Warsz)*. 2009; 57(5):383–91. <https://doi.org/10.1007/s00005-009-0043-8> PMID: [19707721](#).
19. Liu B, Knirel YA, Feng L, Perepelov AV, Senchenkova SN, Reeves PR, et al. Structural diversity in *Salmonella* O antigens and its genetic basis. *FEMS Microbiol Rev*. 2014; 38(1):56–89. <https://doi.org/10.1111/1574-6976.12034> PMID: [23848592](#).
20. Hu D, Liu B, Dijkshoorn L, Wang L, Reeves PR. Diversity in the major polysaccharide antigen of *Acinetobacter baumannii* assessed by DNA sequencing, and development of a molecular serotyping scheme. *PLoS One*. 2013; 8(7):e70329. <https://doi.org/10.1371/journal.pone.0070329> PMID: [23922982](#); PubMed Central PMCID: [PMCPMC3726653](#).
21. Liu B, Knirel YA, Feng L, Perepelov AV, Senchenkova SN, Wang Q, et al. Structure and genetics of *Shigella* O antigens. *FEMS Microbiol Rev*. 2008; 32(4):627–53. <https://doi.org/10.1111/j.1574-6976.2008.00114.x> PMID: [18422615](#).
22. Knirel' lu A, Kochetkov NK. Structure of lipopolysaccharides from gram-negative bacteria. III. Structure of O-specific polysaccharides. *Biokhimiia*. 1994; 59(12):1784–851. PMID: [7533007](#).

23. Momtaz H, Karimian A, Madani M, Safarpour Dehkordi F, Ranjbar R, Sarshar M, et al. Uropathogenic *Escherichia coli* in Iran: serogroup distributions, virulence factors and antimicrobial resistance properties. *Ann Clin Microbiol Antimicrob*. 2013; 12:8. <https://doi.org/10.1186/1476-0711-12-8> PMID: 23627669; PubMed Central PMCID: PMCPMC3651382.
24. Sakazaki R. Studies on the *Hafnia* group of Enterobacteriaceae. *Jpn J Med Sci Biol*. 1961; 14:223–41. PMID: 14496196.
25. Reeves PP, Wang L. Genomic organization of LPS-specific loci. *Curr Top Microbiol Immunol*. 2002; 264(1):109–35. PMID: 12014174.
26. Kenyon JJ, Reeves PR. The Wzy O-antigen polymerase of *Yersinia pseudotuberculosis* O:2a has a dependence on the Wzz chain-length determinant for efficient polymerization. *FEMS Microbiol Lett*. 2013; 349(2):163–70. <https://doi.org/10.1111/1574-6968.12311> PMID: 24164168.
27. Cunneen MM, Liu B, Wang L, Reeves PR. Biosynthesis of UDP-GlcNAc, UndPP-GlcNAc and UDP-GlcNAcA involves three easily distinguished 4-epimerase enzymes, Gne, Gnu and GnaB. *PLoS One*. 2013; 8(6):e67646. <https://doi.org/10.1371/journal.pone.0067646> PMID: 23799153; PubMed Central PMCID: PMCPMC3682973.
28. Beutin L, Strauch E, Zimmermann S, Kaulfuss S, Schaudinn C, Mannel A, et al. Genetical and functional investigation of fliC genes encoding flagellar serotype H4 in wildtype strains of *Escherichia coli* and in a laboratory *E. coli* K-12 strain expressing flagellar antigen type H48. *BMC Microbiol*. 2005; 5:4. <https://doi.org/10.1186/1471-2180-5-4> PMID: 15663798; PubMed Central PMCID: PMCPMC548302.
29. Mortimer CK, Peters TM, Gharbia SE, Logan JM, Arnold C. Towards the development of a DNA-sequence based approach to serotyping of *Salmonella enterica*. *BMC Microbiol*. 2004; 4:31. <https://doi.org/10.1186/1471-2180-4-31> PMID: 15298703; PubMed Central PMCID: PMCPMC514894.
30. Zarnowiec P, Mizera A, Chrapek M, Urbaniak M, Kaca W. Chemometric analysis of attenuated total reflectance infrared spectra of *Proteus mirabilis* strains with defined structures of LPS. *Innate Immun*. 2016; 22(5):325–35. <https://doi.org/10.1177/1753425916647470> PMID: 27189426.
31. Chen YS, Li HR, Zhang W, Hua ZD, Lin XH, Lin MQ, et al. Development of a bead-based suspension array for the detection of pathogens in acute respiratory tract infections. *Exp Biol Med* (Maywood). 2016; 241(14):1551–8. <https://doi.org/10.1177/1535370216647128> PMID: 27190247; PubMed Central PMCID: PMCPMC4994910.
32. Knirel YA, Perepelov AV, Kondakova AN, Senchenkova SN, Sidorczyk Z, Rozalski A, et al. Structure and serology of O-antigens as the basis for classification of *Proteus* strains. *Innate Immun*. 2011; 17(1):70–96. <https://doi.org/10.1177/1753425909360668> PMID: 20305038.
33. Arbatsky NP, Drzewiecka D, Palusiak A, Shashkov AS, Zablotni A, Siwinska M, et al. Structure of a Kdo-containing O polysaccharide representing *Proteus* O79, a newly described serogroup for some clinical *Proteus* genomospecies isolates from Poland. *Carbohydr Res*. 2013; 379:100–5. <https://doi.org/10.1016/j.carres.2013.07.001> PMID: 23896084.
34. Siwinska M, Levina EA, Ovchinnikova OG, Drzewiecka D, Shashkov AS, Rozalski A, et al. Classification of a *Proteus penneri* clinical isolate with a unique O-antigen structure to a new *Proteus* serogroup, O80. *Carbohydr Res*. 2015; 407:131–6. <https://doi.org/10.1016/j.carres.2015.02.003> PMID: 25771295.
35. Wang Q, Torzewska A, Ruan X, Wang X, Rozalski A, Shao Z, et al. Molecular and genetic analyses of the putative *Proteus* O antigen gene locus. *Appl Environ Microbiol*. 2010; 76(16):5471–8. <https://doi.org/10.1128/AEM.02946-09> PMID: 20581173; PubMed Central PMCID: PMCPMC2918944.
36. Bastin DA, Reeves PR. Sequence and analysis of the O antigen gene (rfb) cluster of *Escherichia coli* O111. *Gene*. 1995; 164(1):17–23. PMID: 7590310.
37. Wang L, Reeves PR. Organization of *Escherichia coli* O157 O antigen gene cluster and identification of its specific genes. *Infect Immun*. 1998; 66(8):3545–51. PMID: 9673232; PubMed Central PMCID: PMCPMC108385.
38. Staden R. The Staden sequence analysis package. *Mol Biotechnol*. 1996; 5(3):233–41. PMID: 8837029.
39. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. *Bioinformatics*. 2000; 16(10):944–5. PMID: 11120685.
40. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997; 25(17):3389–402. PMID: 9254694; PubMed Central PMCID: PMCPMC146917.
41. Linard B, Thompson JD, Poch O, Lecompte O. OrthoInspector: comprehensive orthology analysis and visual exploration. *BMC Bioinformatics*. 2011; 12:11. <https://doi.org/10.1186/1471-2105-12-11> PMID: 21219603; PubMed Central PMCID: PMCPMC3024942.
42. Gao Y, MacDonald D, Collins KD, Alaghebandan R, Chen Y. Role of social support in the relationship between sexually transmitted infection and depression among young women in Canada. *J Epidemiol*.

- 2010; 20(4):313–8. <https://doi.org/10.2188/jea.JE20090133> PMID: 20551580; PubMed Central PMCID: PMCPMC3900792.
43. Cao B, Yao F, Liu X, Feng L, Wang L. Development of a DNA microarray method for detection and identification of all 15 distinct O-antigen forms of *Legionella pneumophila*. *Appl Environ Microbiol*. 2013; 79(21):6647–54. <https://doi.org/10.1128/AEM.01957-13> PMID: 23974134; PubMed Central PMCID: PMCPMC3811488.
 44. Pearson MM, Sebahia M, Churcher C, Quail MA, Seshasayee AS, Luscombe NM, et al. Complete genome sequence of uropathogenic *Proteus mirabilis*, a master of both adherence and motility. *J Bacteriol*. 2008; 190(11):4027–37. <https://doi.org/10.1128/JB.01981-07> PMID: 18375554; PubMed Central PMCID: PMCPMC2395036.
 45. Kneidinger B, Larocque S, Brisson JR, Cadotte N, Lam JS. 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(Pt 3):989–95. <https://doi.org/10.1042/BJ20030099> PMID: 12575896; PubMed Central PMCID: PMCPMC1223350.
 46. Mulrooney EF, Poon KK, McNally DJ, Brisson JR, Lam JS. Biosynthesis of UDP-N-acetyl-L-fucosamine, a precursor to the biosynthesis of lipopolysaccharide in *Pseudomonas aeruginosa* serotype O11. *J Biol Chem*. 2005; 280(20):19535–42. <https://doi.org/10.1074/jbc.M500612200> PMID: 15778500.
 47. Kneidinger B, O’Riordan K, Li J, Brisson JR, Lee JC, Lam JS. Three highly conserved proteins catalyze the conversion of UDP-N-acetyl-D-glucosamine to precursors for the biosynthesis of O antigen in *Pseudomonas aeruginosa* O11 and capsule in *Staphylococcus aureus* type 5. Implications for the UDP-N-acetyl-L-fucosamine biosynthetic pathway. *J Biol Chem*. 2003; 278(6):3615–27. <https://doi.org/10.1074/jbc.M203867200> PMID: 12464616.
 48. Zuccotti S, Zanardi D, Rosano C, Sturla L, Tonetti M, Bolognesi M. Kinetic and crystallographic analyses support a sequential-ordered bi bi catalytic mechanism for *Escherichia coli* glucose-1-phosphate thymidyltransferase. *J Mol Biol*. 2001; 313(4):831–43. <https://doi.org/10.1006/jmbi.2001.5073> PMID: 11697907.
 49. Allard ST, Giraud MF, Whitfield C, Graninger M, Messner P, Naismith JH. 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(1):283–95. <https://doi.org/10.1006/jmbi.2000.4470> PMID: 11243820.
 50. Giraud MF, Gordon FM, Whitfield C, Messner P, McMahon SA, Naismith JH. 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*. 1999; 55(Pt 3):706–8. PMID: 10089454.
 51. Giraud MF, McMiken HJ, Leonard GA, Messner P, Whitfield C, Naismith JH. 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*. 1999; 55(Pt 12):2043–6. PMID: 10666586.
 52. Pfoestl A, Hofinger A, Kosma P, Messner P. Biosynthesis of dTDP-3-acetamido-3,6-dideoxy-alpha-D-galactose in *Aneurinibacillus thermoaerophilus* L420-91T. *J Biol Chem*. 2003; 278(29):26410–7. <https://doi.org/10.1074/jbc.M300858200> PMID: 12740380.
 53. Pfoestl A, Zayni S, Hofinger A, Kosma P, Schaffer C, Messner P. Biosynthesis of dTDP-3-acetamido-3,6-dideoxy-alpha-D-glucose. *Biochem J*. 2008; 410(1):187–94. <https://doi.org/10.1042/BJ20071044> PMID: 17941826; PubMed Central PMCID: PMCPMC4379490.
 54. Follens A, Veiga-da-Cunha M, Merckx R, van Schaftingen E, van Eldere J. *acs1* of *Haemophilus influenzae* type a capsulation locus region II encodes a bifunctional ribulose 5-phosphate reductase-CDP-ribitol pyrophosphorylase. *J Bacteriol*. 1999; 181(7):2001–7. PMID: 10094675; PubMed Central PMCID: PMCPMC93610.
 55. Usadel B, Schluter U, Molhoj M, Gipmans M, Verma R, Kossmann J, et al. Identification and characterization of a UDP-D-glucuronate 4-epimerase in *Arabidopsis*. *FEBS Lett*. 2004; 569(1–3):327–31. <https://doi.org/10.1016/j.febslet.2004.06.005> PMID: 15225656.
 56. Vigetti D, Ori M, Viola M, Genasetti A, Karousou E, Rizzi M, et al. Molecular cloning and characterization of UDP-glucose dehydrogenase from the amphibian *Xenopus laevis* and its involvement in hyaluronan synthesis. *J Biol Chem*. 2006; 281(12):8254–63. <https://doi.org/10.1074/jbc.M508516200> PMID: 16418163.
 57. Bengoechea JA, Pinta E, Salminen T, Oertelt C, Holst O, Radziejewska-Lebrecht J, 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(15):4277–87. <https://doi.org/10.1128/JB.184.15.4277-4287.2002> PMID: 12107146; PubMed Central PMCID: PMCPMC135196.

58. Nassau PM, Martin SL, Brown RE, Weston A, Monsey D, McNeil MR, et al. Galactofuranose biosynthesis in *Escherichia coli* K-12: identification and cloning of UDP-galactopyranose mutase. *J Bacteriol.* 1996; 178(4):1047–52. PMID: [8576037](#); PubMed Central PMCID: PMCPMC177764.
59. Lehrer J, Vigeant KA, Tatar LD, Valvano MA. 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(7):2618–28. <https://doi.org/10.1128/JB.01905-06> PMID: [17237164](#); PubMed Central PMCID: PMCPMC1855806.
60. Wang Q, Wang S, Beutin L, Cao B, Feng L, Wang L. Development of a DNA microarray for detection and serotyping of enterotoxigenic *Escherichia coli*. *J Clin Microbiol.* 2010; 48(6):2066–74. <https://doi.org/10.1128/JCM.02014-09> PMID: [20351209](#); PubMed Central PMCID: PMCPMC2884529.
61. Holden KM, Gilbert M, Coloe PJ, Li J, Fry BN. The role of WlaRG, WlaTB and WlaTC in lipooligosaccharide synthesis by *Campylobacter jejuni* strain 81 116. *Microb Pathog.* 2012; 52(6):344–52. <https://doi.org/10.1016/j.micpath.2012.03.004> PMID: [22445818](#).
62. Bogdanovich T, Carniel E, Fukushima H, Skurnik M. Use of O-antigen gene cluster-specific PCRs for the identification and O-genotyping of *Yersinia pseudotuberculosis* and *Yersinia pestis*. *J Clin Microbiol.* 2003; 41(11):5103–12. <https://doi.org/10.1128/JCM.41.11.5103-5112.2003> PMID: [14605146](#); PubMed Central PMCID: PMCPMC262526.
63. Penner JL, Hennessy JN. Separate O-grouping schemes for serotyping clinical isolates of *Proteus vulgaris* and *Proteus mirabilis*. *J Clin Microbiol.* 1980; 12(3):304–9. PMID: [7012166](#); PubMed Central PMCID: PMCPMC273580.
64. Kaca W, Glenska J, Lechowicz L, Grabowski S, Brauner A, Kwinkowski M. Serotyping of *Proteus mirabilis* clinical strains based on lipopolysaccharide O-polysaccharide and core oligosaccharide structures. *Biochemistry (Mosc).* 2011; 76(7):851–61. <https://doi.org/10.1134/S0006297911070169> PMID: [21999547](#).