Carbohydrate Research 528 (2023) 108809

Contents lists available at ScienceDirect

Carbohydrate Research

journal homepage: www.elsevier.com/locate/carres

Structure and gene cluster annotation of the O-antigen of *Aeromonas sobria* strain K928 isolated from common carp and classified into the new *Aeromonas* PGO1 serogroup

Maria Kurzylewska [a], Arkadiusz Bomba [b], Katarzyna Dworaczek [a], Agnieszka Pękala-Safińska [c], Anna Turska-Szewczuk [a,*]

[a] *Department of Genetics and Microbiology, Institute of Biological Sciences, M. Curie-Sklodowska University, Akademicka 19, 20-033, Lublin, Poland*
[b] *Department of Omics Analyses, National Veterinary Research Institute, Partyzantow 57, 24-100, Pulawy, Poland*
[c] *Department of Preclinical Sciences and Infectious Diseases, Faculty of Veterinary Medicine and Animal Science, Poznan University of Life Sciences, Wolynska 35, 60-637, Poznan, Poland*

ARTICLE INFO

Keywords:
Aeromonas
O-antigen
O-antigen gene cluster
Wzx/Wzy pathway
Fuc*p*3N
NMR spectroscopy

ABSTRACT

Aeromonas sobria strain K928 was isolated from a common carp during a Motile Aeromonas Infection/Motile Aeromonas Septicaemia disease outbreak on a Polish fish farm and classified into the new provisional PGO1 serogroup. The lipopolysaccharide of *A. sobria* K928 was subjected to mild acid hydrolysis, and the O-specific polysaccharide, which was isolated by gel-permeation chromatography, was studied using sugar and methylation analyses and ^{1}H and ^{13}C NMR spectroscopy. The following structure of the branched O-specific polysaccharide repeating unit of *A. sobria* K928 was established.

→2)[α-D-Fuc*p*3N*R*Hb-(1→3)]-α-L-Rha*p*-(1→3)-β-L-Rha*p*-(1→4)-α-L-Rha*p*-(1→3)-β-D-Fuc*p*NAc-(1→

The O-antigen gene cluster was identified and characterized in the genome of the *A. sobria* K928 strain after comparison with sequences in the available databases. The composition of the O-antigen genetic region was found to be consistent with the O-polysaccharide structure, and its organization was proposed.

1. Introduction

The aquatic environment is the primary natural habitat of *Aeromonas* bacteria, but they are also able to colonize terrestrial surroundings. *A. salmonicida* and *A. hydrophila*, which are responsible for furunculosis, ulcers, hemorrhages, and septicaemias, are the main pathogens of salmonid fish [1]. In turn, mesophilic species, e.g., *A. encheleia, A. bestriarum*, and *A. sobria*, are the etiological agents of Motile Aeromonas Infection (MAI) [2] and Motile Aeromonas Septicaemia (MAS). The latter is a systemic disease affecting common carp, rainbow trout, channel catfish, eel, and tilapia, causing outbreaks with a high mortality rate on fish farms [3–5]. *Aeromonas* spp. are also human pathogens, with representatives of *A. hydrophila, A. veronii, A. caviae* [6], and A. dhakensis [3] considered to be the most important clinical species. They cause diseases of the digestive system in healthy subjects and can lead to severe diseases, including sepsis, in immunocompromised patients [6].

Aeromonas virulence factors that contribute to the high pathogenic potential of these bacteria include outer membrane proteins (OMP), capsular polysaccharide (CPS), lipopolysaccharide (LPS), S-layer, fimbriae, iron-binding system, secretion systems, and exotoxins [7,8]. LPS is one of the most widely studied bacterial outer membrane glycolipid, relevant for its virulence [9,10]. This negatively charged tripartite molecule of the outer membrane (OM) consists of lipid A, an oligosaccharide core, and the O-antigen (O-specific polysaccharide, OPS). Three types of LPS can be distinguished: the complete molecule, i.e., smooth LPS (S-LPS), rough LPS (R-LPS) composed of the core oligosaccharide (OS) and lipid A, which is the toxicity centre of the molecule triggering an inflammatory response in mammals, and semi-rough LPS, i.e., a variant in which the core OS is substituted with one O-antigen unit [11]. The O-antigen is the outermost part of LPS, determining the serological specificity of bacteria [12], the great diversity of which is caused by the type and sequence of sugar residues, the type of glycosidic linkages, and the occurrence of non-carbohydrate substituents [13].

The National Institute of Health NIH (Japan) serotyping system of Sakazaki and Shimada, which distinguishes 44 serogroups, is the traditional classification scheme of *Aeromonas* based on the O-antigen

* Corresponding author.
E-mail address: anna.turska-szewczuk@mail.umcs.pl (A. Turska-Szewczuk).

https://doi.org/10.1016/j.carres.2023.108809
Received 10 February 2023; Received in revised form 5 April 2023; Accepted 5 April 2023
Available online 11 April 2023

structure [14], but this pool is insufficient to assign all *Aeromonas* strains, and thus new provisional serogroups should be added to the scheme [12,15]. In recent years, infections of trout and carp in Poland have been caused mainly by *Aeromonas* belonging to the classical (O3, O6, O11, O16, O18, O33, O34) and provisional serogroups (PGO1, PGO2, PGO4, PGO6) [16–19].

Molecular serotyping based on the analysis of gene sequence diversity in the O-antigen gene cluster (OGC) seems to be an alternative method in comparison to the traditional serotyping test [20]. The O-antigen is synthesized by a unique set of genes grouped together [20,21], and genetic variation in the region results in different structural forms of the O-antigen [22]. There are three main groups of genes involved in O-antigen synthesis: 1) nucleotide sugar precursor genes for sugars specific to a given polysaccharide, 2) glycosyltransferase genes, which add sugars to an increasing O-unit, and 3) processing genes, including those that encode protein flippases and polymerases. Three pathways are responsible for the synthesis and transport of the O-antigen: the Wzx/Wzy-dependent pathway typical of heteropolymers [13,23,24], the Wzm/Wzt-dependent ABC transporter pathway typical of homopolymers [25–27], and the synthase-dependent secretion system [28]. The Wzx/Wzy-pathway is the most popular scheme for the O-antigen synthesis [23,24,29], which is found in, e.g., *E. coli* [23], *Shigella* spp. [30], *Salmonella* spp. [31], *Hafnia* spp. [32], *Proteus* spp. [24], and *Plesiomonas shigelloides* [25]. An analysis of the *A. hydrophila* OGC regions studied so far confirmed the presence of the Wzx/Wzy-pathway in serogroups O9, O16, O19, O24, O25, O29, O30, and O44 [20]. In *A. hydrophila*, the OGC ranges from 15,000 to 50,000 bp and comprises approximately 15–38 open reading frames (*orfs*) [20]. The region is usually mapped between housekeeping genes *acrB* and *oprM*, in comparison to enterobacterial OGCs, where *galF* and *gnd* flanked genes have been commonly found [23,30,31]. The *acrB* and *oprM* genes encode the A subunit of the multidrug pump and the outer membrane protein, respectively [20,33]; however, they do not always surround the OGC region of *Aeromonas*. In addition to *acrB,* both *wbpO* (UDP-glucose 6-dehydrogenase) and *rmlA* have also been found in the upstream position of the region, whereas *oprM* and *waaL* have been identified in the downstream position [20].

In this work, we report the structure of the O-specific polysaccharide of *A. sobria* strain K928, i.e., a representative of the PGO1 serogroup prevailing in Polish aquaculture, and characterize the O-antigen gene cluster determining its synthesis.

2. Experimental

2.1. Bacterial strain and cultivation

A. sobria strain K928 was isolated from common carp tissues during an outbreak of motile aeromonad septicaemia/motile aeromonad infection (MAS/MAI) on a Polish fish farm, and classified into the PGO1 serogroup after agglutination tests with sets of *Aeromonas* O-serogroup antisera of the NIH scheme [14] and additional antisera developed for selected Polish isolates representing 20 new provisional serogroups (PGO1—PGO20) [16,17].

For the LPS analysis, *A. sobria* strain K928 was cultivated in an aerated tryptic soy broth (TSB) medium for 72 h at 28 °C, and the biomass was harvested by centrifugation (8000×*g*, 35 min) followed by washing with 0.85% saline and with distilled water.

2.2. Isolation and degradation of the lipopolysaccharide

Cell biomass of *A. sobria* K928 was suspended in 50 mM phosphate buffer (pH 7.0) containing 5 mM $MgCl_2$ and treated with lysozyme, nucleases, and proteinase K as previously described [18], and then subjected to LPS isolation with the classical hot phenol/water method [34]. After extraction, the water and phenol layers were separated by centrifugation (3000×*g*, 1 h) and dialyzed against tap water for 3–5 days until the smell of phenol disappeared; this was followed by dialysis overnight against distilled water. The high molecular weight S-LPS species were recovered from the phenol phase after ultracentrifugation (105,000×*g*, 3 h) and freeze-drying. The yield of LPS was ~2.5%.

An LPS sample (100 mg) was hydrolyzed with aq 2% HOAc at 100 °C for 3.5 h, the lipid precipitate was removed by centrifugation (13,000×*g*, 20 min), and the OPS was obtained by fractionation of the supernatant by gel-permeation chromatography (GPC) on a Sephadex G-50 fine column (1.8 × 80 cm) in 1% acetic acid and monitored with a differential refractometer (Knauer, Germany). The high-molecular-weight O-polysaccharide fraction was obtained in a yield 11% of the LPS mass subjected to hydrolysis.

2.3. OPS composition analysis

For neutral and amino sugar analysis, an OPS sample was hydrolyzed with 2 M CF_3CO_2H (100 °C, 4 h), reduced with $NaBD_4$, and peracetylated with a 1:1 (v/v) Ac_2O-pyridine mixture (85 °C, 0.5 h). Monosaccharides were identified by GLC-MS of alditol acetates on an Agilent Technologies 7890 A gas chromatograph (USA) connected to a 5975C MSD detector (inert XL EI/CI, Agilent Technologies, USA). The chromatograph was equipped with an HP-5MS capillary column (Agilent Technologies, 30 m × 0.25 mm, flow rate of 1 ml min^{-1}, He as the carrier gas). The temperature program for all the derivatives was as follows: 150 °C for 5 min, then 150–310 °C at 5 °C min^{-1}, and the final temperature was maintained for 10 min.

Methylation of the OPS (1.0 mg) was performed with CH_3I in dimethyl sulfoxide in the presence of sodium hydroxide as described elsewhere [35]. After extraction with chloroform/water (1:1, v/v), the products were recovered from chloroform and evaporated under nitrogen. The PMAA derivatives obtained after hydrolysis with 2 M CF_3CO_2H at 120 °C for 2 h, N-acetylation, reduction with $NaBD_4$, and peracetylation were analysed by GLC-MS.

The absolute configuration of Rha and FucNAc was determined by GLC of acetylated (*S*)-(+)-2-octyl- and racemic 2-octyl-glycoside derivatives using authentic sugar standards as described previously [36]. In turn, the absolute configuration of Fuc3N was determined based on the analysis of glycosylation effects on ^{13}C resonances in the O-polysaccharide.

The absolute configuration of 3-hydroxybutanoic acid (Hb) was determined according to the published method [37] with some modifications: after hydrolysis of an OPS sample (1 mg) with 2 M CF_3CO_2H (120 °C, 4 h), the product was extracted with EtOAc and, after evaporation under nitrogen, subjected to solvolysis with 2 M HCl in *S*-(+)-2-octanol at 80 °C for 16 h followed by trimethylsilylation. The derivatives obtained were analysed by GLC-MS and compared with the retention time of *O*-TMS (*S*)-2-octyl esters of authentic (*S*)- and (*R*)-3-hydroxybutanoates used as references.

2.4. NMR spectroscopy

NMR spectra were recorded for an OPS sample (8 mg) in 99.98% D_2O (after deuterium-exchange by freeze-drying from 99.9% D_2O) at 32 °C on a 500 MHz NMR Varian Unity Inova instrument and calibrated with external acetone (δ_H 2.225, δ_C 31.45). Standard Varian software (Vnmrj V. 4.2 rev.) was used to acquire and process the NMR data. Homonuclear and heteronuclear two-dimensional experiments: $^1H,^1H$ TOCSY, $^1H,^1H$ DQF-COSY, $^1H,^1H$ NOESY, $^1H,^{13}C$ HSQC, and $^1H,^{13}C$ HMBC were conducted for signal assignments and determination of the sugar sequence. The mixing time of 200 and 100 m s was used in the NOESY and TOCSY experiments, respectively. The $^1H,^{13}C$ HSQC spectrum (band-selective gHSQCAD) measured without ^{13}C decoupling was used to determine the $^1J_{C1,H1}$ coupling constants for the anomeric carbons. The heteronuclear multiple-bond correlation (HMBC) experiment was optimized for $J_{C,H}$ = 8 Hz, with 2-step low-pass filter 130 and 165 Hz to suppress one-bond correlations.

2.5. DNA isolation, whole genome sequencing, and bioinformatics analyses

Genomic DNA was isolated using the bacterial DNA Extraction Kit (EuriX, Poland) according to the manufacturer's protocol and measured with Quant-iT PicoGreen dsDNA (Life Technologies, P11496). For sequencing library construction, genomic DNA was sheared on an ultrasonicator (Covaris E210) and prepared by the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (New England Biolabs). Sequencing was performed on a MiSeq instrument (Illumina, USA) using the v3 (2 × 300 bp) reagent kit.

Raw reads after adapter trimming (fastp 0.12.4 [38]) and quality control (FastQC 0.11.9 [39] yielded 420,2 Mbp. The draft genome was assembled using SPAdes 3.15.3 [40]. The quality of the assembly was evaluated using QUAST 5.0.2 [41]. The genome was annotated using prokka 1.14.5 [42] with TIGRFAMs, Pfam-A, and HAMAP databases. A set of known OGC GenBank sequences deposited with accession numbers MH449673 to MH449686 [20] was used to determine the prokka '–proteins' option.

The DNA sequence of the *A. sobria* K928 O-antigen gene cluster has been deposited in GenBank under the Accession Number OQ411005. Visualization of the K928 OGC was generated by Clinker 0.0.26 [43].

For species identification, the 16S rRNA gene was BLAST-ed (blastn 2.12.0) to the SILVA 138 SSU database. MLPA (multilocus phylogenetic typing analysis) was done based on the selected housekeeping genes *gyrB*, *rpoD*, *recA*, *dnaJ*, *gyrA*, *dnaX*, and *atpD* [44]. These 7 genes from isolate K928 and all *Aeromonas* strains available from NCBI RefSeq were concatenated and aligned using mafft [45] (v7.490). An unrooted maximum likelihood phylogenetic tree (Supplementary Fig. S1) was generated using IQ-TREE [46] (2.1.4) software with bootstrap values of 1000 replicates and the '-m TEST' option to determine the best-fit substitution model.

3. Results and discussion

3.1. SDS-PAGE, isolation of the O-polysaccharide and chemical analyses

The lipopolysaccharide of *A. sobria* strain K928 was isolated from enzymatically treated cell biomass after extraction with the classical procedure and recovered from the phenol phase in a yield of ~2.5% of the dry bacterial cell mass. SDS-PAGE (data not shown) revealed the typical ladder-like pattern characteristic of LPS species containing an O-polysaccharide differing in the number of repeating units.

The OPS was prepared by mild acid degradation of the LPS, after precipitation of lipid A, and separation of the obtained supernatant by GPC on Sephadex G50 fine. The yield of the high-molecular-weight OPS fraction was 11% of the LPS portion subjected to hydrolysis.

Full acid hydrolysis of the O-polysaccharide followed by reduction, acetylation, and analysis by GLC-MS resulted in identification of 6-deoxymannose (Rha), 2-amino-2,6-dideoxygalactose (FucN) and 3-amino-3,6-dideoxygalactose (Fuc3N), which contained a 3-hydroxybutanoyl group (Hb) amido-linked at C-3 (Fuc3NHb), with a peak area ratio of ~1.0 : 1.6 : 1.2. The determination of the absolute configuration of the monosaccharides with the use of a published method [36] showed the L configuration of Rha and the D configuration of FucN. The D configuration of Fuc3N was determined by the analysis of the ^{13}C glycosylation effects in the O-polysaccharide. The *R* configuration of the 3-hydroxybutanoyl Hb group was established by GLC-MS of *O*-TMS (*S*)-2-octyl esters after comparing with the retention time of the authentic standards [37].

The methylation analysis of the O-polysaccharide was performed with methyl iodide in dimethyl sulfoxide in the presence of solid sodium hydroxide [35]. This revealed 4-substituted Rha, 3-substituted Rha, 2, 3-disubstituted Rha, 3-substituted FucN, and terminal Fuc3NHb, with a peak area ratio of ~1.25 : 1.0 : 2.0 : 3.3 : 2.6, identified by GLC-MS as partially methylated alditol and aminoalditol acetates.

3.2. Elucidation of the O-polysaccharide structure by NMR

The ^{1}H NMR spectrum of the O-polysaccharide (Fig. 1) contained five signals for anomeric protons at δ 5.19, 5.15, 4.89, 4.85, and 4.75, with an integral intensity ratio of 1.0 : 1.0 : 0.98 : 1.0 : 0.96; six C*H*$_3$–CH groups of H-6 of 6-deoxyhexoses (5 signals) and H-4 of Hb (1 signal) in the range of δ 1.21–1.35, one C*H*$_2$ group of Hb at δ 2.51, and one signal for an *N*-acetyl group at δ 2.12, signals of C*H*–OH of H-3 of the 3-hydroxybutanoyl group at δ 4.24, and ring proton signals in the range of δ 3.42–4.49; some of the signals overlapped. These data indicated that the O-polysaccharide has a regular structure composed of a pentasaccharide repeating unit.

The ^{1}H and ^{13}C NMR resonances of the OPS were completely assigned in 2D ^{1}H,^{1}H DQF-COSY, TOCSY, NOESY, ^{1}H,^{13}C HSQC, and ^{1}H,^{13}C HMBC experiments. The ^{1}H and ^{13}C NMR chemical shifts are collected in Table 1. Based on intra-residue H,H and H,C correlations and coupling constant values estimated from the 2D NMR spectra, the spin systems assigned to sugar residues demonstrated their pyranose form.

The ^{1}H,^{1}H DQF-COSY, TOCSY, and NOESY spectra revealed five spin systems for sugars (**A-E** units) having the *manno* or *galacto* configuration, and an additional non-sugar spin system corresponding to the *N*-3-hydroxybutanoyl group (**Hb**). The $J_{H1,H2}$ coupling constant values of <4 Hz and ~8 Hz confirmed the different anomeric configurations of *galacto*pyranoses: the α-configuration of unit **A** and the β-configuration of unit **C**, respectively.

In turn, the anomeric configuration of *manno*pyranoses (**B**, **D** and **E** units) was established on the basis of their C-5 positions and $^{1}J_{C1,H1}$ coupling constant values.

The ^{1}H,^{1}H TOCSY spectrum (Fig. 2, red) contained cross-peaks of the H-1 signals with H-2, H-3, and H-4 for units **A** and **C**, indicating the *galacto* configuration. The remaining H-5 and H-6 resonances were assigned from the H-3/H-5, H-4/H-5, and H-4/H-6 correlations in the NOESY spectrum. Additionally, in the ^{1}H,^{13}C HSQC spectrum (Fig. 3), two correlation signals at δ 4.41/52.2 (H-3/C-3) and 4.01/52.4 (H-2/C-2) of the protons at the nitrogen-bearing carbons to the corresponding carbons and the H-6/C-6 signals at δ 1.21/16.3 and 1.26/16.6 were assigned to the spin systems **A** and **C**, respectively. The $^{1}J_{C1,H1}$ coupling constant values of 173 Hz and 163 Hz confirmed the α-anomeric configuration of the **A** unit and the β-configuration of the **C** unit, respectively. In the NOESY spectrum (Fig. 2, black), these data were further corroborated by the correlations of H-1/H-2 for unit **A** and by H-1/H-3 and H-1/H-5 for unit **C**. The **A** and **C** spin systems were eventually identified as α-Fuc3N and β-FucN, respectively.

The position of the N-acyl (Hb) and N-acetyl groups was confirmed in the ^{1}H,^{13}C HMBC spectrum by the correlations between the proton at the nitrogen-bearing carbon: H-3 of α-Fuc3N at δ 4.41 and H-2 of β-FucN at δ 4.01 with the corresponding carbonyl group signals at δ 175.2 and 175.4, respectively, and between the latter signals and methylene and methyl protons at δ 2.51 and 2.12, respectively.

The **B**, **D**, and **E** units (H-1/C-1 cross-peaks at δ 5.19/102.5, 4.85/103.4, and 4.75/101.5, respectively) were typical of *manno*pyranose, as indicated by the vicinal coupling constant values: $^{3}J_{1,2}$ (~2 Hz), $^{3}J_{2,3}$ (3.5 Hz), $^{3}J_{3,4}$ (8.5 Hz), $^{3}J_{4,5}$ (~9 Hz), and $^{3}J_{5,6}$ (6 Hz). They were assigned to Rha*p* residues based on the chemical shift values of their H-6 in the range of δ 1.29–1.35 and C-6 at δ 18.0 [47,48]. In the TOCSY spectrum (Fig. 2, red), correlation signals between H-1 and H-2, H-2 and H-3,4,5,6, and H-6 and H-3,4,5 were observed for Rha*p* residues; however, some signals overlapped, i.e. H-2, H-3, and H-5 of unit **D** and H-4 and H-5 of unit **E**. The long-range correlations observed in the ^{1}H,^{13}C HMBC spectrum (Fig. 4) facilitated recognition of ^{13}C resonances of Rha*p* residues, especially for unit **D**, by analysing the correlations for both anomeric protons (H-1/C-2, H-1/C-3 and H-1/C-5 at δ 4.85/71.0, 4.85/70.2, and 4.85/68.7, respectively) and H-6 protons (H-6/C-4 and H-6/C-5 at δ 1.29/83.7 and 1.29/68.7, respectively). The contacts that were found completed the assignment of the ^{13}C Rha*p* residue

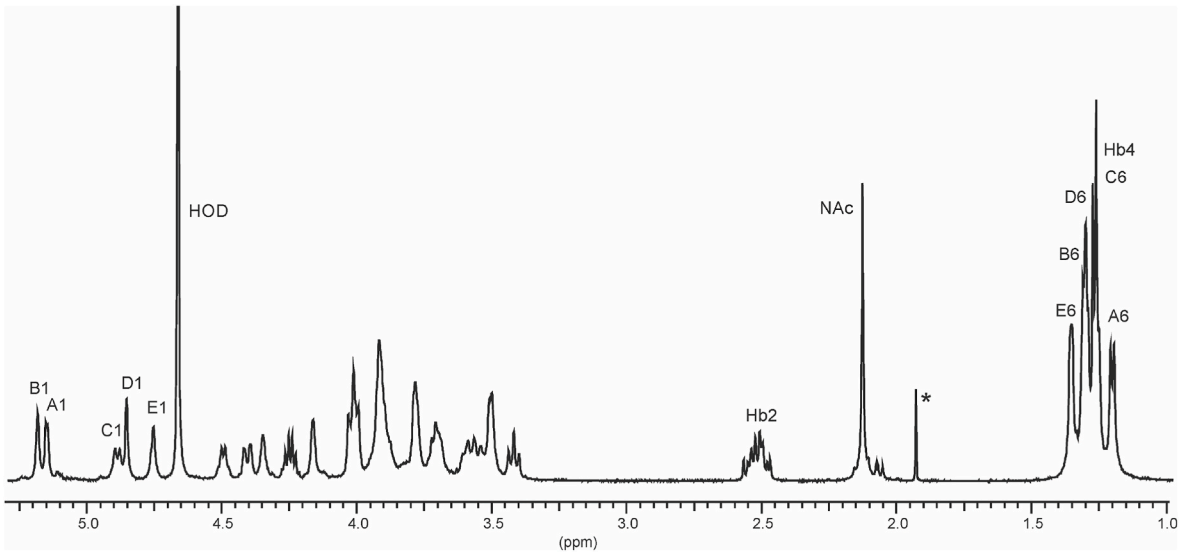


Fig. 1. ^{1}H NMR spectrum of the OPS of *A. sobria* strain K928. The spectrum was recorded at 32 °C in D_2O at 500 MHz. The capital letters and Arabic numerals refer to atoms in the sugar residues denoted as shown in Table 1. NAc - N-acetyl group, asterisk - free acetic acid.

Table 1
^{1}H (500.4 MHz) and ^{13}C NMR (125 MHz) data (δ, ppm) for the OPS of *A. sobria* strain K928.

Sugar residue		H-1 C-1	H-2 C-2	H-3 C-3	H-4 C-4	H-5 C-5	H-6 C-6	NAc
α-D-Fuc*p*3NAcyl-(1→	**A**	5.15	4.03	4.41	3.79	4.49	1.21	
$^{1}J_{C,H}$ = 173 Hz		96.5	67.1	52.2	71.8	68.1	16.3	
→2,3)-α-L-Rha*p*-(1→	**B**	5.19	4.35	4.01	3.42	3.91	1.30	
$^{1}J_{C,H}$ = 176 Hz		102.5	75.8	76.1	72.1	70.6	18.0	
→3)-β-D-Fuc*p*NAc-(1→	**C**	4.89	4.01	3.56	3.78	3.72	1.26	2.12
$^{1}J_{C,H}$ = 163 Hz		103.1	52.4	81.0	71.8	71.9	16.6	23.6; 175.4
→4)-α-L-Rha*p*-(1→	**D**	4.85	3.91	3.91	3.59	3.89	1.29	
$^{1}J_{C,H}$ = 173 Hz		103.4	71.0	70.2	83.7	68.7	18.0	
→3)-β-L-Rha*p*-(1→	**E**	4.75	4.16	3.69	3.50	3.50	1.35	
$^{1}J_{C,H}$ = 163 Hz		101.5	71.7	81.9	72.2	73.2	18.0	
(*R*)-3-hydroxybutanoyl	**Hb**		2.51	4.24	1.27			
		175.2	46.1	66.3	23.1			

resonances inferred from the ^{1}H,^{13}C HSQC spectrum (Fig. 3), especially for some coinciding signals. For the final determination of the chemical shift values, it was also important to take into account the data from the methylation analysis and the effect of glycosylation on the chemical shifts of carbon atoms.

The position of C-5 for the **B** and **D** units at δ 70.6 and 68.7, respectively, demonstrated that they were α-Rha*p*, whereas unit **E** was β-Rha*p* (δ 73.2), in comparison with the chemical shifts of δ 70.0 and 73.0 for C-5 in α-Rha*p* and β-Rha*p*, respectively [49]. The NOESY spectrum (Fig. 2, black) showed intra-residue cross-peaks: H-1/H-2 for the α-Rha*p* residues (**B** and **D**) and H-1/H-2, H-1/H-3, and H-1/H-5 for β-Rha*p* **E** [47,50,51].

The downfield shifted signals of carbon atoms C-2 and C-3 of unit **B**, C-3 of unit **C** and **E**, and C-4 of unit **D**, compared with their positions in the corresponding non-substituted monosaccharides [52,53], confirmed the glycosylation pattern of the monosaccharides in the O-unit. These data demonstrated that the O-polysaccharide repeating unit is branched with residue **B** at the branching point. On the basis of the insignificant displacements of the signals for the carbon atoms C-2,3,4,6 of unit **A** [47, 54], a terminal position of this residue was deduced.

The two-dimensional NOESY experiment showed cross-peaks for the B H-1/E H-3, E H-1/D H-4, D H-1/C H-3, C H-1/B H-2, and A H-1/B H-3 pairs of transglycosidic protons, thus indicating a sequence of the sugars in the linear part of the O-unit and attachment of α-Fuc3N (unit **A**) at position C-3 of the disubstituted α-Rha*p* residue (unit **B**).

In the ^{1}H,^{13}C HMBC spectrum (Fig. 4), the following correlations between anomeric protons and transglycosidic carbons were observed: Rha*p* **B** H-1/Rha*p* **E** C-3 at δ 5.19/81.9, Rha*p* **E** H-1/Rha*p* **D** C-4 at δ 4.75/83.7, Rha*p* **D** H-1/Fuc*p*NAc **C** C-3 at δ 4.85/81.0, Fuc*p*NAc **C** H-1/Rha*p* **B** C-2 at δ 4.89/75.8, and Fuc*p*3NAcyl **A** H-1/Rha*p* **B** C-3 at δ 5.15/76.1. Moreover, in the spectrum the following correlations between anomeric carbons and transglycosidic protons were also visible: Rha*p* **B** C-1/Rha*p* **E** H-3 at δ 102.5/3.69, Rha*p* **E** C-1/Rha*p* **D** H-4 at δ 101.5/3.59, Rha*p* **D** C-1/Fuc*p*NAc **C** H-3 at δ 103.4/3.56, Fuc*p*NAc **C** C-1/Rha*p* **B** H-2 at δ 103.1/4.35, and Fuc*p*3NAcyl **A** C-1/Rha*p* **B** H-3 at δ 96.5/4.01.

Therefore, it was concluded that the OPS of *A. sobria* strain K928 has a branched pentasaccharide repeating unit containing three Rha*p* residues (units **B**, **D**, **E**) and one Fuc*p*NAc residue (unit **C**) in the main chain, while Fuc3NAcyl (unit **A**) is located in the side chain.

The analysis of the glycosylation effects on the ^{13}C NMR chemical shifts allowed establishing the absolute configuration of Fuc*p*3N (unit **A**) and confirmed the absolute configuration of Fuc*p*N (unit **C**).

In the disaccharide fragment **A**-(1 → 3)-**B**, α-Fuc*p*3N-(1 → 3)-α-L-Rha*p*, the small positive α-effect on C-1 of residue **A** (+4.5 ppm) and C-3 of residue **B** (+5.2 ppm) and the large negative β-effect (- 1.0 ppm) on C-4 of **B** indicated different D-L absolute configurations of the linked monosaccharides. In the case of the same L-L absolute configurations, the positive α-effect on C-1 of residue **A** and C-3 of residue **B** would have been much higher: >7.7 ppm and >7.1 ppm, respectively, and the negative β-effect on C-4 of **B** would have been much lower (- 0.6 ± 0.6 ppm) [54,55]. Based on these data, α-Fuc*p*3N **A** was confirmed to have

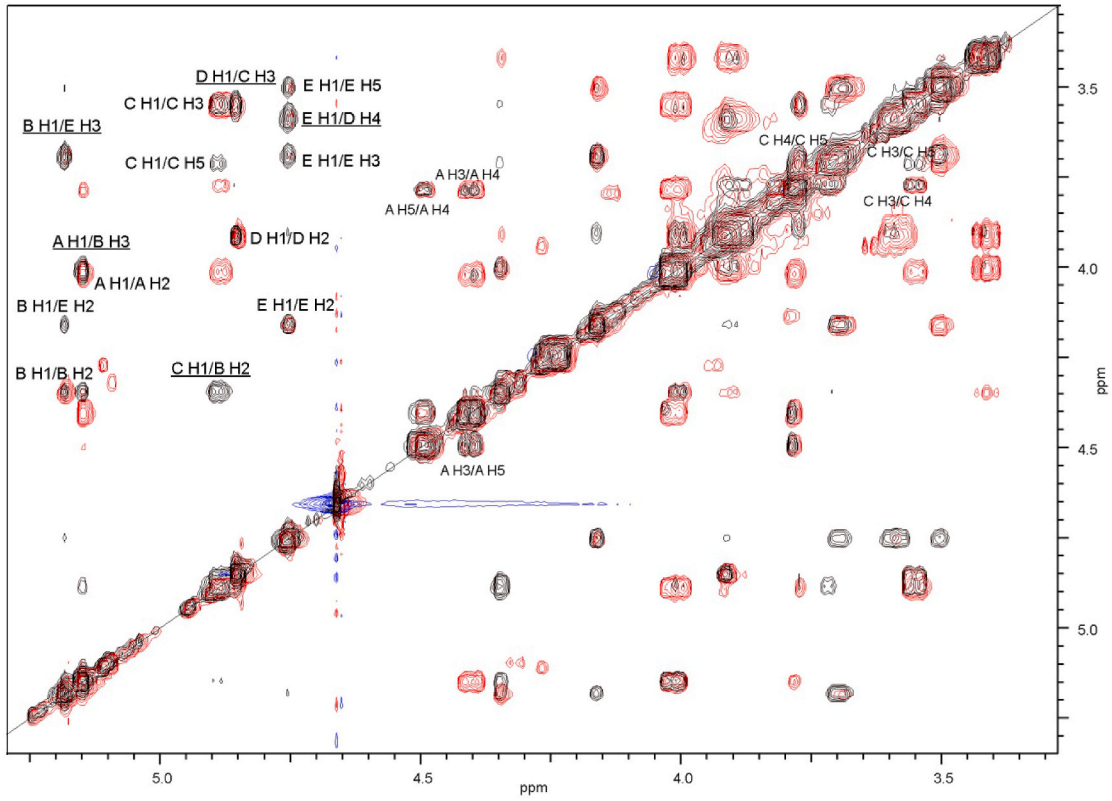


Fig. 2. Parts of ^{1}H,^{1}H TOCSY and ^{1}H,^{1}H NOESY spectra of the OPS from *A. sobria* strain K928. Overlay of the regions of the ^{1}H,^{1}H TOCSY (red) and the ^{1}H,^{1}H NOESY (black) spectra. The NOE correlations between anomeric protons and protons at the glycosidic linkages are underlined. Some other important H/H correlations are depicted as well. The capital letters and Arabic numerals refer to protons in the sugar residues denoted as shown in Table 1.

the D absolute configuration.

In the disaccharide fragment **D**-(1 → 3)-**C**, α-L-Rha*p*-(1 → 3)-β-Fuc*p*N, the large positive α-effect on C-1 of residue **D** (+8.3 ppm) and C-3 of residue **C** (+8.6 ppm) and the small negative β-effect (- 0.1 ppm) on C-4 of **C** indicated different D-L absolute configurations of the linked monosaccharides. In the case of the same L-L absolute configurations, the positive α-effect on C-1 of residue **D** and C-3 of residue **C** would have been much lower: <4 ppm and <6.0 ppm, respectively, and the negative β-effect on C-4 of **C** would have been much higher (- 3.4 ppm). Based on these data, α-Fuc*p*N **C** was confirmed to have the D absolute configuration.

In conclusion, the data showed that the O-specific polysaccharide from *A. sobria* strain K928 has the structure presented below.

The composition of the O-antigen from *A. sobria* strain K928 is unique among bacterial O-polysaccharides (Bacterial Carbohydrate Structure Database: http://glyco.ac.ru/bcsdb) [56], however, it is similar to the recently published structure of the O-polysaccharide of *A. encheleia* strain A4, which was identified as belonging to the same provisional serogroup PGO1, prevailing among the fish-pathogenic aeromonads in Polish aquaculture. The difference between these O-polysaccharides lies in one sugar, i.e. β-D-Fuc*p*NAc, which is replaced by β-D-Qui*p*NAc in the O-polysaccharide of *A. encheleia* A4. Moreover, both O-antigens contain α-D-Fuc3N with an N-linked (*R*)-3-hydroxybutanoyl (Fuc3N*R*Hb) group located in the side chains of the branched O-units. It can be assumed that the terminal position of the α-D-Fuc3N*R*Hb residue is essential for the immunospecificity of *Aeromonas* spp. strains classified into the PGO1 serogroup.

3.3. Characterization of the O-antigen gene cluster OGC

The draft genome sequence of *A. sobria* strain K928 contains 78 contigs with mean coverage of ~92x, a genome size of 4,533,522 bp, and GC content of 57.98%. The size of the putative O-antigen gene cluster OGC in the *A. sobria* K928 (serogroup PGO1) genome is 32,382 bp, and 27 *orfs* (designated as *orf1* - *orf27*) have been identified, whose functions have been assigned based on their similarity to related genes from the available databases (NCBI nr/nt, InterPro-related databases, KEGG). The GC pair content in the OGC is 46.78%, which is lower than the average GC content of the whole K928 genome (57.98%) and is similar to that found in the so far sequenced *A. hydrophila* OGCs [20]. The K928 OGC was located between the housekeeping genes, downstream of *acrB* (*orf1*) and upstream of *oprM (orf27)*, and both share 98–99% identity with their homologs found in likewise organized OGCs of several *A. hydrophila* strains classified into the O7, O9, O10, O13, O23, O30, and O35 O-serogroups [20]. The *oprM* gene is preceded by the *orf26* gene, which encodes the histone-like nucleoid structuring (H-NS) protein, which is implicated in the transcriptional regulation of gene expression (by repression) in response to environmental changes (pH, osmolarity, and temperature) and is important for the pathogenicity of Gram-negative bacteria, including *E. coli*, *Salmonella* spp., *Shigella* spp., and *Vibrio* spp. [57,58]. The *hsn* homolog identified in *A. sobria* K928 OGC and other *A. hydrophila* OGC studied so far may serve a similar function, e.g. in the regulation of exopolysaccharide gene expression, production of pili and fimbriae, and biofilm formation [20,58].

All of the *orfs* had the same transcriptional direction from *oprM* to *acrB* except for *oprM*, which showed the opposite one. The O-antigen repeating unit of *A. sobria* K928 consists of five sugar residues: three Rha, FucNAc, and terminal Fuc3NHb. The functions of putative genes for the synthesis of nucleotide precursors of the sugars, glycosyltransferases, and O-antigen processing genes in the studied OGC cluster were predicted on the basis of homology by searching available databases [59]. These data are summarized in Table 2. The OGC structure is

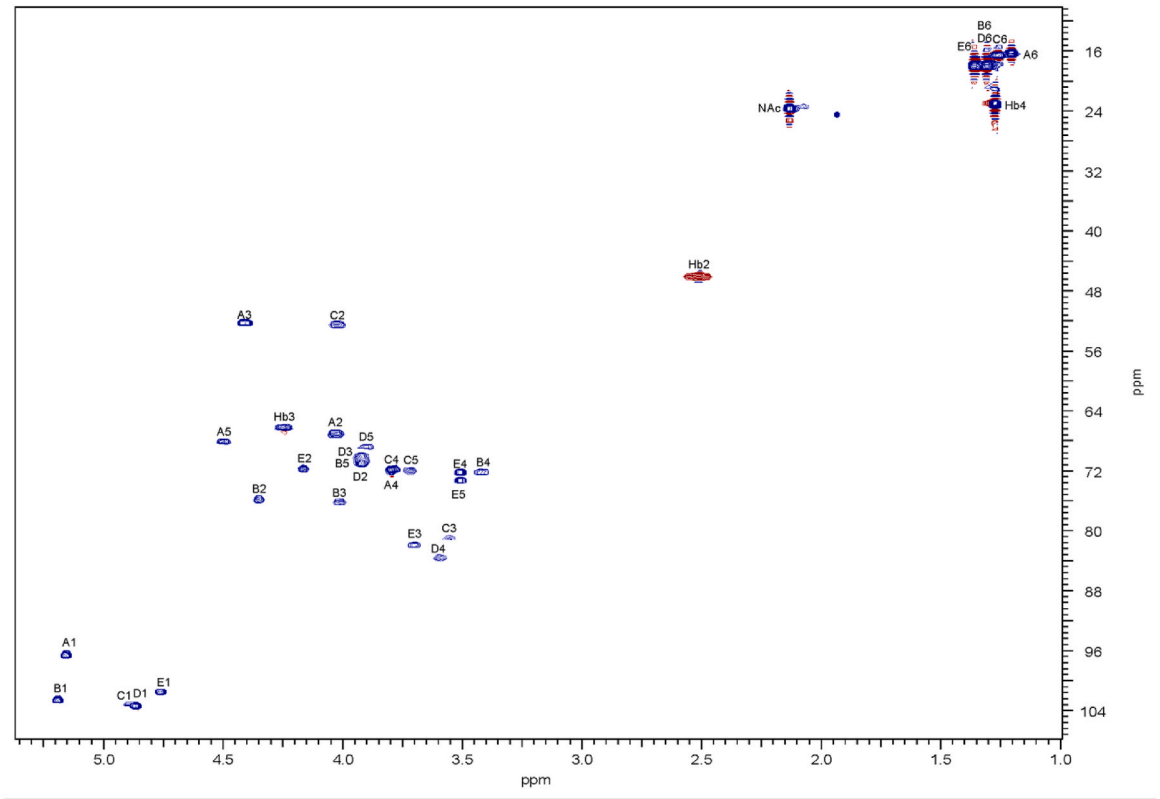


Fig. 3. $^{1}H,^{13}C$ HSQC (500 × 125 MHz) spectrum of the OPS of *A. sobria* strain K928. The spectrum was recorded at 32 °C in D_2O as a solvent. The capital letters and Arabic numerals refer to proton/carbon pairs in the respective sugar residues denoted as shown in Table 1.

shown in Fig. 5.

Orf2, 3, 4, and *5* were assigned to the *rmlB, rmlD, rmlA,* and *rmlC* genes, respectively, based on their high identity with their homologs from *Aeromonas* spp. (94–99%) and *E. coli*. This set of genes encode four enzymes crucial for the biosynthesis of nucleotide-activated dTDP-L-rhamnose, and are usually found towards the 5'-end of the O-antigen gene cluster in the order *rmlBDAC* [20]. Similar location and arrangement of genes were found in the studied OGC.

The next proteins encoded by *orf6, orf7*, and *orf9* share 94%, 55%, and 95% identity levels with FdtA (*A. hydrophila*), FdtC (*Acinetobacter baumannii*), and FdtB (*A. veronii*), respectively, which are known to be involved in the synthesis of dTDP-3,6-dideoxy-(3-hydroxybutanoylamino)-D-galactose (dTDP-D-Fuc3NHb), i.e., the sugar residue identified in the OPS of *A. sobria* K928. In this process, FdtA (dTDP-6-deoxy-3,4-keto-hexulose isomerase) and FdtB (dTDP-6-deoxy-D-*xylo*-hexos-3-ulose aminotransferase) catalyze the synthesis of dTDP-D-Fuc3N, and FdtC (dTDP-D-Fuc3N acyltransferase) is responsible for the addition of a 3-hydroxybutanoyl group to dTDP-D-Fuc3N to give dTDP-D-Fuc3NHb [60,61]. Moreover, the dTDP-D-Fuc3NHb biosynthesis pathway is preceded by the activity of RmlA and RmlB, encoded in the K928 OGC by *orf4* and *orf2*, respectively [60]. Furthermore, *orf7* and *orf8* (*phaB*) found between *fdtA* and *fdtB* encode proteins sharing 45% and 91% identity with butyryltransferase (WbsC) of *E. coli* STEC O91 [62] and 3-oxo-acyl-APC-reductase of *A. hydrophila,* respectively. It is likely that, due to its function, the latter enzyme may also be involved in the formation of a non-sugar (acyl) subsituent [63].

In the K928 OGC, there are five genes (*orf11, 13, 14, 15*, and *17*) encoding putative sugar- and glycosyltransferases of *Aeromonas* spp., *Salmonella* spp., and *Shigella* spp. The *orf13, orf14,* and *orf15* genes encode three out of the four glycosyltransferases (GTs), as can be expected given the pentasaccharide O-unit. The *orf15*-encoded protein belongs to the glycosyltransferase class 2 family and shares 88% identity with a putative GT of *A. caviae*. Based on BLASTp analysis, 47–48% identity has been found between *orf15*-encoded protein and WbsY and RfbG of *S. boydii* O11 and *S. flexneri* type 2a, which are putative dTDP-rhamnosyltranserases. Therefore, based on the data alignment, the *orf15* gene is proposed to encode dTDP-rhamnosyl transferase for the formation of L-Rha-α-(1 → 3)-L-Rha, analogously to the function of WbsY of *S. boydii* O11 [64]. In turn, *orf13* and *orf14* encode putative glycosyltransferases 2, which share 33% and 59% identity with WbsW [65] and WbsX of *S. boydii* O11, respectively [22,64]. Based on the similarity to the proteins available in GenBank and considering those whose function is known and the fact that the *S. boydii* O11 O-antigen contains three Rha residues, the *orf13* and *orf14* genes can be proposed to encode putative rhamnosyltransferases. The products of two other genes, i.e. *orf11* and *orf17*, are referred to as putative GTs. The *orf11*-encoded protein shares 45% and 26% identity with *E. coli* and WdcO of *S. enterica arizonae* O63 glycosyltransferases, respectively. Interestingly, the localization of *wdcO* in the *Salmonella* O63 OGC, whose O-antigen also contains D-Fuc3NAc, is similar to that studied here, i.e. between the *wzx* and *wzy* genes [31].

The assembly of an O-unit occurs via sequencing transfer of monosaccharides from the corresponding nucleotide precursors to the first sugar that is attached to a membrane-linked Und-PP carrier. O-antigen biosynthesis pathways, regardless of which of the three known synthesis modes they represent, are initiated by the transfer of a sugar phosphate from an NDP-sugar to undecaprenyl phosphate (Und-P), and in most species OGC includes an initial transferase (IT) gene for this first step [23,66]. In the K928 OGC, we propose this function to be assigned to *orf17*. The *orf17* gene shares 92% identity with an undecaprenyl-phosphate beta-N-acetyl-D-fucosamine-phosphotransferase and is located upstream of *wzz* in the K928 OGC. The IT may be thus responsible for initializing the sequential synthesis during the O-unit assembly by transfer of the sugar from a corresponding nucleotide precursor (UDP-D-FucNAc) to a membrane-linked undecaprenyl phosphate carrier, resulting in Und-PP-D-FucNAc, which possibly is the initial sugar for the O unit assembly [23,66].

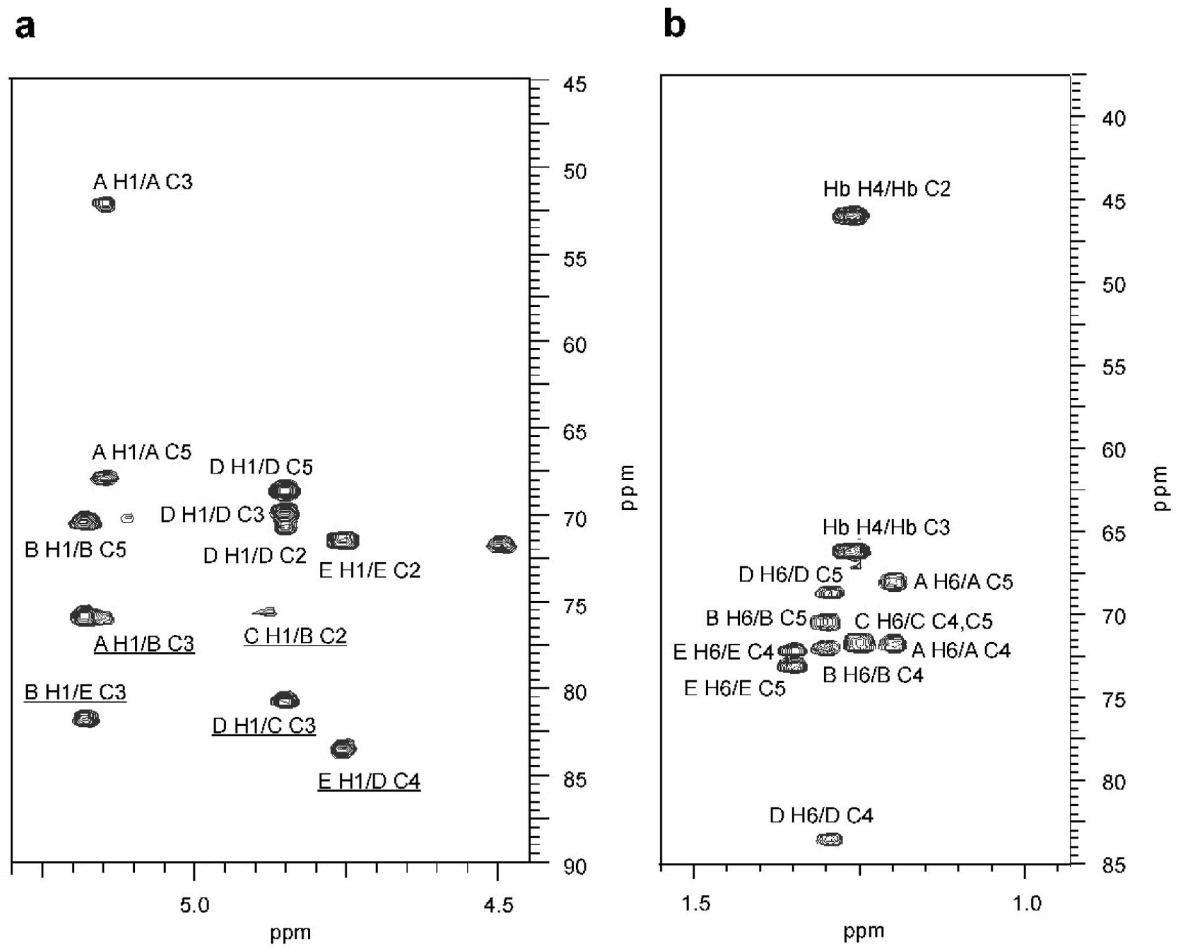
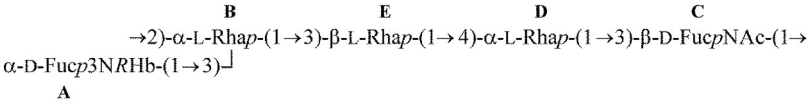


Fig. 4. Regions of the $^{1}H,^{13}C$ HMBC spectrum of the OPS of *A. sobria* strain K928. The long-range heteronuclear correlations for (a) anomeric protons and (b) H-6 protons are marked. Inter-residue correlations important for establishing the sugar sequence in the O-antigen repeating unit are underlined. The capital letters and Arabic numerals refer to atoms in the sugar residues denoted as shown in Table 1.



The *orf16* and *orf18* genes, with 40% and 49% identity to WbpK and WbpM from *Pseudomonas aeruginosa* PAO1 [67], respectively, are implicated in the biosynthesis of D-FucNAc, where UDP-D-GlcNAc is the initial substrate of the process. Although it has been misannotated as a bifunctional UDP-D-GlcNAc 4,6-dehydratase/C4-reductase for many years, WbpM is currently postulated to be a UDP-D-GlcNAc 4,6-dehydratase converting UDP-D-GlcNAc into UDP-4-keto-D-QuiNAc, whose intermediate is then converted to UDP-D-FucNAc by WbpK activity [68]. Other homologs of WbpM are represented by PglF and FlaA1 found in *Campylobacter jejuni* [69] and *Helicobacter pylori* [68], respectively.

Orf10, orf12, orf24, orf20, and *orf25* are suggested to encode proteins responsible for O-antigen processing and translocation. The *orf10* gene is proposed as a putative Wzx (flippase) sharing 88% identity with an O-antigen translocase from *A. salmonicida*. The second one, *orf12*, is proposed as a putative Wzy polymerase, i.e. a protein which shares 39% identity with the *Nitratifractor salsuginis* O-antigen ligase family protein [70]. The third *orf24* is suggested as a *wzz* gene encoding a protein regulating the O-antigen chain length (66% identity with Wzz/FepE/Etk N-terminal domain-containing protein of *A. veronii*). The *orf20* gene encodes an additional protein with the putative function of polysaccharide chain length determinator, whose function is considered below [71]. The *orf25* gene with 95% identity with O-antigen ligase family protein from *Aeromonas* sp. CU5 is assigned to a WaaL encoding gene, whose protein contains twelve predicted membrane-spanning segments and carries out the ligation of the O-antigen to the lipid A/core OS. Considering the similar organization of the genes in the O-antigen gene clusters of several *A. hydrophila* described so far and the fact that, in addition to *orf12*, there is *orf25* encoding protein with a more likely ligase function, we propose the product of *orf12* as the putative polymerase Wzy [20]. The function of *orf10*-and *orf12*-encoded proteins in the O-antigen translocation and polymerization has been predicted taking into account 12 and 9 predicted transmembrane segments, which are the typical topologies of flippase and polymerase, respectively, thus indicating the Wzx/Wzy pathway of the O-antigen biosynthesis in *A. sobria* K928 [71].

The set of five *orfs*: *orf19, orf20, orf21, orf22*, and *orf23* should be considered together. Three orfs: *orf21, orf22*, and *orf23* with 93–97% identity to the *gfcB*, *gfcC*, and *gfcD* genes belong to the Yjb family (lipo) proteins and the group 4 capsule polysaccharide formation (lipo)protein Gfc, both previously found in *A. hydrophila* [72]. As reported by Thurlow

Table 2
Characteristics of the *orfs* in the *A. sobria* K928 strain (serogroup PGO1) O-antigen gene cluster.

orf no.	Gene	Gene position in OGC sequence	No of aa	G + C content [%]	Similar protein(s) (strain, GenBank Accession No.)	% Identical/% similar (no. of aa overlap)	Putative function of protein
1	*acrB*	1–3150	1049	61.1	efflux RND transporter permease subunit, *Aeromonas veronii* (WP_139740608.1)	99/99 (1039)	Multidrug efflux pump subunit AcrB
2	*rmlB*	3725–4810	361	56.6	dTDP-glucose 4,6-dehydratase, *Aeromonas caviae* (WP_257710168.1)	97/98 (358)	dTDP-d-glucose-4,6-dehydratase
3	*rmlD*	4810–5697	295	60.6	dTDP-4-dehydrorhamnose reductase, *Aeromonas* sp. CU5, (WP_098969878.1)	99/98 (293)	dTDP-4-dehydrorhamnose reductase
4	*rmlA*	5810–6688	292	53.5	glucose-1-phosphate thymidylyltransferase RfbA, *Aeromonas salmonicida* (WP_125597703.1)	99/100 (290)	Glucose-1-phosphate thymidylyltransferase
5	*rmlC*	6751–7293	180	43.5	dTDP-4-dehydrorhamnose 3,5-epimerase, *Aeromonas hydrophila* (WP_080688753.1)	94/97 (179)	dTDP-4-dehydrorhamnose 3,5-epimerase
6	*fdtA*	7297–7710	138	35.5	FdtA/QdtA family cupin domain-containing protein, *Aeromonas hydrophila* (WP_252615986.1)	94/97 (138)	dTDP-6-deoxy-3,4-keto-hexulose isomerase
7	*fdtC*	7691–8209	172	32	FdtC, acetyltransferase, *Acinetobacter baumannii* (UEH20402.1) WbsC, putative butyryltransferase, *Escherichia coli* O91 (AAK60452.1)	55/74 (150) 45/60 (169)	Putative butyryltransferase/GNAT family N-acetyltransferase
8	*phaB*	8202–8912	236	38.3	3-oxo-acyl-APC-reductase, *Aeromonas hydrophila* (WP_252615988.1)	91/95 (235)	3-oxo-acyl-APC-reductase
9	*fdtB*	8937–10040	367	46.0	DegT/DnrJ/EryC1/StrS family aminotransferase, *Aeromonas veronii* (WP_193859829.1)	95/97 (364)	Aminotransferase (dTDP-3-amino-3,6-dideoxy-alpha-d-galactopyranose transaminase)
10	*wzx*	10037–11287	416	45.2	O-antigen translocase, *Aeromonas salmonicida* (WP_125607491.1)	88/93 (412)	Flippase
11	*GT2*	11320–12765	481	32.8	Glycosyltransferase, *Escherichia coli* (EIA0234640.1) WdcO, glycosyltransferase, *Salmonella arizonae* O63 (AFW04734.1)	45/68 (347) 26/46 (115)	Glycosyltransferase
12	*wzy*	12989–14065	358	32.6	O-antigen ligase protein, *Nitratifractor salsuginis* (WP_169308536.1)	39/57 (111)	Polymerase
13	*wbsW*	14058–14876	272	30.4	WbsW, *Shigella boydii* (AAS98032.1) Glycosyltransferase, *Aeromonas caviae* (WP_080560547)	33/52 (237) 82/89 (270)	Glycosyltransferase class 2
14	*wbsX*	14890–16014	374	34.9	WbsX, *Shigella boydii* O11 (AAS98033.1)	59/74 (367)	Glycosyltransferase
15	*wbsY*	16020–16919	299	35.8	WbsY, *Shigella boydii* O11 (AAS98034.1) WvaA, glycosyltransferase, *Aeromonas caviae* (WP_010675326.1) RfbG, *Shigella flexneri* type 2a (CAA50773.1)	47/66 (297) 88/93 (297) 48/67 (300)	Glycosyltransferase class 2
16	*wbpK*	16934–17890	317	46.0	WbpK, *Pseudomonas aeruginosa* PAO1 (AAC45865.1)	40/57 (298)	NAD-dependent epimerase/dehydratase
17	*GT1*	17884–18441	185	45.7	Undecaprenyl-phosphate beta-N-acetyl-d-fucosamine-phosphotransferase, *Aeromonas hydrophila* (BDC81682.1)	92/95 (185)	Undecaprenyl-phosphate beta-N-acetyl-d-fucosamine-phosphotransferase
18	*wbpM*	18501–20462	653	51.8	WbpM, *Pseudomonas aeruginosa* (AAF23989.1)	49/66 (620)	UDP-N-acetylglucosamine-4,6-dehydratase
19	*wza*	20719–23343	874	48.3	SLBB domain-containing protein, *Aeromonas veronii* (WP_159143412.1)	95/97 (848)	Polysaccharide export protein
20	*wzz*	23421–24380	319	46.1	WzzB/FepE family protein, *Aeromonas veronii* (QHC09997.1)	94/96 (319)	Polysaccharide chain length determinant protein
21	*gfcB*	24855–25535	226	41.3	YjbF family lipoprotein /Group 4 capsule polysaccharide formation lipoprotein GfcB, *Aeromonas hydrophila* (WP_162791230.1)	94/97 (224)	O-antigen capsule production lipoprotein
22	*gfcC*	25532–26293	253	44.4	Capsule biosynthesis GfcC family protein, *Aeromonas hydrophila* (WP_252616001.1)	97/98 (251)	Putative O-antigen capsule production periplasmic protein
23	*gfcD*	26293–28383	696	43.7	YjbH domain-containing protein, *Aeromonas hydrophila* (WP_080938066.1)	93/96 (690)	Putative outer membrane lipoprotein

(*continued on next page*)

Table 2 (*continued*)

orf no.	Gene	Gene position in OGC sequence	No of aa	G + C content [%]	Similar protein(s) (strain, GenBank Accession No.)	% Identical/% similar (no. of aa overlap)	Putative function of protein
24	*wzz*	28453–29526	357	37.1	Wzz/FepE/Etk N-terminal domain-containing protein, *Aeromonas veronii* (WP_245147503.1)	66/79 (353)	O-antigen chain length determinant protein
25	*waaL*	29625–31358	577	49.1	O-antigen ligase family protein, *Aeromonas* sp. CU5 (WP_098969870.1)	95/95 (566)	Ligase
26	*hns*	31972–32382	136	53.8	H-NS histone family protein, *Aeromonas sobria* (WP_101315486.1)	99/99 (135)	DNA-binding protein H-NS
27	*oprM*	32908–34272	454	63.0	Efflux transporter outer membrane subunit, *Aeromonas sobria* (WP_101324081.1)	98/98 (450)	Outer membrane protein OprM

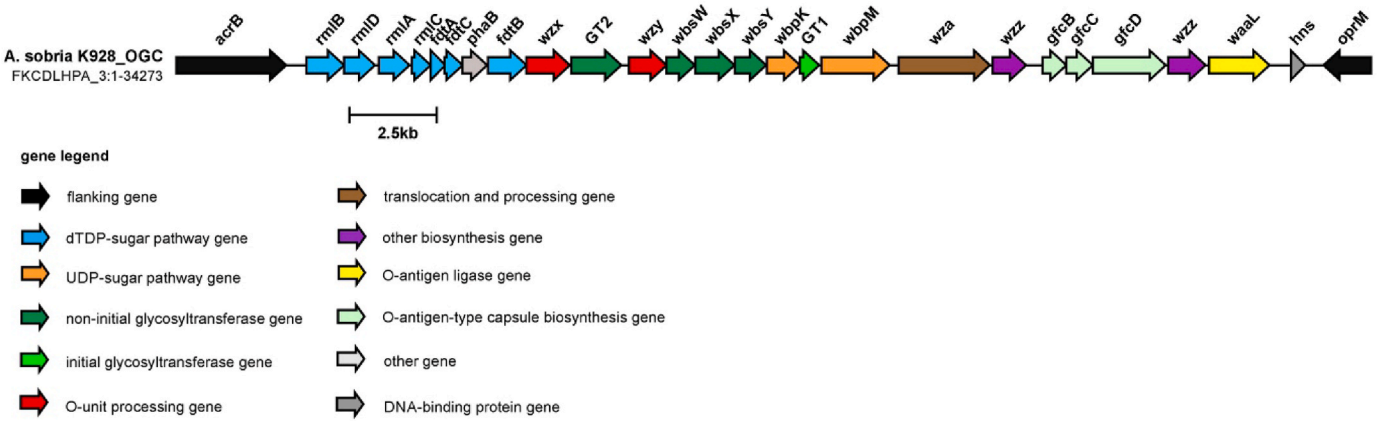


Fig. 5. Composition and organization of the O-antigen gene cluster OGC of *A. sobria* K928. Visualization of the OGC region was generated by Clinker [43].

et al., 2019, the genes form the common operon *gfcABCD* (or the truncated variant *gfcBCD*) and encode proteins for the O-antigen–type capsule assembly, which results in the high virulence of *A. hydrophila* ML09-119, i.e. the strain responsible for epidemic MAS outbreaks in catfish. The operon *gfcABCD/gfcBCD* [72] has homologs in other strains, e.g. *Vibrio cholerae* O139 *wbfDCB* and *E. coli yjbEFGH* [73,74]. The *gfcB, gfcC*, and *gfcD* are variants of *yjbC, yjbB, yjbA* in *E. coli* K-12 substrain MG1655 [74] and are flanked by *wzz* and *waaL* [72]. Similar organization of the operon has been found in the *A. sobria* K928 OGC. Moreover, the *gfcB* (*ymcC)* gene accompanied by other *orfs*, yet without annotated function, have been found in *A. hydrophila* OGCs [20]. By searching databases, we found similarities between the *orf21,22,23* genes and undefined *orfs*. In the K928 OGC, *orf19* and *orf20* assigned to *wza* and *wzz* are located upstream operon *gfcBCD* and encode proteins presumably involved in the polymerization and export of the O-antigen–type capsule, whose expression may contribute to the virulence of *A. hydrophila* [72]. Studies showed that removal of the *gfc* operon in *S. sonnei* followed by loss of the O-antigen-type capsule resulted in avirulence [72,75].

4. Conclusions

The composition and proposed organization of the O-antigen gene cluster identified and characterized in the genome of *A. sobria* K928 was found to be consistent with the structure of the branched O-specific polysaccharide established in this work using chemical methods (sugar and methylation analyses) and one- and two-dimensional NMR spectroscopy. Chemical studies of O-antigens together with the gene cluster organization will help improve diagnostic tools for *Aeromonas* serotyping, epidemiological monitoring, and prevention of diseases. Such comprehensive studies can provide insight into determinants responsible for phenotypic alterations within the serotype(s), most often identified as the causative agent of MAS/MAI diseases in cultured fish, and may contribute to the selection of appropriate strains/antigens as vaccine candidates.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors wish to thank Doctor Paweł Sowiński for recording the NMR spectra (Intercollegiate NMR Laboratory, Department of Chemistry, Technical University of Gdansk, Poland) and MSc Hubert Pietras for his technical support in LPS isolation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.carres.2023.108809.

References

[1] A. Fernández-Bravo, M.J. Figueras, An update on the genus *Aeromonas*: taxonomy, epidemiology, and pathogenicity, Microorganisms 8 (2020) 129, https://doi.org/10.3390/microorganisms8010129.

[2] A. Kozinska, L. Guz, The effect of various *Aeromonas bestiarum* vaccines on non-specific immune parameters and protection of carp (*Cyprinus carpio* L.), Fish Shellfish Immunol. 16 (2004) 437–445, https://doi.org/10.1016/j.fsi.2003.08.003.

[3] Y. Zhou, L. Yu, Z. Nan, P. Zhang, B. Kan, D. Yan, J. Su, Taxonomy, virulence genes and antimicrobial resistance of *Aeromonas* isolated from extra-intestinal and intestinal infections, BMC Infect. Dis. 19 (2019) 158, https://doi.org/10.1186/s12879-019-3766-0.

[4] T.S. Le, T.H. Nguyen, H.P. Vo, V.C. Doan, H.L. Nguyen, M.T. Tran, T.T. Tran, P. C. Southgate, D. İpek Kurtböke, Protective effects of bacteriophages against *Aeromonas hydrophila* species causing motile aeromonas Septicemia (MAS) in striped catfish, Antibiotics 7 (2018) 16, https://doi.org/10.3390/antibiotics7010016.

[5] P.K. Saharia, I.A. Hussain, H. Pokhrel, B. Kalita, G. Borah, R. Yasmin, Prevalence of motile aeromonas septicaemia (MAS) in fish culture systems of the central brahmaputra valley zone of Assam, India, Aquacult. Res. 52 (2020) 1201–1214, https://doi.org/10.1111/are.14979.

[6] R.B.G. Pessoa, W.F. de Oliveira, M.T. dos, S. Correia, A. Fontes, L.C.B.B. Coelho, *Aeromonas* and human health disorders: clinical approaches, Front. Microbiol. 13 (2022), 868890, https://doi.org/10.3389/fmicb.2022.868890.

[7] J.M. Tomás, The main *Aeromonas* pathogenic factors, ISRN Microbiol (2012) 1–22, https://doi.org/10.5402/2012/256261, 2012.

[8] J. Matys, A. Turska-Szewczuk, A. Sroka-Bartnicka, Role of bacterial secretion systems and effector proteins - insights into *Aeromonas* pathogenicity mechanisms, Acta Biochim. Pol. 67 (2020) 283–293, https://doi.org/10.18388/abp.2020_5410.

[9] Y. Li, P. Xia, J. Wu, A. Huang, G. Bu, F. Meng, F. Kong, X. Cao, X. Han, G. Yu, X. Pan, S. Yang, X. Zeng, X. Du, The potential sensing molecules and signal cascades for protecting teleost fishes against lipopolysaccharide, Fish Shellfish Immunol. 97 (2020) 235–247, https://doi.org/10.1016/j.fsi.2019.12.050.

[10] L.R. Stromberg, N.W. Hengartner, K.L. Swingle, R.A. Moxley, S.W. Graves, G. A. Montaño, H. Mukundan, Membrane insertion for the detection of lipopolysaccharides: exploring the dynamics of amphiphile-in-lipid assays, PLoS One 11 (2016), e0156295, https://doi.org/10.1371/journal.pone.0156295.

[11] L. Mazgaeen, P. Gurung, Recent advances in lipopolysaccharide recognition systems, Int. J. Mol. Sci. 21 (2020) 379, https://doi.org/10.3390/ijms21020379.

[12] E. Mendoza-Barberá, S. Merino, J. Tomás, Surface glucan structures in *Aeromonas* spp, Mar. Drugs 19 (2021) 1–19, https://doi.org/10.3390/MD19110649.

[13] F. Di Lorenzo, K.A. Duda, R. Lanzetta, A. Silipo, C. De Castro, A. Molinaro, A journey from structure to function of bacterial lipopolysaccharides, Chem. Rev. 122 (2021) 15767–15821, https://doi.org/10.1021/acs.chemrev.0c01321.

[14] R. Sakazaki, T. Shimada, O-serogrouping scheme for mesophilic *Aeromonas* strains, Jpn. J. Med. Sci. Biol. 37 (1984) 247–255, https://doi.org/10.1016/S0923-2508(98)80323-9.

[15] X. Pan, L. Lin, Y. Xu, X. Yuan, J. Yao, W. Yin, G. Hao, J. Shen, Draft genome sequence of *Aeromonas hydrophila* strain BSK-10 (serotype O97), isolated from Carassius carassius with motile aeromonad septicemia in China, Genome Announc. 5 (5) (2017), https://doi.org/10.1128/genomeA.00497-17 e00497-17.

[16] A. Kozińska, Genotypic and Serological Analysis of Domestic Mesophilic Isolates Aeromonas Sp. In Terms of Pathogenicity and the Type of Disease Symptoms Caused by Them in Fish, Habilitation thesis, The National Veterinary Institute—The National Research Institute, Puławy, Poland, 2009.

[17] A. Kozińska, A. Pękala, Serotyping of *Aeromonas* species isolated from Polish fish farms in relation to species and virulence phenotype of the bacteria, Bull. Vet. Inst. Pulawy 54 (2010) 315–320.

[18] M. Kurzylewska, K. Dworaczek, A. Turska-Szewczuk, Structure of the lipopolysaccharide O-antigen of *Aeromonas encheleia* strain A4 representing the new PGO1 serogroup of aeromonads prevailing in Polish aquaculture, Carbohydr. Res. 519 (2022), 108602, https://doi.org/10.1016/j.carres.2022.108602.

[19] K. Dworaczek, M. Kurzylewska, M. Laban, D. Drzewiecka, A. Pękala-safińska, A. Turska-szewczuk, Structural studies of the lipopolysaccharide of *Aeromonas veronii* bv. *sobria* strain K133 which represents new provisional serogroup PGO1 prevailing among mesophilic *Aeromonads* on polish fish farms, Int. J. Mol. Sci. 22 (2021) 4272, https://doi.org/10.3390/ijms22084272.

[20] H. Cao, M. Wang, Q. Wang, T. Xu, Y. Du, H. Li, C. Qian, Z. Yin, L. Wang, Y. Wei, P. Wu, X. Guo, B. Yang, B. Liu, Identifying genetic diversity of O antigens in *Aeromonas hydrophila* for molecular serotype detection, PLoS One 13 (2018), e0203445, https://doi.org/10.1371/journal.pone.0203445.

[21] C. DebRoy, P.M. Fratamico, X. Yan, G.M. Baranzoni, Y. Liu, D.S. Needleman, R. Tebbs, C.D. O'Connell, A. Allred, M. Swimley, M. Mwangi, V. Kapur, J. A. Raygoza Garay, E.L. Roberts, R. Katani, 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 11 (2016), e0147434, https://doi.org/10.1371/journal.pone.0147434.

[22] L. Feng, S.N. Senchenkova, W. Wang, A.S. Shashkov, B. Liu, S.D. Shevelev, D. Liu, Y.A. Knirel, L. Wang, Structural and genetic characterization of the *Shigella boydii* type 18 O antigen, Gene 355 (2005) 79–86, https://doi.org/10.1016/j.gene.2005.06.001.

[23] B. Liu, A. Furevi, A.V. Perepelov, X. Guo, H. Cao, Q. Wang, P.R. Reeves, Y.A. Knirel, L. Wang, G. Widmalm, Structure and genetics of *Escherichia coli* O antigens, FEMS Microbiol. Rev. 44 (2020) 655–683, https://doi.org/10.1093/femsre/fuz028.

[24] X. Yu, A. Torzewska, X. Zhang, Z. Yin, D. Drzewiecka, H. Cao, B. Liu, Y.A. Knirel, A. Rozalski, L. Wang, Genetic diversity of the O antigens of *Proteus* species and the development of a suspension array for molecular serotyping, PLoS One 12 (2017), e0183267, https://doi.org/10.1371/journal.pone.0183267.

[25] D. Xi, X. Wang, K. Ning, Q. Liu, F. Jing, X. Guo, B. Cao, O-antigen gene clusters of *Plesiomonas shigelloides* serogroups and its application in development of a molecular serotyping scheme, Front. Microbiol. 10 (2019) 741, https://doi.org/10.3389/fmicb.2019.00741.

[26] O.I. Naumenko, H. Zheng, A.S. Shashkov, Y. Sun, S.N. Senchenkova, L. Bai, J. Wang, H. Wang, Q. Li, Y.A. Knirel, Y. Xiong, *Escherichia albertii* EA046 (O9) harbors two polysaccharide gene clusters for synthesis of the O-antigen by the Wzx/Wzy-dependent pathway and a mannan shared by *Escherichia coli* O8 by the Wzm/Wzt-dependent pathway, Int. J. Biol. Macromol. 142 (2020) 609–614, https://doi.org/10.1016/j.ijbiomac.2019.09.135.

[27] Y. Bi, E. Mann, C. Whitfield, J. Zimmer, Architecture of a channel-forming O-antigen polysaccharide ABC transporter, Nature 176 (2018) 139–148, https://doi.org/10.1038/nature25190.

[28] H. Wang, H. Zheng, Q. Li, Y. Xu, J. Wang, P. Du, X. Li, X. Liu, L. Zhang, N. Zou, G. Yan, Z. Zhang, H. Jing, J. Xu, Y. Xiong, Defining the genetic features of O-antigen biosynthesis gene cluster and performance of an O-antigen serotyping scheme for *Escherichia albertii*, Front. Microbiol. 8 (2017) 1857, https://doi.org/10.3389/fmicb.2017.01857.

[29] S.T. Islam, J.S. Lam, Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent pathway, Can. J. Microbiol. 60 (2014) 697–716, https://doi.org/10.1139/cjm-2014-0595.

[30] B. Liu, Y.A. Knirel, L. Feng, A.V. Perepelov, S.N. Senchenkova, Q. Wang, P. R. Reeves, L. Wang, Structure and genetics of *Shigella* O antigens, FEMS Microbiol. Rev. 32 (2008) 627–653, https://doi.org/10.1111/j.1574-6976.2008.00114.x.

[31] B. Liu, Y.A. Knirel, L. Feng, A.V. Perepelov, S.N. Senchenkova, P.R. Reeves, L. Wang, Structural diversity in *Salmonella* O antigens and its genetic basis, FEMS Microbiol. Rev. 38 (2013) 56–89, https://doi.org/10.1111/1574-6976.12034.

[32] Z. Duan, T. Niedziela, C. Lugowski, B. Cao, T. Wang, L. Xu, B. Yang, B. Liu, L. Wang, Genetic diversity of O-antigens in *Hafnia alvei* and the development of a suspension array for serotype detection, PLoS One 11 (2016) 1–16, https://doi.org/10.1371/journal.pone.0155115.

[33] A.V. Perepelov, A.V. Filatov, W. Zhu, A.S. Shashkov, M. Wang, X. Guo, Structure and gene cluster of the O-antigen of *Enterobacter cloacae* G3422, Carbohydr. Res. 510 (2021), 108440, https://doi.org/10.1016/j.carres.2021.108440.

[34] O. Westphal, K. Jann, Bacterial lipopolysaccharides. Extraction with phenol water and further applications of the procedure, in: R.L. Whistler (Ed.), Methods Carbohydr. Chem., Academic Press, New York, 1965, pp. 83–91.

[35] I. Ciucanu, F. Kerek, A simple and rapid method for the permethylation of carbohydrates, Carbohydr. Res. 131 (1984) 209–217, https://doi.org/10.1016/0008-6215(84)85242-8.

[36] K. Leontein, B. Lindberg, J. Lönngren, Assignment of absolute configuration of sugars by g.l.c. of their acetylated glycosides formed from chiral alcohols, Carbohydr. Res. 62 (1978) 359–362, https://doi.org/10.1016/S0008-6215(00)80882-4.

[37] L. Kenne, B. Lindberg, M.M. Rahman, Structural studies of the *Vibrio mimicus* W-26768 O-antigen polysaccharide, Carbohydr. Res. 243 (1993) 131–138, https://doi.org/10.1016/0008-6215(93)84086-L.

[38] S. Chen, Y. Zhou, Y. Chen, J. Gu, Fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics 34 (2018) 884–890, https://doi.org/10.1093/bioinformatics/bty560.

[39] S. Andrews, FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. (Accessed 9 January 2023) accessed.

[40] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V. M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, A.V. Pyshkin, A.V. Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev, P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, J. Comput. Biol. 19 (2012) 455–477, https://doi.org/10.1089/cmb.2012.0021.

[41] A. Mikheenko, A. Prjibelski, V. Saveliev, D. Antipov, A. Gurevich, Versatile genome assembly evaluation with QUAST-LG, Bioinformatics 34 (2018) 142–150, https://doi.org/10.1093/bioinformatics/bty266.

[42] T. Seemann, Prokka: rapid prokaryotic genome annotation, Bioinformatics 30 (2014) 2068–2069, https://doi.org/10.1093/bioinformatics/btu153.

[43] C.L.M. Gilchrist, Y.H. Chooi, Clinker & clustermap.js: automatic generation of gene cluster comparison figures, Bioinformatics 37 (2021) 2473–2475, https://doi.org/10.1093/bioinformatics/btab007.

[44] Y. Du, H. Li, Z. Yin, A. Rozalski, A. Torzewska, P. Yang, C. Qian, T. Xu, H. Cao, P. Wu, L. Jiang, X. Guo, D. Huang, B. Liu, Development of a molecular serotyping scheme and a multiplexed luminex-based array for Providencia, J. Microbiol. Methods 153 (2018) 14–23, https://doi.org/10.1016/j.mimet.2018.08.009.

[45] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (2013) 772–780, https://doi.org/10.1093/molbev/mst010.

[46] L.T. Nguyen, H.A. Schmidt, A. Von Haeseler, B.Q. Minh, IQ-TREE, A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies, Mol. Biol. Evol. 32 (2014) 268–274, https://doi.org/10.1093/molbev/msu300.

[47] E.L. Zdorovenko, G.V. Zatonsky, G.M. Zdorovenko, L.A. Pasichnyk, A.S. Shashkov, Y.A. Knirel, Structural heterogeneity in the lipopolysaccharides of *Pseudomonas syringae* with O-polysaccharide chains having different repeating units, Carbohydr. Res. 336 (2001) 329–336, https://doi.org/10.1016/S0008-6215(01)00268-3.

[48] A. Turska-Szewczuk, A. Kozińska, R. Russa, O. Holst, The structure of the O-specific polysaccharide from the lipopolysaccharide of *Aeromonas bestiarum* strain 207, Carbohydr. Res. 345 (2010) 680–684, https://doi.org/10.1016/j.carres.2009.12.030.

[49] A.S. Shashkov, S.N. Senchenkova, P. Laux, B.C. Ahohuendo, M.L. Kecskés, K. Rudolph, Y.A. Knirel, Structure of the O-chain polysaccharide of the lipopolysaccharide of *Xanthomonas campestris* pv. *manihotis* GSPB 2755 and GSPB 2364, Carbohydr. Res. 323 (2000) 235–239, https://doi.org/10.1016/S0008-6215(99)00262-1.

[50] G.M. Lipkind, A.S. Shashkov, S.M. Suren, N. Kochetkov, The nuclear overhauser effect and structural factors determining the conformations of disaccharide

glycosides, Carbohydr. Res. 181 (1988) 1–12, https://doi.org/10.1016/0008-6215(88)84018-7.
[51] S.N. Senchenkova, A.S. Shashkov, M.L. Kecskés, B.C. Ahohuendo, Y.A. Knirel, K. Rudolph, Structure of the O-specific polysaccharides of the lipopolysaccharides of *Xanthomonas campestris* pv. *vignicola* GSPB 2795 and GSPB 2796, Carbohydr. Res. 329 (2000) 831–838, https://doi.org/10.1016/S0008-6215(00)00250-0.
[52] K. Bock, C. Pedersen, Carbon-13 nuclear magnetic resonance spectroscopy of monosaccharides, Adv. Carbohydr. Chem. Biochem. 41 (1983) 27–66, https://doi.org/10.1016/S0065-2318(08)60055-4.
[53] P.E. Jansson, L. Kenne, G. Widmalm, Computer-assisted structural analysis of polysaccharides with an extended version of casper using ^{1}H- and ^{13}C-n.m.r. data, Carbohydr. Res. 188 (1989) 169–191, https://doi.org/10.1016/0008-6215(89)84069-8.
[54] G.M. Lipkind, A.S. Shashkov, Y.A. Knirel, E.V. Vinogradov, N.K. Kochetkov, A computer-assisted structural analysis of regular polysaccharides on the basis of ^{13}C-N.M.R. data, Carbohydr. Res. 175 (1988) 59–75, https://doi.org/10.1016/S0008-6215(01)00214-2.
[55] A.S. Shashkov, G.M. Lipkind, Y.A. Knirel, N.K. Kochetkov, Stereochemical factors determining the effects of glycosylation on the ^{13}C chemical shifts in carbohydrates, Magn. Reson. Chem. 26 (1988) 735–747, https://doi.org/10.1002/mrc.1260260904.
[56] P.V. Toukach, K.S. Egorova, Carbohydrate structure database merged from bacterial, archaeal, plant and fungal parts, Nucleic Acids Res. 44 (2016) D1229–D1236, https://doi.org/10.1093/nar/gkv840.
[57] S. Fitzgerald, S.C. Kary, E.Y. Alshabib, K.D. MacKenzie, D.M. Stoebel, T.C. Chao, A. D.S. Cameron, Redefining the H-NS protein family: a diversity of specialized core and accessory forms exhibit hierarchical transcriptional network integration, Nucleic Acids Res. 48 (2020) 10184–10198, https://doi.org/10.1093/nar/gkaa709.
[58] H. Wang, J.C. Ayala, A.J. Silva, J.A. Benitez, The histone-like nucleoid structuring protein (H-NS) Is a repressor of *Vibrio cholerae* exopolysaccharide biosynthesis (*vps*) genes, Appl. Environ. Microbiol. 78 (2012) 2482–2488, https://doi.org/10.1128/AEM.07629-11.
[59] A.V. Perepelov, A.V. Filatov, M. Wang, A.S. Shashkov, L. Wang, Y.A. Knirel, Structure and gene cluster of the O-antigen of *Enterobacter cloacae* G3421, Carbohydr. Res. 427 (2016) 55–59, https://doi.org/10.1016/j.carres.2016.03.008.
[60] O.V. Sizova, A.S. Shashkov, A.N. Kondakova, Y.A. Knirel, R.Z. Shaikhutdinova, S. A. Ivanov, A.A. Kislichkina, L.A. Kadnikova, A.G. Bogun, S.V. Dentovskaya, Full structure and insight into the gene cluster of the O-specific polysaccharide of *Yersinia intermedia* H9-36/83 (O:17), Carbohydr. Res. 460 (2018) 51–56, https://doi.org/10.1016/j.carres.2018.02.014.
[61] J.B. Thoden, M.F. Goneau, M. Gilbert, H.M. Holden, Structure of a sugar N -formyltransferase from *Campylobacter jejuni*, Biochemistry 52 (2013) 6114–6126, https://doi.org/10.1021/bi4009006.
[62] S. Perelle, F. Dilasser, J. Grout, P. Fach, Identification of the O-antigen biosynthesis genes of *Escherichia coli* O91 and development of a O91 PCR serotyping test, J. Appl. Microbiol. 93 (2002) 758–764, https://doi.org/10.1046/j.1365-2672.2002.01743.x.
[63] J.M. Shively, G.C. Cannon, S. Heinhorst, D.A. Bryant, S. DasSarma, D. Bazylinski, J. Preiss, A. Steinbüchel, R. Docampo, C. Dahl, Bacterial and Archaeal Inclusions, ELS, 2011, pp. 1–11, https://doi.org/10.1002/9780470015902.a0000302.pub3.
[64] J. Tao, L. Feng, H. Guo, Y. Li, L. Wang, The O-antigen gene cluster of *Shigella boydii* O11 and functional identification of its wzy gene, FEMS Microbiol. Lett. 234 (2004) 125–132, https://doi.org/10.1016/j.femsle.2004.03.021.
[65] N. Mullane, P. O'Gaora, J.E. Nally, C. Iversen, P. Whyte, P.G. Wall, S. Fanning, Molecular analysis of the *Enterobacter sakazakii* O-antigen gene locus, Appl. Environ. Microbiol. 74 (2008) 3783–3794, https://doi.org/10.1128/AEM.02302-07.
[66] J. Wang, H.B. Jiao, X.F. Zhang, Y.Q. Zhang, N. Sun, Y. Yang, Y. Wei, B. Hu, X. Guo, Two enteropathogenic *Escherichia coli* strains representing novel serotypes and investigation of their roles in adhesion, J. Microbiol. Biotechnol. 31 (2021) 1191–1199, https://doi.org/10.4014/jmb.2105.05016.
[67] J.S. Lam, M. Matewish, K.K.H. Poon, Lipopolysaccharides of Pseudomonas aeruginosa, in: J.-L. Ramos (Ed.), *Pseudomonas*. Biosynth. Macromol. Mol. Metab., Springer Science+Business Media, LLC, Granada, Spain, 2004, pp. 3–53, https://doi.org/10.1007/978-1-4419-9088-4.
[68] J.D. King, D. Kocíncová, E.L. Westman, J.S. Lam, Lipopolysaccharide biosynthesis in *Pseudomonas aeruginosa*, Innate Immun. 15 (2009) 261–312, https://doi.org/10.1177/1753425909106436.
[69] U. Vogel, K. Beerens, T. Desmet, Nucleotide sugar dehydratases: structure, mechanism, substrate specificity, and application potential, J. Biol. Chem. 298 (2022), 101809, https://doi.org/10.1016/j.jbc.2022.101809.
[70] J.D. Klena, E. Pradel, C.A. Schnaitman, Comparison of lipopolysaccharide biosynthesis genes *rfaK, rfaL, rfaY*, and *rfaZ* of *Escherichia coli* K-12 and *Salmonella typhimurium*, J. Bacteriol. 174 (1992) 4746–4752, https://doi.org/10.1128/jb.174.14.4746-4752.1992.
[71] Y. Li, J. Huang, X. Wang, C. Xu, T. Han, X. Guo, Genetic characterization of the O-antigen and development of a molecular serotyping scheme for *Enterobacter cloacae*, Front. Microbiol. 11 (2020) 1–12, https://doi.org/10.3389/fmicb.2020.00727.
[72] C.M. Thurlow, M.J. Hossain, D. Sun, P. Barger, L. Foshee, B.H. Beck, J.C. Newton, J.S. Terhune, M.A. Saper, M.R. Liles, The gfc operon is involved in the formation of the O antigen capsule in *Aeromonas hydrophila* and contributes to virulence in channel catfish, Aquaculture 512 (2019), 734334, https://doi.org/10.1016/j.aquaculture.2019.734334.
[73] M.R. Larson, K. Biddle, A. Gorman, S. Boutom, I. Rosenshine, M.A. Saper, *Escherichia coli* O127 group 4 capsule proteins assemble at the outer membrane, PLoS One 16 (2021), e0259900, https://doi.org/10.1371/journal.pone.0259900.
[74] L. Ferrières, S.N. Aslam, R.M. Cooper, D.J. Clarke, The yjbEFGH locus in *Escherichia coli* K-12 is an operon encoding proteins involved in exopolysaccharide production, Microbiology 153 (2006) 1070–1080, https://doi.org/10.1099/mic.0.2006/002907-0.
[75] M. Caboni, T. Pédron, O. Rossi, D. Goulding, D. Pickard, F. Citiulo, C. A. MacLennan, G. Dougan, N.R. Thomson, A. Saul, P.J. Sansonetti, C. Gerke, An O antigen capsule modulates bacterial pathogenesis in *Shigella sonnei*, PLoS Pathog. 11 (2015), e1004749, https://doi.org/10.1371/journal.ppat.1004749.