

RESEARCH ARTICLE

Comparison of O-Antigen Gene Clusters of All O-Serogroups of *Escherichia coli* and Proposal for Adopting a New Nomenclature for O-Typing

Chitrita DebRoy^{1*}, Pina M. Fratamico², Xianghe Yan², GianMarco Baranzoni², Yanhong Liu², David S. Needleman², Robert Tebbs³, Catherine D. O'Connell³, Adam Allred³, Michelle Swimley³, Michael Mwangi¹, Vivek Kapur¹, Juan A. Raygoza Garay¹, Elisabeth L. Roberts¹, Robab Katani¹

1 *E. coli* Reference Center, Department of Veterinary and Biomedical Sciences, The Pennsylvania State University, University Park, Pennsylvania, United States of America, **2** Eastern Regional Research Center, Agricultural Research Service, U.S. Department of Agriculture, Wyndmoor, Pennsylvania, United States of America, **3** Animal Health & Food Safety, Life Sciences Solutions, Thermo Fisher Scientific, Austin, Texas, United States of America

* rcd3@psu.edu



OPEN ACCESS

Citation: DebRoy C, Fratamico PM, Yan X, Baranzoni G, Liu Y, Needleman DS, et al. (2016) 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(1): e0147434. doi:10.1371/journal.pone.0147434

Editor: Dov Joseph Stekel, University of Nottingham, UNITED KINGDOM

Received: July 21, 2015

Accepted: January 3, 2016

Published: January 29, 2016

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: GenBank Accessions are provided in [S1 Table](#).

Funding: Thermo Fisher Scientific partially supported sequencing of the O- antigen gene clusters and purchased some of the reference standard strains. Robert Tebbs and Michelle Swimley contributed on a technical level through generating whole genome sequencing results for some of the strains. Adam Aldred was involved in genome assembly.

Abstract

Escherichia coli strains are classified based on O-antigens that are components of the lipopolysaccharide (LPS) in the cell envelope. O-antigens are important virulence factors, targets of both the innate and adaptive immune system, and play a role in host-pathogen interactions. Because they are highly immunogenic and display antigenic specificity unique for each strain, O-antigens are the biomarkers for designating O-types. Immunologically, 185 O-serogroups and 11 OX-groups exist for classification. Conventional serotyping for O-typing entails agglutination reactions between the O-antigen and antisera generated against each O-group. The procedure is labor intensive, not always accurate, and exhibits equivocal results. In this report, we present the sequences of 71 O-antigen gene clusters (O-AGC) and a comparison of all 196 O- and OX-groups. Many of the designated O-types, applied for classification over several decades, exhibited similar nucleotide sequences of the O-AGCs and cross-reacted serologically. Some O-AGCs carried insertion sequences and others had only a few nucleotide differences between them. Thus, based on these findings, it is proposed that several of the *E. coli* O-groups may be merged. Knowledge of the O-AGC sequences facilitates the development of molecular diagnostic platforms that are rapid, accurate, and reliable that can replace conventional serotyping. Additionally, with the scientific knowledge presented, new frontiers in the discovery of biomarkers, understanding the roles of O-antigens in the innate and adaptive immune system and pathogenesis, the development of glycoconjugate vaccines, and other investigations, can be explored.

Competing Interests: Robert Tebbs, Catherine D. O'Connell, Michelle Swimley and Adam Allred were employees of Thermo Fisher Scientific when this study was performed. All other authors declare they have no competing interests.

Introduction

O-antigens are part of the lipopolysaccharide (LPS) on the outer envelope of *Escherichia coli*. LPS exhibits a tripartite structure, including the lipid A, core oligosaccharide, and the O-polysaccharides or O-antigens. The O-antigen domain is composed of repeating units of one or more sugar residues, exhibiting remarkable diversity in structure. Variation in the combination, position, stereochemistry, and links between these sugars and the presence or absence of non-carbohydrate entities makes them the most variable region in the cell [1, 2]. Since O-antigens that define the serogroups are important virulence factors and targets of both the innate and adaptive immune systems, their roles in both human and veterinary medicine have evoked considerable interest.

A method based on the identification of the combination of three principal cell surface components, the O-antigens, flagellar H-antigens, and capsular K-antigens was developed for subtyping *E. coli* strains. Since few laboratories had capabilities to type K-antigens, serotyping based on O- and H-antigens became the “gold standard” for *E. coli* typing. In the 1940s, Kaufmann [3–5] classified *E. coli* by serological methods, and by 1945 he successfully classified *E. coli* on the basis of the antigenic properties. Ørskov *et al.* [6] presented a comprehensive serotyping system for *E. coli* strains for 164 O-groups, which has been the basis for O-classification for taxonomic and epidemiological studies and for distinguishing strains during outbreaks and for surveillance.

O-groups O1–O187 have been defined, although O-groups O31, O47, O67, O72, O94 and O122 are no longer valid and have been withdrawn [7, 8], and four groups have been divided into subtypes: O18ab/ac, O28ab/ac, O112ab/ac and O125ab/ac, giving a total of 185 O-groups. In addition, there are 11 other OX-groups informally used by several laboratories (including ours), thus making 196 designated O-groups. Serotyping, the standard method for detecting the O-groups, is based on agglutination reactions of the O-antigen and antisera generated against each of the O-types. Serotyping is labor intensive and error-prone due to cross-reactivity between adsorbed O-antigen antisera produced in rabbits. Some strains are non-typeable, and others can be rough or autoagglutinating, making these cultures un-typeable.

Genes required for the biosynthesis of *E. coli* O-antigens are located on the chromosomal O-antigen gene cluster (O-AGC) flanked between a conserved 39-bp JUMPstart sequence (upstream), which is downstream of *galF* (UTP-glucose-1-phosphate uridylyltransferase) and *gnd* (6-phosphogluconate dehydrogenase) [9, 10]. The O-antigen biosynthesis genes in the O-AGC vary considerably for each serogroup. There are three mechanisms known for the processing of the O-antigen that generally consists of 10–25 repeating units of two to seven sugar residues. There is one mechanism that is O-antigen polymerase, Wzy dependent, where individual repeat units of O-polysaccharides are assembled at the cytoplasmic face of the inner membrane and are transported across the membrane by O-antigen flippase, Wzx. Polymerization of new units of polysaccharides occurs in the periplasmic face of the inner membrane by Wzy (O-antigen polymerase) and is typical for heteropolysaccharides. The majority of *E. coli* O-antigens are Wzx/Wzy-dependent. With the ABC-transporter-dependent pathway, typical for homopolymers, the extension of the O-antigen repeat unit occurs entirely on the cytoplasmic face of the inner membrane by glycosyl transferases followed by transport across the membrane by the ABC transporter system [11]. The third system is the synthase-dependent exopolysaccharide secretion system in which the glycosyl transferases are responsible for transport of the polysaccharide across the membrane; this system is not well comprehended. Although, key components of this pathway have recently been identified in *E. coli*, they only appear to function in the transport of specific exopolysaccharides [12].

In the last decade, significant progress has been made in identifying the *E. coli* O-groups by molecular methods, especially for serogroups associated with diseases in humans and animals. The sequences of the O-unit processing genes, the *wzx* (O-antigen flippase) and *wzy* (O-antigen polymerase) are relatively unique for each individual O-type. Therefore, these two genes were targeted for PCR assays and microarrays to identify the *E. coli* O-groups [13–17]. Lin *et al.* [18] combined PCR with the Luminex system to identify ten pathogenic Shiga toxin-producing *E. coli* O-groups. The amplified *wzx* and *wzy* targets were bound to fluorescent microspheres conjugated with complementary DNA probes in the Luminex system. Multiplex assays targeting several O-serogroup genes [15, 19] and virulence genes have been developed [20, 21]. While for Wzy-dependent O-AGCs, the PCR assays targeted the *wzx* and *wzy* genes, *wzm* and *wzt* genes have been targeted for the detection of for ABC transporter-dependent O-AGC, O8, O9, O52 and O101 [16, 22–24]. Microarrays for genosero-typing were designed for detecting O-groups, H-types, and virulence genes that allowed comprehensive typing of *E. coli* strains using the GeneAtlas system from Affymetrix [25, 26]. Other methods such as flow cytometry [27], immunoassays [28, 29] and microarrays using antibodies [30, 31] have also been developed for rapid detection of Shiga toxin-producing *E. coli* O-groups.

The objectives of this study were to compare the nucleotide sequences of all 196 O-AGCs of *E. coli* in conjunction with their serological reactions. The gene sequences of 71 O-AGCs were determined and submitted to GenBank and the comparative genetics of 196 O-AGCs of *E. coli* are presented with suggestions for updating the nomenclature for *E. coli* O-groups. This study may be leveraged to discover biomarkers for developing rapid, convenient, and accurate methods for O-group determination. The sequences could be potentially utilized to study the comparative evolution of O-antigens of bacteria that may occur through gene deletion, acquisition, or inactivation, mechanisms of host adaptation and immune system evasion, expression of virulence, and development of glycoconjugate vaccines for diseases, as well as for other purposes.

Materials and Methods

Bacterial strains and culture conditions

The reference control standard strains that were sequenced are used routinely for O-serotyping at the *E. coli* Reference Center at the Pennsylvania State University [6]. The strains were obtained from Statens Serum Institut (SSI) in Denmark that is affiliated with the World Health Organization Collaborating Centre for Reference and Research on *Escherichia* and *Klebsiella*. The strains are listed in [S1 Table](#). All bacteria were grown in Luria Bertani (LB) broth or on LB agar plates at 37°C.

Genome sequencing, assembly, and annotation

Genomic DNA was isolated using the PureLink Genomic DNA Mini kit (Thermo Fisher Scientific, Inc., Waltham, MA). The concentration of DNA was measured by absorbance readings at 260 nm and 280 nm using the Nanodrop ND100 UV-Vis spectrophotometer (Nanodrop Technologies, Wilmington, DE). DNA libraries for sequencing on the Ion Torrent Personal Genome Machine (PGM) (Thermo Fisher Scientific, Inc.) were prepared following the manufacturer's recommended library construction procedures. Ion Torrent PGM Ion 316 or 318 v2 chips with either the 200-bp or 400-bp OneTouch kits were used for generating sequence data. The *de novo* assembly of whole genomes into the final contigs was performed with CLC Genomics Workbench 7.0 (CLC Bio, Aarhus, Denmark) using the default settings. The published primers complementary to JUMPStart and *gnd* [32] were mapped to the final contigs with a minimum sequence identity of 70% over a window of 20 nucleotides. When necessary joining of O-AGC contigs was performed by using Sanger sequencing and joining of long PCR

amplicons as described [32]. GeneWise [33] was used to predict gene structure and check for frameshifts and sequencing errors. In addition, Prokka 1.10 software [34] in combination with manual annotation was used to finalize the gene structure of the O-AGCs before submission to GenBank. The HMMTOP 2.0 transmembrane topology prediction server [35] was used to identify potential transmembrane helices from the amino acid sequences.

Construction of the phylogenetic tree

A phylogenetic tree of the 196 O-AGCs was generated using the DNA sequences between the JUMPstart and GND primers. Both the alignment and the phylogenetic tree were generated using CLC Genomics Workbench 8.5.1. To create the alignment, the following parameters were selected: Gap open cost = 10.0, Gap extension cost = 1.0, and selecting the very accurate progressive alignment. To create the phylogenetic tree, the Maximum Likelihood Phylogeny tool was selected and analysis was performed under the assumption of the Jukes Cantor substitution model within the software program. The Neighbor Joining construction method was selected. To determine the reliability of the tree, 100 bootstrap replicates were performed.

GenBank accession numbers

All O-antigen cluster sequences were deposited in the NCBI GenBank database and the accession numbers are listed in [S1 Table](#).

Results and Discussion

Structure of the O-AGCs

To characterize the genetic diversity of the O-AGCs, the DNA sequences generated either from the current study or from nucleotide sequences published in GenBank (Accession numbers are listed in [S1 Table](#)), including insertion elements, and other non-coding regions between the JUMPstart and *gnd* regions from 196 O-AGCs were compared using the maximum likelihood phylogenetic tool of the CLC Genomics Workbench. The comparative phylogenetic tree is depicted in [Fig 1](#). Since insertion elements play an important role in the evolution of O-AGCs, these were included to present a more complete comparison of the relationship among the clusters [32, 36]. The number of genes in the O-AGC varied between five (O174) and 18 (O108) and the lengths ranged from 5.6 kb (O174) to 27.7 kb (O55) ([S1 Fig](#)). The genes encoding for the O-antigens belong to three major categories. The nucleotide sugar biosynthesis genes that are involved in the synthesis of O-antigen nucleotide sugar precursors, the glycosyl transferases, that transfer the various sugar precursors to form the oligosaccharide, and the O-antigen processing proteins, the flippase (*Wzx*), O-antigen polymerase (*Wzy*) and polysaccharide ABC transporter, O-antigen ABC transporter permease *Wzm*, and O-antigen ABC transporter ATP-binding protein *Wzt*.

Serogroups, O14 and O57, do not carry O-AGC-related genes between *galF* and *gnd* loci, and therefore, could not be mapped. Serogroup O14 is known to be rough and cannot be serotyped [6], and has been previously reported to lack an O-AGC [37, 38]. Antisera raised against O14:K7 (a rough strain) have been shown to cross-react against *E. coli* and other Enterobacteriaceae due to the presence of the enterobacterial common antigen to which the antisera react [38]. Similarly, other investigators could not locate an O-AGC in O57 [37, 39].

Nucleotide sugar biosynthesis genes

Nucleotide sugar biosynthesis genes exhibit a high level of identity among the different O-groups and often group together in the cluster. A notable number of these genes are conserved

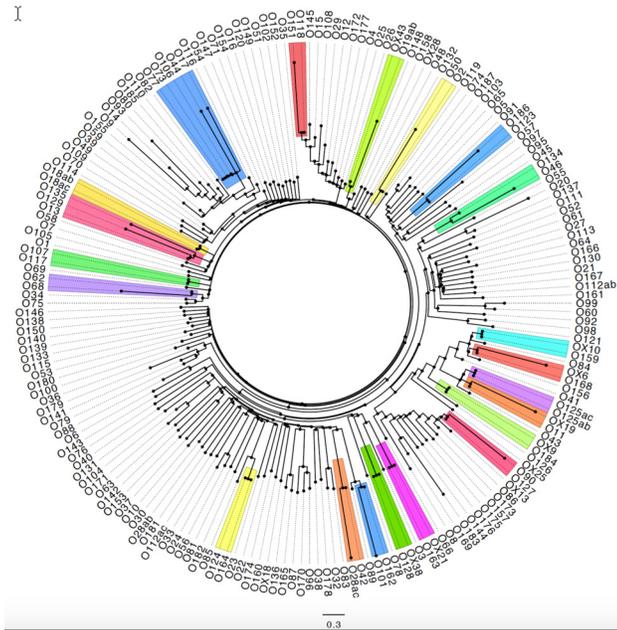


Fig 1. Phylogenetic tree for all O-AGCs of *E. coli*. The O-AGCs that show 98–99.9% relatedness are highlighted.

doi:10.1371/journal.pone.0147434.g001

in various species. There are four genes, *rmlB* (dTDP-glucose 4, 6-dehydratase), *rmlD* (dTDP-4-dehydrorhamnose reductase), *rmlA* (glucose-1-phosphate thymidyltransferase), and *rmlC* (dTDP-4-dehydrorhamnose 3,5-epimerase) that are involved in the biosynthesis of dTDP-L-rhamnose. In 49 O-AGCs, these are grouped as *rmlBDAC* (S1 Fig). In 30 O-AGCs, part of the group is separated or missing. In O2, O50, O54, O62, O71, O109, O119 and O177, *rmlC* is separated from the group due to insertion of other genes between *rmlA* and *rmlC*, and in others, only two of the genes in the group such as *rmlDA* or *rmlBA* are present. The *manB* gene encoding for phosphomannomutase and *manC* encoding for mannose-1-phosphate guanyltransferase responsible for the biosynthesis of GDP-D-mannose [40] are present in 56 O-AGCs. The two genes involved in biosynthesis of UDP-L-FucNAc derived from UDP-GlcNAc, *fnlA* (UDP-glucose epimerase) and *fnlC* (UDP-N-acetylglucosamine 2-epimerase), were identified in 15 O-AGCs. *VioA* and *VioB* that carry out transamination of dTDP-6-deoxy-D-xylo-4-hexulose to dTDP-4-amino-4,6-dideoxy-D-glucose (*VioN*) and *VioB* that N-acetylates *VioN* to dTDP-*VioNAc* were found to be associated with the O-AGC for O39, O49, and O116.

Glycosyl transferases

Glycosyl transferases are responsible for adding sugar residues to the O-antigens during their synthesis. Numerous combinations of an extensive range of sugars are present in O-antigens, with specific linkages among them. Therefore, heterogeneous groups of highly specific glycosyl transferases are associated with the O-AGCs. These were identified based on sequence similarities to other sugar transferases that are found within the O-AGCs.

O-antigen processing genes

The O-antigen processing genes, *wzx* (flippase) and *wzy* (polymerase), are highly specific for each O-group and are present in most of the O-AGCs. The O-antigen is synthesized when a glycosyl-1-phosphoryl residue is transferred to an undecaprenyl phosphate acceptor to form

an undecaprenyl-PP-sugar intermediate. Transfer of additional sugar units to this undecaprenyl results in an undecaprenyl-PP-oligosaccharide intermediate to which repeating sugar units are sequentially transferred, and are then translocated and flipped across the membrane by Wzx [40]. Both Wzx and Wzy are hydrophobic proteins with transmembrane helices, and they show high variation in sequence. These genes are involved in the synthesis and translocation of O-antigens using the Wzy-dependent pathway. The O-AGCs of 185 O-groups carry the O-antigen flippase (*wzx*) and O-antigen polymerase (*wzy*) genes, as confirmed in our analyses. Eleven O-AGCs: O8, O9, O52, O60, O89, O92, O95, O97, O99, O101 and O162 are ABC transporter-dependent for O-antigen processing and carry *wzm* and *wzt* that assist in the transport process. The mechanisms of O-antigen biosynthesis in O8 and O9, that have capsules, have been extensively studied [11]. Although O8 has *wzx* and *wzy* genes in the O-AGC, the genes in the cluster are directed to form a capsule and the O-polysaccharides are transported using an ATP-binding ABC-transporter process [11, 24, 40, 41]. The O-AGCs of O89, O101 and O162 are notably identical as discussed later and therefore, there are nine unique O-AGCs that are ABC transporter-dependent.

Relatedness among O-AGCs

Analysis of the phylogenetic relatedness among 196 O-AGCs demonstrated that twenty sets of O-groups were 98–99.9% identical in their nucleotide sequences as highlighted in Fig 1. Diagrammatic representations of the genes representing these 21 sets of identical O-AGCs, are presented in Fig 2. These are O2/O50, O13/O129/O135, O17/O44/O73/O77/O106, O42/O28ac, O46/O134, O62/O68, O90/O127, O101/O162, O107/O117, O118/O151, O123/O186, O124/O164, O125ab/O125ac, OX6/O168, OX9/O184, OX10/O159, OX19/O11, OX21/O163, OX38/O128, OX43/O19, O118ab/O118ac. The O-AGCs of O62 and O68 differ due to the presence of an insertion element located within the third codon from the end of the *rmlA* gene in O62 otherwise, they are almost identical [32] (Fig 1). As mentioned above, insertion elements play a role in the evolution of O-AGCs.

The comparative serological cross-reactivity data for these sets of identical O-groups, as observed and recorded for the last 50 years of serotyping at the *E. coli* Reference Center are listed in Table 1. Although the nucleotide sequences may be identical in certain O-groups, the serological reactions with rabbit antisera may not show any cross-reactivity as observed for strains belonging to O2/O50, O46/O134, O118/O151, and OX19/O11. This could be due to post-translational modification of proteins that may be responsible for the epitopes in antigens. Recently Joensen *et al.* [42] presented information on cross-reactions of the O-groups that have 98–100% identical *wzx* and *wzy* genes. Although there are some differences in cross-reactions they observed between identical O-groups that are different from ours, some are similar. For example, serogroups O107 and O117 and serogroups O123 and O186, show serological cross-reactivity in both studies; however, Joensen and co-workers [42] stated that serogroups O2 and O50 cross-reacted serologically, while in the current study no cross-reaction was observed. Cross-reactions between O-groups vary considerably, and may depend on the polyclonal antisera generated in different rabbits. Further research may elucidate the mechanism of antigen-antibody reactions for these O-groups. Some of the O-groups such as O90 sometimes cross-react with O127 but not vice versa; O101 may sometimes cross-react with O162, but O162 does not cross-react with O101, and the reason for this is unclear. Many of the genetically similar O-groups that are related do cross-react as shown in Table 1. Strains that react serologically with O17 antisera were sometimes found to cross-react with antisera generated against O73, O77 and O106 but never with O44, suggesting that the epitopes for the immunologic reactions may vary based on the whole genome composition. O-AGCs of O118 and O151

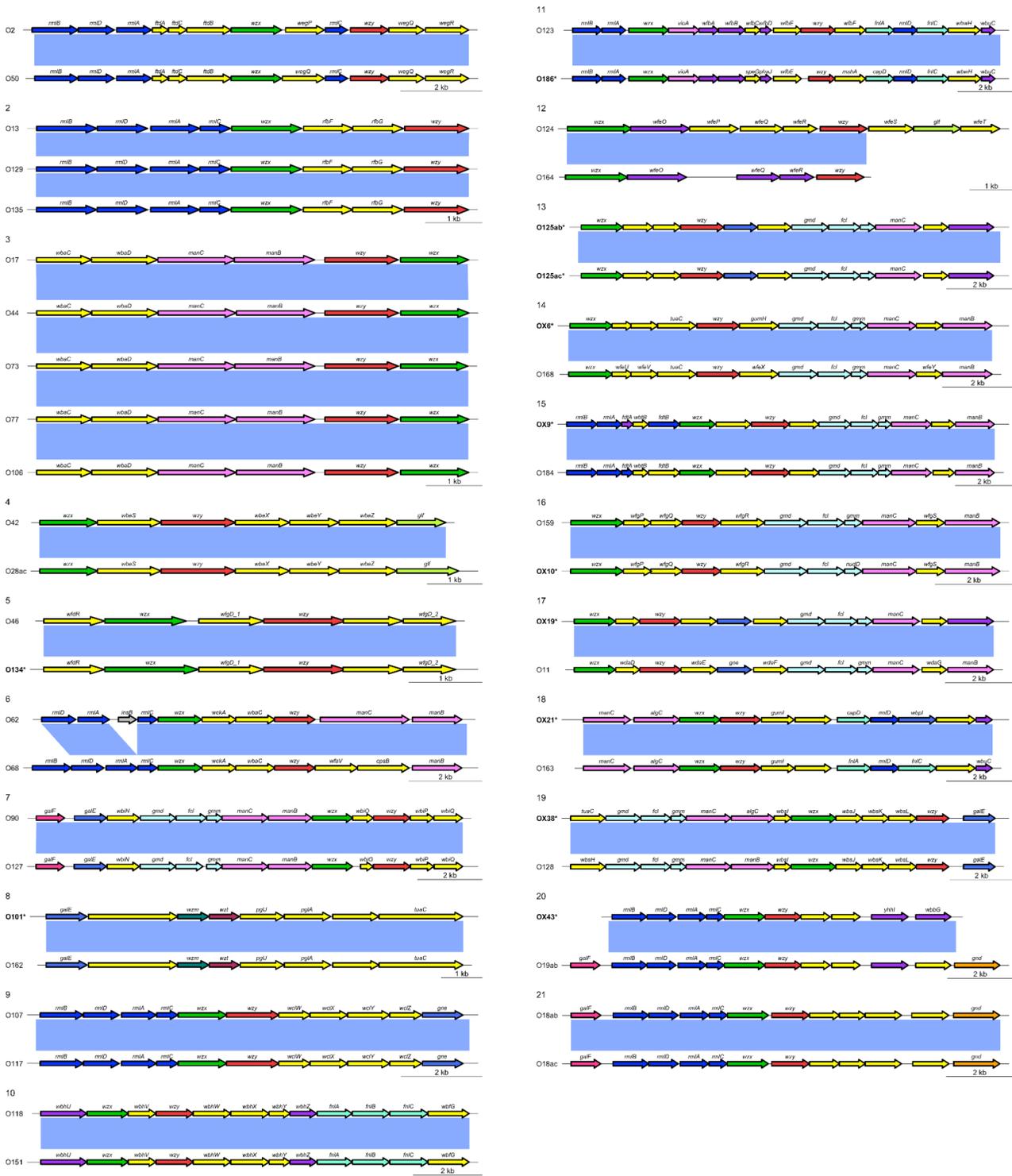


Fig 2. Comparison of identical O-AGCs.

doi:10.1371/journal.pone.0147434.g002

Table 1. Comparison of O-AGC of O-groups that are 98–99.9% identical.

Set #	Computational analysis	Serological cross-reaction	Comparison Status of sequences	Availability ^a
1	O2/O50	No	<i>in silico</i>	O2 (n = 2650)/ O50 (n = 46)
2	O13/O129/O135	Yes	<i>in silico</i>	O13 (n = 280)/ O129 (n = 42)/ O135 (n = 45)
3	O17/O44 /O73/ O77/ O106	Yes	<i>in silico</i>	O17 (n = 322)/ O44 (n = 109)/ O73 (n = 604)/ O77 (n = 253)/ O106 (n = 229)
4	O42/O28ac	Insufficient data	<i>in silico</i>	O42 (n = 41)/O28 (n = 42)
5	O46/O134	No	This study	O46 (n = 60)/O134 (n = 28)
6	O62/O68	O68 cross-reacts with O62 but not vice versa	Liu Y et al. 2015	O62 (n = 2)/O68 (n = 127)
7	O90/O127	O90 cross-reacts with O127 but not vice versa	<i>in silico</i>	O90 (n = 25)/O127 (n = 51)
8	O101 /O162	O101 cross-react with O162 but not vice versa	<i>in silico</i>	O101 (n = 734)/O162 (n = 98)
9	O107/O117	Yes	<i>in silico</i>	O107 (n = 59)/O117 (n = 239)
10	O118/ O151	No	<i>in silico</i>	O118 (n = 182)/O151 (n = 61)
11	O123/O186	Yes	<i>in silico</i>	O123 (n = 45)/O186 (n = 4)
12	O124/ O164	Occasional cross-reaction	<i>in silico</i>	O124 (n = 91)/O164 (n = 11)
13	O125ab/O125ac	Insufficient data	This study	O125 (n = 54)
14	OX6/ O168	Yes	This study	OX6 (n = 46)/O168 ((n = 99)
15	OX9/O184	Insufficient data	This study	OX9 (n = 124)/O184 (n = 1)
16	OX10/O159	Yes	This study	OX10 (n = 81)/O159 (n = 161)
17	OX19/O11	No	This study	OX19 (n = 28)/O11 (n = 527)
18	OX21/O163	Yes	This study	OX21 (n = 26)/O163 (n = 137)
19	OX38/O128	Insufficient data	This study	OX38 (n = 15)/O128 (n = 329)
20	OX43/O19	Yes	This study	OX43 (n = 40)/O19 (n = 385)
21	O18ab/O18ac	Insufficient Data	<i>in silico</i>	O18 (n = 777)

a, Number of cultures in database of *E. coli* Reference Center (ECRC) (1967–2015)

doi:10.1371/journal.pone.0147434.t001

exhibit identical nucleotide sequences, except O151 carries substitutions in two nucleotides thereby altering two amino acids in the proteins that are translated [43], they do not cross-react serologically.

Iguchi *et al.* [37] assigned O-AGCs of all 184 O-groups of *E. coli* into 16 groups based on similarities in nucleotide sequences. Most of the groups they describe match with our results except for O153 and O137 [37]. No significant similarities in nucleotides sequences were observed for O153/O178 in the current study (Fig 1, S1 Fig). PCR assays developed targeting *wzx* and *wzy* genes from GenBank sequences submitted in this report (KJ755551) for O153 were highly specific for clinical isolates belonging to serogroup O153 (Fig 3). Therefore, grouping O-AGCs O153 and O178 based on 99.9% identity may not be accurate [37]. Similarly, O-group O137, reported to be 99.7% identical to O20 [37], was not corroborated in the present investigation. The sequence of O137 (KJ755548) generated in the current study matches 100% with the nucleotide sequence published earlier for this O-group [44] (GenBank accession number GU068043) and is not identical to O20 (S1 Fig). While O89, O101, and O162 were grouped based on nucleotide similarities [37], our data show that O89 shares 96.6% identity over 66.6% coverage to O101 and O162, which are 99% identical over 100% coverage as determined by BLASTn [45] (Fig 1). Serogroup O89 is also serologically distinct from O101 and O162. Therefore, we believe O89 to be a distinct O-group. The O-AGCs of O169 and O183 were found to be 97% identical over 64% coverage (Fig 1), and thus are only partially similar. No serological

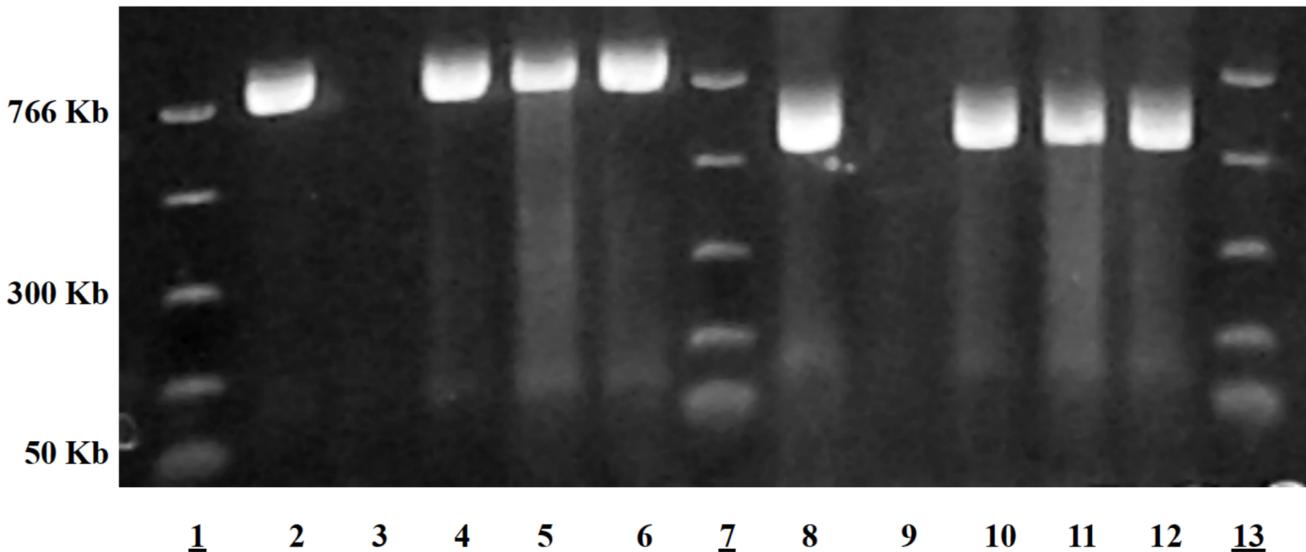


Fig 3. O153 wzx and wzy genes amplified by PCR using primers from sequences presented in this investigation. Lanes 1, 7, 13: Molecular weight markers. Lane 2: Positive control for O153 targeting wzx gene, Lane 3: Negative control, Lanes 4,5,6: wzx amplified for three clinical isolates. Lane 9: Positive control for O153 targeting wzy gene. Lane 10: Negative control, Lane 11,12,13: wzy amplified for three clinical isolates.

doi:10.1371/journal.pone.0147434.g003

cross-reaction between O169 and O183 was observed, and therefore, they could be considered as distinct O-groups.

The sequencing data generated will assist in developing platforms for molecular genotyping of *E. coli*. In order to develop the scheme, there is a need to consider merging or eliminating the designations of O-groups that have identical O-AGCs. Since O-AGCs for O14 and O57 could not be identified in their genomes, it will be difficult to designate these O-groups until target genes that may potentially be involved in the synthesis of O-AGCs are identified for these O-groups. Whole genome sequencing and gene expression studies with knock-out mutants for rough strains may also elucidate the complexity involving O-antigen synthesis for O14. It should be considered that the serogroups that are similar in nucleotide sequence and cross-react serologically (Table 2) may be merged to eliminate redundancy. O125ab/O125ac may be designated O125 and O18ab/O18ac may be merged as O18 as these O-groups have

Table 2. O-groups that may be potentially merged based on similarities in O-AGC nucleotide sequence and serological cross-reactions.

Set #	Identity 98–99.9%	Suggested O group
1	O42/O28ac	O42
2	O13/O129/O135	O13
3	O107/O117	O107
4	O123/O186	O123
5	O125ab/O125ac	O125
6	O18ab/O18ac	O18
7	OX6/O168	O168
8	OX10/O159	O159
9	OX21/O163	O163
10	OX38/O128	O128
11	OX43/O19	O19

doi:10.1371/journal.pone.0147434.t002

been found to be identical [37]. O19ab can be designated as O19. The carbohydrate structures of O13/O129/O135 have been found to be similar and related to *Shigella flexneri* [46]. The sequence of serogroup O13 is 99% identical with 100% coverage to O129 it is 99% identical with 82% coverage for O135. Strains belonging to these serogroups cross-react serologically; therefore, O13 may be merged with O129 and O135 and the merged O-groups designated as O13. O28ac and O42 are identical except for three point mutations exhibited in O42 in the *wbeX* and *wbeY* genes [21], and these serogroups cross-react. Thus, these may be merged and designated as O42. O107 and O117 may be merged as O107. O17/O44/O73/O77/O106 have identical nucleotide sequences and share a common four-sugar backbone O-subunit structure with each other and *Salmonella enterica* serogroup O:6,14 (H) [47]. All of these O-groups except *E. coli* O77 O-antigen, have substitutions of one or two glucose side branches at various positions in the O-unit backbone and cross-react with each other except for O44. Three genes were identified in the *E. coli* O44 genome within a putative prophage that are presumably involved in the glucosylation of the basic tetrasaccharide unit [47]. This may be the reason why O44 strains never serologically cross-react with the others in the group (O17/O73/O77/O106). Since the antigenic specificities for these O-groups are quite distinct, further investigations need to be conducted to determine if these O-groups can be merged. However, for genosero-typing assays these O-groups may not be distinguishable.

Eight O-groups have been previously designated as OX1-OX8 by Ewing *et al* [48]. OX1 is now designated as O170, OX2 as O169, and OX3 as O174. OX4 and OX6 were found to be similar to O146 and O171, respectively, OX5 is now designated as O168, and OX7 as O175 [8, 49]. In this investigation, many of the OX-groups were found to be identical to established groups and may be eliminated. OX6 can be designated as O171, OX10 as O159, OX21 as O163, OX38 as O128, and OX43 as O19. The other OX groups, including OX13, OX18, OX25, OX28, and OX38, were found to have unique O-AGCs. Although additional studies are needed, we propose that these OX-groups may be designated as new O-groups chronologically following the designation of the Statens Serum Institut that have now listed O188 serogroups. (<http://www.ssi.dk/English/SSI%20Diagnostica/Products%20from%20SSI%20Diagnostica/Bacterial%20strains/E%20coli.aspx>). It is likely that more O-groups will be discovered, as the nucleotide sequences of the large number of non-typeable strains may exhibit unique sequences that cannot be designated as any of the established O-groups [50]. We may be able to assign O-groups to non-typeable strains based on the genosero-typing as they may exhibit SNPs or mutations in the O-AGCs hampering the serological reaction, resulting in their designation as non-serotype-able [51]. Whole genome sequencing of the strains may reveal factors responsible for synthesis of the antigenic domains of the O-antigens.

Based on the nucleotide sequences of the O-AGCs, genosero-typing can be achieved by targeting the unique sequences for each O-group. While *wzx* and *wzy* are suitable targets for most of the O-groups, and among the O-groups that do not carry *wzx* and *wzy* genes, unique regions within the *wzm* and *wzt* genes could be utilized for detecting O-groups O8, O9, O52, O60, O89/O101/O162, O92, O95, and O97. Joensen *et al.* [42] recently presented serotyping based on *in silico* whole genome sequences. The publicly available web tool, SerotypeFinder hosted by the Center for Genomic Epidemiology (www.genomicsepitidemiology.org) is available for O-genosero-typing. The O-antigen genes *wzx*, *wzy*, *wzm*, and *wzt* and flagellin genes can be detected easily based on sequence data, and thus, this tool can be an alternate faster and cheaper method than serotyping. Other methods are also likely to develop from the information presented that may lead to more accurate and rapid O-typing of *E. coli*.

Supporting Information

S1 Fig. Structure of O-AGC of all 196 O-serogroups. The O-AGCs of all 196 O- and OX-groups are diagrammatically represented. The nucleotide sequences of 71 O-groups marked with asterisk and in bold font were determined in the present investigation. (PDF)

S1 Table. Strains and GenBank accession numbers for O-AGCs of all known *E. coli* O-groups. (DOCX)

Author Contributions

Conceived and designed the experiments: CD PMF XY GMB YL DSN. Performed the experiments: XY GMB YL ELR. Analyzed the data: CD PMF XY GMB YL DSN MM VK JARG ELR RK. Contributed reagents/materials/analysis tools: RT CDO AA MS JARG MM. Wrote the paper: CD PMF YL DSN XY GMB JARG RK.

References

1. 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):627–653. doi: [10.1111/j.1574-6976.2008.00114.x](https://doi.org/10.1111/j.1574-6976.2008.00114.x) PMID: [18422615](https://pubmed.ncbi.nlm.nih.gov/18422615/).
2. Stenutz R, Weintraub A, Widmalm G. The structures of *Escherichia coli* O-polysaccharide antigens. *FEMS Microbiol Rev.* 2006; 30(3):382–403. doi: [10.1111/j.1574-6976.2006.00016.x](https://doi.org/10.1111/j.1574-6976.2006.00016.x) PMID: [16594963](https://pubmed.ncbi.nlm.nih.gov/16594963/).
3. Kaufmann F. Ueber neue tthermolabile Ko'rperantigen der Colibakterien. *Acta Pathol Microbiol Scand.* 1943; 20:21–44.
4. Kaufmann F. Zur Serologie der Coli-Gruppe. *Acta Pathol Microbiol Scand.* 1944; 21:20–45.
5. Kaufmann F. The serology of the *Coli* group. *J. Immunol.* 1947; 57:71–100. PMID: [20264689](https://pubmed.ncbi.nlm.nih.gov/20264689/)
6. Orskov I, Orskov F, Jann B, Jann K. Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. *Bacteriol Rev.* 1977; 41(3):667–710. PMID: [334154](https://pubmed.ncbi.nlm.nih.gov/334154/); PubMed Central PMCID: PMC414020.
7. Orskov F, Orskov I. Serotyping of *Escherichia coli*. *Methods Microbiol.* 1984; 14:43–112.
8. Scheutz F, Cheasty T, Woodward D, Smith HR. Designation of O174 and O175 to temporary O groups OX3 and OX7, and six new *E. coli* O groups that include Verocytotoxin-producing *E. coli* (VTEC): O176, O177, O178, O179, O180 and O181. *APMIS.* 2004; 112(9):569–584. doi: [10.1111/j.1600-0463.2004.apm1120903.x](https://doi.org/10.1111/j.1600-0463.2004.apm1120903.x) PMID: [15601305](https://pubmed.ncbi.nlm.nih.gov/15601305/).
9. Hobbs M, Reeves PR. The JUMPstart sequence: a 39 bp element common to several polysaccharide gene clusters. *Mol Microbiol.* 1994; 12(5):855–856. PMID: [8052136](https://pubmed.ncbi.nlm.nih.gov/8052136/).
10. Wang L, Reeves PR. Organization of *Escherichia coli* O157 O antigen gene cluster and identification of its specific genes. *Infect Immun.* 1998; 66(8):3545–3551. PMID: [9673232](https://pubmed.ncbi.nlm.nih.gov/9673232/); PubMed Central PMCID: PMC108385.
11. Greenfield LK, Whitfield C. Synthesis of lipopolysaccharide O-antigens by ABC transporter-dependent pathways. *Carbohydr Res.* 2012; 356:12–24. doi: [10.1016/j.carres.2012.02.027](https://doi.org/10.1016/j.carres.2012.02.027) PMID: [22475157](https://pubmed.ncbi.nlm.nih.gov/22475157/).
12. Whitney JC and Howell PL. Synthase-dependent exopolysaccharide secretion in Gram-negative bacteria. *Trends Microbiol.* 2013; 21(2):63–72. doi: [10.1016/j.tim.2012.10.001](https://doi.org/10.1016/j.tim.2012.10.001) PMID: [23117123](https://pubmed.ncbi.nlm.nih.gov/23117123/); PubMed Central PMCID: PMC4113494.
13. Bugarel M, Beutin L, Martin A, Gill A, Fach P. Micro-array for the identification of Shiga toxin-producing *Escherichia coli* (STEC) seropathotypes associated with hemorrhagic colitis and hemolytic uremic syndrome in humans. *Int J Food Microbiol.* 2010; 142(3):318–329. doi: [10.1016/j.ijfoodmicro.2010.07.010](https://doi.org/10.1016/j.ijfoodmicro.2010.07.010) PMID: [20675003](https://pubmed.ncbi.nlm.nih.gov/20675003/).
14. DebRoy C, Roberts E, Fratamico PM. Detection of O antigens in *Escherichia coli*. *Anim Health Res Rev.* 2011; 12(2):169–185. doi: [10.1017/S1466252311000193](https://doi.org/10.1017/S1466252311000193) PMID: [22152292](https://pubmed.ncbi.nlm.nih.gov/22152292/).
15. DebRoy C, Roberts E, Valadez AM, Dudley EG, Cutter CN. Detection of Shiga toxin-producing *Escherichia coli* O26, O45, O103, O111, O113, O121, O145, and O157 serogroups by multiplex polymerase chain reaction of the *wzx* gene of the O-antigen gene cluster. *Foodborne Pathog Dis.* 2011; 8(5):651–652. doi: [10.1089/fpd.2010.0769](https://doi.org/10.1089/fpd.2010.0769) PMID: [21548768](https://pubmed.ncbi.nlm.nih.gov/21548768/).

16. Han W, Liu B, Cao B, Beutin L, Kruger U, Liu H, et al. DNA microarray-based identification of serogroups and virulence gene patterns of *Escherichia coli* isolates associated with porcine postweaning diarrhea and edema disease. *Appl Environ Microbiol*. 2007; 73(12):4082–4088. doi: [10.1128/AEM.01820-06](https://doi.org/10.1128/AEM.01820-06) PMID: [17449692](https://pubmed.ncbi.nlm.nih.gov/17449692/); PubMed Central PMCID: PMC1932722.
17. Liu Y, Fratamico P. *Escherichia coli* O antigen typing using DNA microarrays. *Mol Cell Probes*. 2006; 20(3–4):239–244. doi: [10.1016/j.mcp.2006.01.001](https://doi.org/10.1016/j.mcp.2006.01.001) PMID: [16537102](https://pubmed.ncbi.nlm.nih.gov/16537102/).
18. Lin A, Nguyen L, Lee T, Clotilde LM, Kase JA, Son I, et al. Rapid O serogroup identification of the ten most clinically relevant STECs by Luminex microbead-based suspension array. *J Microbiol Methods*. 2011; 87(1):105–110. doi: [10.1016/j.mimet.2011.07.019](https://doi.org/10.1016/j.mimet.2011.07.019) PMID: [21835211](https://pubmed.ncbi.nlm.nih.gov/21835211/).
19. Sanchez S, Llorente MT, Echeita MA, Herrera-Leon S. Development of three multiplex PCR assays targeting the 21 most clinically relevant serogroups associated with Shiga toxin-producing *E. coli* infection in humans. *PLOS ONE*. 2015; 10(1):e0117660. doi: [10.1371/journal.pone.0117660](https://doi.org/10.1371/journal.pone.0117660) PMID: [25629697](https://pubmed.ncbi.nlm.nih.gov/25629697/); PubMed Central PMCID: PMC4309606.
20. Fratamico PM, DebRoy C, Miyamoto T, Liu Y. PCR detection of enterohemorrhagic *Escherichia coli* O145 in food by targeting genes in the *E. coli* O145 O-antigen gene cluster and the Shiga toxin 1 and Shiga toxin 2 genes. *Foodborne Pathog Dis*. 2009; 6(5):605–611. doi: [10.1089/fpd.2008.0254](https://doi.org/10.1089/fpd.2008.0254) PMID: [19435408](https://pubmed.ncbi.nlm.nih.gov/19435408/).
21. Fratamico PM, Yan X, Liu Y, DebRoy C, Byrne B, Monaghan A, et al. *Escherichia coli* serogroup O2 and O28ac O-antigen gene cluster sequences and detection of pathogenic *E. coli* O2 and O28ac by PCR. *Can J Microbiol*. 2010; 56(4):308–316. doi: [10.1139/w10-010](https://doi.org/10.1139/w10-010) PMID: [20453897](https://pubmed.ncbi.nlm.nih.gov/20453897/).
22. Feng L, Senchenkova SN, Yang J, Shashkov AS, Tao J, Guo H, et al. Synthesis of the heteropolysaccharide O antigen of *Escherichia coli* O52 requires an ABC transporter: structural and genetic evidence. *J Bacteriol*. 2004; 186(14):4510–4519. doi: [10.1128/JB.186.14.4510-4519.2004](https://doi.org/10.1128/JB.186.14.4510-4519.2004) PMID: [15231783](https://pubmed.ncbi.nlm.nih.gov/15231783/); PubMed Central PMCID: PMC438562.
23. Liu B, Wu F, Li D, Beutin L, Chen M, Cao B, et al. Development of a serogroup-specific DNA microarray for identification of *Escherichia coli* strains associated with bovine septicemia and diarrhea. *Vet Microbiol*. 2010; 142(3–4):373–378. doi: [10.1016/j.vetmic.2009.10.019](https://doi.org/10.1016/j.vetmic.2009.10.019) PMID: [19932572](https://pubmed.ncbi.nlm.nih.gov/19932572/).
24. Wang L, Briggs CE, Rothmund D, Fratamico P, Luchansky JB, Reeves PR. Sequence of the *E. coli* O104 antigen gene cluster and identification of O104 specific genes. *Gene*. 2001; 270(1–2):231–236. PMID: [11404020](https://pubmed.ncbi.nlm.nih.gov/11404020/).
25. Geue L, Monecke S, Engelmann I, Braun S, Slickers P, Ehrlich R. Rapid microarray-based DNA genotyping of *Escherichia coli*. *Microbiol Immunol*. 2014; 58(2):77–86. doi: [10.1111/1348-0421.12120](https://doi.org/10.1111/1348-0421.12120) PMID: [24298918](https://pubmed.ncbi.nlm.nih.gov/24298918/).
26. Lacher DW, Gangiredla J, Jackson SA, Elkins CA, Feng PC. Novel microarray design for molecular serotyping of Shiga toxin-producing *Escherichia coli* strains isolated from fresh produce. *Appl Environ Microbiol*. 2014; 80(15):4677–4682. doi: [10.1128/AEM.01049-14](https://doi.org/10.1128/AEM.01049-14) PMID: [24837388](https://pubmed.ncbi.nlm.nih.gov/24837388/); PubMed Central PMCID: PMC4148803.
27. Hegde NV, Jayarao BM, DebRoy C. Rapid detection of the top six non-O157 Shiga toxin-producing *Escherichia coli* O groups in ground beef by flow cytometry. *J Clin Microbiol*. 2012; 50(6):2137–2139. doi: [10.1128/JCM.00137-12](https://doi.org/10.1128/JCM.00137-12) PMID: [22493328](https://pubmed.ncbi.nlm.nih.gov/22493328/); PubMed Central PMCID: PMC3372132.
28. Hegde NV, Cote R, Jayarao BM, Muldoon M, Lindpaintner K, Kapur V, et al. Detection of the top six non-O157 Shiga toxin-producing *Escherichia coli* O groups by ELISA. *Foodborne Pathog Dis*. 2012; 9(11):1044–1048. doi: [10.1089/fpd.2012.1231](https://doi.org/10.1089/fpd.2012.1231) PMID: [23134286](https://pubmed.ncbi.nlm.nih.gov/23134286/).
29. Medina MB, Shelver WL, Fratamico PM, Fortis L, Tillman G, Narang N, et al. Latex agglutination assays for detection of non-O157 Shiga toxin-producing *Escherichia coli* serogroups O26, O45, O103, O111, O121, and O145. *J Food Prot*. 2012; 75(5):819–826. doi: [10.4315/0362-028X.JFP-11-430](https://doi.org/10.4315/0362-028X.JFP-11-430) PMID: [22564929](https://pubmed.ncbi.nlm.nih.gov/22564929/).
30. Gehring A, Barnett C, Chu T, DebRoy C, D'Souza D, Eaker S, et al. A high-throughput antibody-based microarray typing platform. *Sensors*. 2013; 13(5):5737–5748. doi: [10.3390/s130505737](https://doi.org/10.3390/s130505737) PMID: [23645110](https://pubmed.ncbi.nlm.nih.gov/23645110/); PubMed Central PMCID: PMC3690026.
31. Hegde NV, Praul C, Gehring A, Fratamico P, DebRoy C. Rapid O serogroup identification of the six clinically relevant Shiga toxin-producing *Escherichia coli* by antibody microarray. *J Microbiol Methods*. 2013; 93(3):273–276. doi: [10.1016/j.mimet.2013.03.024](https://doi.org/10.1016/j.mimet.2013.03.024) PMID: [23570904](https://pubmed.ncbi.nlm.nih.gov/23570904/).
32. Liu Y, Yan X, DebRoy C, Fratamico PM, Needleman DS, Li RW, et al. *Escherichia coli* O-antigen gene clusters of serogroups O62, O68, O131, O140, O142, and O163: DNA sequences and similarity between O62 and O68, and PCR-based serogrouping. *Biosensors (Basel)*. 2015; 5(1):51–68. doi: [10.3390/bios5010051](https://doi.org/10.3390/bios5010051) PMID: [25664526](https://pubmed.ncbi.nlm.nih.gov/25664526/); PubMed Central PMCID: PMC4384082.
33. Birney E, Clamp M, Durbin R. GeneWise and Genomewise. *Genome Res*. 2004; 14(5):988–995. doi: [10.1101/gr.1865504](https://doi.org/10.1101/gr.1865504) PMID: [15123596](https://pubmed.ncbi.nlm.nih.gov/15123596/); PubMed Central PMCID: PMC479130.

34. Seemann T, Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014; 30(14):2068–2069. doi: [10.1093/bioinformatics/btu153](https://doi.org/10.1093/bioinformatics/btu153) PMID: [24642063](https://pubmed.ncbi.nlm.nih.gov/24642063/).
35. Tusnady GE, Simon I. The HMMTOP transmembrane topology prediction server. *Bioinformatics*. 2001; 17(9):849–850. PMID: [11590105](https://pubmed.ncbi.nlm.nih.gov/11590105/).
36. Cheng J, Wang Q, Wang W, Wang Y, Wang L, Feng L. Characterization of *E. coli* O24 and O56 O antigen gene clusters reveals a complex evolutionary history of the O24 gene cluster. *Curr Microbiol*. 2006; 53(6):470–476. doi: [10.1007/s00284-006-0032-7](https://doi.org/10.1007/s00284-006-0032-7) PMID: [17072668](https://pubmed.ncbi.nlm.nih.gov/17072668/).
37. Iguchi A, Iyoda S, Kikuchi T, Ogura Y, Katsura K, Ohnishi M, et al. A complete view of the genetic diversity of the *Escherichia coli* O-antigen biosynthesis gene cluster. *DNA Res*. 2015; 22(1):101–107. doi: [10.1093/dnares/dsu043](https://doi.org/10.1093/dnares/dsu043) PMID: [25428893](https://pubmed.ncbi.nlm.nih.gov/25428893/); PubMed Central PMCID: [PMC4379981](https://pubmed.ncbi.nlm.nih.gov/PMC4379981/).
38. Jensen SO, Reeves PR. Deletion of the *Escherichia coli* O14:K7 O antigen gene cluster. *Can J Microbiol*. 2004; 50(4):299–302. doi: [10.1139/w04-010](https://doi.org/10.1139/w04-010) PMID: [15213754](https://pubmed.ncbi.nlm.nih.gov/15213754/).
39. Coimbra RS, Grimont F, Lenormand P, Burguiere P, Beutin L, Grimont PA. Identification of *Escherichia coli* O-serogroups by restriction of the amplified O-antigen gene cluster (*rfb*-RFLP). *Res Microbiol*. 2000; 151(8):639–654. PMID: [11081579](https://pubmed.ncbi.nlm.nih.gov/11081579/).
40. Samuel G, Reeves P. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydrate research*. 2003; 338(23):2503–19. PMID: [14670712](https://pubmed.ncbi.nlm.nih.gov/14670712/).
41. Bronner D, Clarke BR, Whitfield C. Identification of an ATP-binding cassette transport system required for translocation of lipopolysaccharide O-antigen side-chains across the cytoplasmic membrane of *Klebsiella pneumoniae* serotype O1. *Mol Microbiol*. 1994; 14(3):505–519. PMID: [7533882](https://pubmed.ncbi.nlm.nih.gov/7533882/).
42. Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F. Rapid and easy in silico serotyping of *Escherichia coli* using whole genome sequencing (WGS) data. *J Clin Microbiol*. 2015; 53(8):2410–2426. doi: [10.1128/JCM.00008-15](https://doi.org/10.1128/JCM.00008-15) PMID: [25972421](https://pubmed.ncbi.nlm.nih.gov/25972421/).
43. Liu Y, Fratamico P, Debroy C, Bumbaugh AC, Allen JW. DNA sequencing and identification of serogroup-specific genes in the *Escherichia coli* O118 O antigen gene cluster and demonstration of antigenic diversity but only minor variation in DNA sequence of the O antigen clusters of *E. coli* O118 and O151. *Foodborne Pathog Dis*. 2008; 5(4):449–457. doi: [10.1089/fpd.2008.0096](https://doi.org/10.1089/fpd.2008.0096) PMID: [18673069](https://pubmed.ncbi.nlm.nih.gov/18673069/).
44. Wang Q, Ruan X, Wei D, Hu Z, Wu L, Yu T, et al. Development of a serogroup-specific multiplex PCR assay to detect a set of *Escherichia coli* serogroups based on the identification of their O-antigen gene clusters. *Mol Cell Probes*. 2010; 24(5):286–290. doi: [10.1016/j.mcp.2010.06.002](https://doi.org/10.1016/j.mcp.2010.06.002) PMID: [20561581](https://pubmed.ncbi.nlm.nih.gov/20561581/).
45. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990; 215(3):403–410. doi: [10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: [2231712](https://pubmed.ncbi.nlm.nih.gov/2231712/).
46. Perepelov AV, Shevelev SD, Liu B, Senchenkova SN, Shashkov AS, Feng L, et al. Structures of the O-antigens of *Escherichia coli* O13, O129, and O135 related to the O-antigens of *Shigella flexneri*. *Carbohydr Res*. 2010; 345(11):1594–1599. doi: [10.1016/j.carres.2010.04.023](https://doi.org/10.1016/j.carres.2010.04.023) PMID: [20546712](https://pubmed.ncbi.nlm.nih.gov/20546712/).
47. Wang W, Perepelov AV, Feng L, Shevelev SD, Wang Q, Senchenkova SN, et al. A group of *Escherichia coli* and *Salmonella enterica* O antigens sharing a common backbone structure. *Microbiology*. 2007; 153(Pt 7):2159–2167. doi: [10.1099/mic.0.2007/004192-0](https://doi.org/10.1099/mic.0.2007/004192-0) PMID: [17600060](https://pubmed.ncbi.nlm.nih.gov/17600060/).
48. Ewing WH, Tatum HW, US Communicable Disease Center. Studies on the serology of the *Escherichia coli* group. 1956. <http://www.worldcat.org/title/studies-on-the-serology-of-the-escherichia-coli-group/oclc/654668490>
49. Ewing W. The Genus *Escherichia*. EPAE WH, editor. Minneapolis, Minnesota: Burgess Publishing Co.; 1972. 67–107 p.
50. Duda KA, Lindner B, Brade H, Leimbach A, Brzuszkiewicz E, Dobrindt U, et al. The lipopolysaccharide of the mastitis isolate *Escherichia coli* strain 1303 comprises a novel O-antigen and the rare K-12 core type. *Microbiology*. 2011; 157(Pt 6):1750–1760. doi: [10.1099/mic.0.046912-0](https://doi.org/10.1099/mic.0.046912-0) PMID: [21372091](https://pubmed.ncbi.nlm.nih.gov/21372091/).
51. DebRoy C, Roberts E, Davis M, Bumbaugh A. multiplex polymerase chain reaction assay for detection of nonserotypable Shiga toxin-producing *Escherichia coli* strains of serogroup O147. *Foodborne Pathog Dis*. 2010; 7(11):1407–1414. doi: [10.1089/fpd.2010.0614](https://doi.org/10.1089/fpd.2010.0614) PMID: [20617939](https://pubmed.ncbi.nlm.nih.gov/20617939/).