DNA Research, 2015, 22(1), 101–107
doi: 10.1093/dnares/dsu043
Advance Access Publication Date: 26 November 2014
Full Paper

OXFORD

Full Paper

A complete view of the genetic diversity of the *Escherichia coli* O-antigen biosynthesis gene cluster

Atsushi Iguchi[1,*], Sunao Iyoda[2], Taisei Kikuchi[3], Yoshitoshi Ogura[4,5], Keisuke Katsura[5], Makoto Ohnishi[2], Tetsuya Hayashi[4,5], and Nicholas R. Thomson[6,7]

[1]Department of Animal and Grassland Sciences, Faculty of Agriculture, University of Miyazaki, Miyazaki 889-2192, Japan, [2]Department of Bacteriology I, National Institute of Infectious Diseases, Tokyo 162-8640, Japan, [3]Division of Parasitology, Department of Infectious Diseases, Faculty of Medicine, University of Miyazaki, Miyazaki 889-1692, Japan, [4]Division of Microbiology, Department of Infectious Diseases, Faculty of Medicine, University of Miyazaki, Miyazaki 889-1692, Japan, [5]Division of Bioenvironmental Science, Frontier Science Research Center, University of Miyazaki, Miyazaki 889-1692, Japan, [6]Pathogen Genomics, The Wellcome Trust Sanger Institute, Cambridge CB10 1SA, UK, and [7]Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK

*To whom correspondence should be addressed. Tel/Fax. +81 985-58-7507. E-mail: iguchi@med.miyazaki-u.ac.jp

Edited by Dr Katsumi Isono

Received 16 September 2014; Accepted 3 November 2014

Abstract

The O antigen constitutes the outermost part of the lipopolysaccharide layer in Gram-negative bacteria. The chemical composition and structure of the O antigen show high levels of variation even within a single species revealing itself as serological diversity. Here, we present a complete sequence set for the O-antigen biosynthesis gene clusters (O-AGCs) from all 184 recognized *Escherichia coli* O serogroups. By comparing these sequences, we identified 161 well-defined O-AGCs. Based on the *wzx*/*wzy* or *wzm*/*wzt* gene sequences, in addition to 145 singletons, 37 serogroups were placed into 16 groups. Furthermore, phylogenetic analysis of all the *E. coli* O-serogroup reference strains revealed that the nearly one-quarter of the 184 serogroups were found in the ST10 lineage, which may have a unique genetic background allowing a more successful exchange of O-AGCs. Our data provide a complete view of the genetic diversity of O-AGCs in *E. coli* showing a stronger association between host phylogenetic lineage and O-serogroup diversification than previously recognized. These data will be a valuable basis for developing a systematic molecular O-typing scheme that will allow traditional typing approaches to be linked to genomic exploration of *E. coli* diversity.

Key words: *E. coli*, O-antigen biosynthesis gene cluster, horizontal gene transfer, O serogroup, genomic diversity

1. Introduction

Cell-surface polysaccharides play an essential role in the ability of bacteria to survive and persist in the environment and in host organisms.[1] The O-antigen polysaccharide constitutes the outermost part of the lipopolysaccharide (LPS) present in the outer membrane of Gram-negative bacteria. The chemical composition and structure of the O-antigen exhibit high levels of variation even within a single species.[2–5] This observation is corroborated by the huge serological

variation of somatic O antigens. Currently, the O serogrouping, sometimes combined with H (flagellar) antigens and K (capsular polysaccharide) antigens, is a standard method for subtyping of *Escherichia coli* strains in taxonomical and epidemiological studies. In particular, identification of strains of the same O serogroup is a prerequisite to start any actions for outbreak investigations and surveillance.

Thus far, the World Health Organization Collaborating Centre for Reference and Research on *Escherichia* and *Klebsiella* based at the Statens Serum Institut (SSI) in Denmark (http://www.ssi.dk/English.aspx) has recognized 184 *E. coli* O serogroups. It is generally believed that the O serogrouping of *E. coli* strains provides valuable information for identifying pathogenic clonal groups, especially for public health surveillance. For example, O157 is a leading O serogroup associated with enterohemorrhagic *E. coli* (EHEC) and is a significant food-borne pathogen worldwide.[6,7] Other important EHEC O serogroups include O26, O103, and O111.[8] The Shiga toxin-producing *E. coli* O104:H4 was found responsible for a large human food-borne disease outbreak in Europe, 2011.[9] Another notable example is strains of serogroup O25; extended-spectrum beta lactamase (ESBL)-producing, multidrug-resistant *E. coli* O25:H4 has emerged worldwide to cause a wide variety of community and nosocomial infections.[10]

In *E. coli*, the genes required for O-antigen biosynthesis are clustered at a chromosomal locus flanked by the colanic acid biosynthesis gene cluster (*wca* genes) and the histidine biosynthesis (*his*) operon. Generally, the O-antigen biosynthesis genes fall into three classes: (i) the nucleotide sugar biosynthesis genes, (ii) the sugar transferase genes, and (iii) those for O-unit translocation and chain synthesis (*wzx*/*wzy* in the Wzx/Wzy-dependent pathway and *wzm*/*wzt* in the Wzm/Wzt-dependent ABC transporter pathway).[11] To date, >90 types of O-antigen biosynthesis gene cluster (O-AGC) sequences have been determined, with the majority derived from major human and animal pathogens.[12] Sequence comparisons of these O-AGCs indicate a great variety of genetic structures. Several studies have provided evidence to show that horizontal transfer and replacement of a part or all of the O-AGC have caused shifts in O serogroups.[13–15] Alternatively, point mutations in the glycosyltransferase genes in the O-AGC or acquisition of alternative O-antigen modification genes, which are located outside of the O-AGC, have also been shown to result in structural alterations of O antigen and concomitant change in the serotype of the isolate.[16,17]

Genes or DNA sequences specific for each O serogroup can be used as targets for the identification of O serogroups via molecular approaches, such as PCR-based and hybridization-based methods. Such systems have already been developed by several researchers to target specific O-antigen types.[12,18–20] In particular, molecular assays targeting major O serogroups are routinely used in EHEC surveillance for clinical or food sample screening. Considering the range of diseases caused by *E. coli* strains belonging to many different serogroups, a more comprehensive and detailed O-AGC information for the complete set of *E. coli* O serogroups is of significant clinical importance for generating a rational molecular typing scheme. This molecular typing scheme, which could be performed *in silico* directly on sequence data, also offers a mechanism with which to link the ever-expanding genomic data to our extensive epidemiological and biological knowledge of this pathogen, based on O-antigen typing. Moreover, these data will also provide a much better understanding of the complex mechanisms by which a huge diversity in O serogroups have arisen. Here, we present a complete sequence set for the O-AGCs from all 184 *E. coli* O serogroups, which include recently added serogroups (O182–O187), providing a complete picture of the O-AGC diversity in *E. coli*.

2. Materials and methods

2.1. Bacterial strains, culture condition, and DNA preparation

Reference strains of all 184 recognized *E. coli* O serogroups were obtained from SSI (see Supplementary Table S1). Cells were grown to the stationary phase at 37°C in Luria–Bertani medium. Genomic DNA was purified using the Wizard Genomic DNA purification kit (Promega) according to the manufacturer's instructions.

2.2. O-AGC sequences and comparative analyses

One hundred and eight *E. coli* O-AGC sequences were determined by Sanger-based capillary sequencing and/or Illumina MiSeq sequencing from PCR products covering O-AGCs (Supplementary Table S1). The O-AGC regions of the reference strains were amplified by PCR using 10 ng of genomic DNA as template with the Tks Gflex DNA polymerase (Takara Bio Inc.) by 25 amplification cycles for 10 s at 98°C and for 16 m at 69°C, and with a combination of three forward primers (TATGCCAGCGGCACCAAACG, ATACCGGCGATGAAAGCC, and GCGGGTGGGATTAAGTCTCT) designed on the *hisFI* genes and two reverse primers (GTGATGCAGGAATCCTCTGT and CCACGCTAATTACGCCATCTT) designed on the *wcaM* genes, or strain-specific primers designed based on the draft genome sequences determined using the MiSeq system from reference strains. Identification and functional annotation of the CDSs were performed based on the results of homology searches against the public, non-redundant protein database using BLASTP. The sequences reported in this article have been deposited in the GenBank database (accession no. AB811596-AB811624, AB812020-AB812085, and AB972413-AB972425). The other 76 *E. coli* O-AGC sequences were obtained from public databases. For a list of accession numbers, see Supplementary Table S1.

2.3. Phylogenetic analysis

Multilocus sequence typing (MLST) was carried out according to the protocol described on the *E. coli* MLST website (http://mlst.warwick.ac.uk/mlst/dbs/Ecoli), and the phylogenetic relationships of reference strains were analysed based on the concatenated sequences (3,423 bp) of seven housekeeping genes (*adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA*) used for MLST. Multiple alignments of DNA and amino acid sequences were constructed by using the CLUSTAL W program.[21] Phylogenetic trees were constructed by using the neighbour-joining algorithm using the MEGA4 softrware.[22]

3. Results

3.1. Genetic structures of the O-AGCs from all *E. coli* O serogroups

Of the 184 known O serogroups, 76 complete O-AGC sequences were obtained from public databases. The sequence of the other 108 O-AGC was determined in this study from *E. coli* O-serogroup reference strains (Supplementary Table S1). Our analysis of these sequences confirmed several previously observed characteristics of O-AGCs in *E. coli* (Supplementary Fig. S1). In brief, O-AGCs are located between the *wca* and *his* operons. This region contains three housekeeping genes: *galF* (encoding UTP-glucose-1-phosphate uridylyltransferase), *gnd* (6-phosphogluconate dehydrogenase), and *ugd* (UDP-glucose 6-dehydrogenase), and most genes for O-antigen biosynthesis in each cluster are directly flanked by *galF* and *gnd*/*ugd*, while *gne* (UDP-GalNAc-4-epimerase) and *wzz* (O-antigen chain

length determination protein) located immediately outside of the region between *galF* and *gnd/ugd* (see Supplementary Fig. S1). The exceptions for this are the O-AGCs for O serogroups O14 and O57, which contain no O-antigen genes at the typical locus. However, it is known that the *E. coli* O14 reference strain Su4411-41 shows an O rough phenotype and lacks the O-AGC.[23] For O57, a further analysis is also required to investigate the presence of O-antigen structure in the LPS of the reference strain. Our data revealed that the O-AGCs located between *galF* and *gnd* ranged in size from 4.5 kbp (O155, including four genes) to 19.5 kbp (O108, including 18 genes).

3.1.1. Nucleotide sugar biosynthesis genes

Genes required for the deoxythymidine diphosphate (dTDP)-sugar biosynthesis pathway (*rmlBDAC)* to synthesize dTDP-L-rhamnose (dTDP-L-Rha), the precursor of L-Rha, were widely distributed in the O-AGCs (conserved in 56 O-serogroup O-AGCs; see Supplementary Fig. S1). The *vioAB* operon, for the biosynthesis of dTDP-*N*-acetylviosamine (dTDP-VioNAc), the precursor of VioNAc, was present in three O-serogroup O-AGCs; the *fnlABC* operon for the synthesis of uridine diphosphate (UDP)-*N*-acetyl-L-fucosamine (UDP-L-FucNAc), the precursor of L-FucNAc, was in 11 O-serogroup O-AGCs; the *fnlA-qnlBC* genes for the synthesis of UDP-*N*-acetyl-L-quinovosamine (UDP-L-QuiNAc), the precursor of L-QuiNAc, were in four O-serogroup O-AGCs; the *nnaDBCA* genes for synthesis of cytidine monophosphate (CMP)-*N*-acetylneuraminate (CMP–NeuNAc), the precursor of *N*-acetylneuraminic acid (Neu5Ac or sialic acid), were found in six O-serogroup O-AGCs (Supplementary Fig. S1). In addition, a gene set comprising seven genes putatively involved in the synthesis of di-*N*-acetyl-8-epilegionaminic acid (8eLeg5Ac7Ac) were found in three O-serogroup O-AGCs. For at least 49 O serogroups, gene sets for nucleotide sugar biosynthesis were not found in their O-AGCs (Supplementary Fig. S1), suggesting that, in these serogroups, nucleotide sugars required for O-antigen biosynthesis were synthesized by pathways encoded by the genes located outside of the O-AGCs.

3.1.2. Glycosyltransferase

Each O-AGC contained two to six genes encoding putative glycosyltransferases for synthesizing O-antigen subunits and a total of 611 glycosyltransferase genes identified in all O-AGCs. Pfam analysis revealed that at least 25 types of glycosyltransferase-related domains were found in the 611 glycosyltransferase genes (Supplementary Table S2). 'Glycosyl transferases group 1' (PF00534) and 'Glycosyl transferase family 2' (PF00535) were the most widely distributed domains, which were found in 216 and 253 genes, respectively. Except for the five genes belonging to 'Glycosyltransferase family 52' (PF07922), which were found in five of the six *nnaDBCA*-containing O-AGCs (O24, O56, O104, O131, and O171), there were no relationships between the type of glycosyltransferase-related domain and the gene set for sugar synthesis in each O-AGC.

3.1.3. O-antigen subunit translocation and chain synthesis

All O-AGCs carried either *wzx/wzy* or *wzm/wzt* gene pairs. Of the 182 O-AGCs (the above-mentioned O14 and O57 were excluded from the 184 clusters analysed in this study), 171 carried the *wzx/wzy* genes, and the other 11 carried the *wzm/wzt* genes (Supplementary Fig. S1 and Table S1). Detailed sequence comparisons of the *wzx/wzy* and *wzm/wzt* genes are described below.

3.2. Grouping the O-AGCs by sequence

On the basis of sequences and genetic structures of the entire O-AGC regions, in addition to 145 unique O-AGCs from different *E. coli* O serogroups, the O-AGCs from 37 O serogroups could be placed into 16 groups (named Gp1–Gp16) with the members of each group having identical or very similar O-AGC genes (mostly sharing ≥95% DNA sequence identity) (Fig. 1). This included nine groups with members of different serogroups but which carried identical O-AGC gene sets (Gp1–Gp9) and one group, Gp10, where two strains (O13 and O129) of the three-member group carried an identical O-AGC gene set (sharing 98.3–99.9% DNA sequence identity) (Fig. 1). The reason(s) why they belong to different O serogroups even though they have identical O-AGCs are discussed in the Discussion section. Indels or exchange of one or more genes was also shown to explain the differences between O135 and other members of Gp10 and members Gp11–Gp16, which otherwise carried highly conserved orthologous genes (summarized in Fig. 1). Simple insertions of insertion sequence (IS) elements containing one or two transposase genes were found in three groups without any gene disruption: an IS*629* insertion in O18ab of Gp12, IS*Ec11* in O164 of Gp13, and IS*1* in O62 of Gp14. IS element-associated replacement of the right-end portion of the O-AGC had occurred in three groups, Gp14, Gp15, and Gp16, resulting in the replacement (or deletion) of glycosyltransferase gene(s). Exchange of the *wzx* gene had also occurred in Gp16. These data suggest that IS elements are important drivers for generating O-antigen biosynthesis gene replacement and therefore diversity.

3.3. Diversity and specificity of the *wzx/wzy* or *wzm/wzt* genes among the *E. coli* O-AGCs

As previously proposed,[12] most *wzx/wzy* or *wzm/wzt* orthologues showed high levels of sequence diversity and their sequences were unique to each O-AGC or O-AGC group described above (Fig. 2 and Supplementary Fig. S2). DNA sequence identities of the closest pairs were <70%, except for the O96/O170 pair, the *wzx* genes of which showed 86% DNA sequence identity. Within the 16 O-AGC groups, the orthologous *wzx/wzy* or *wzm/wzt* genes also showed high sequence conservation (≥95% DNA sequence identity, but mostly ≥97% identity), except for Gp16 that shared only the *wzy* gene (Fig. 2).

3.4. Phylogenetic relationships of *E. coli* O-serogroup reference strains

Based on the concatenated nucleotide sequences of seven housekeeping genes used for MLST, we determined the evolutionary relationships of all *E. coli* O-serogroup reference strains (Fig. 3). This analysis revealed that the members of five groups sharing the common O-AGCs (Gp8, Gp10, Gp11, Gp14, and Gp15) and two members (O17 and O77) of Gp9 were found in closely related lineages. However, the members of other groups (and three members of Gp9) were found in distinct evolutionary lineages. For example, O20 and O137, both carrying the Gp1 O-AGC, were found in two distinct lineages, each belonging to phylogroups A and E/D, respectively, and five serogroups (O17/O77, O44, O73, and O106) belonging to Gp9 were found in multiple lineages (A, E/D, and B1).

The systematic phylogenetic analysis of all *E. coli* O-serogroup reference strains further revealed that one-quarter of the reference strains (46/184) belonged to a single clonal group (≥99.9% sequence identity), which was represented by sequence type (ST) 10 and its very close relatives in phylogroup A (Fig. 3 and Supplementary Fig. S3). Additionally, three clonal groups containing five or more reference

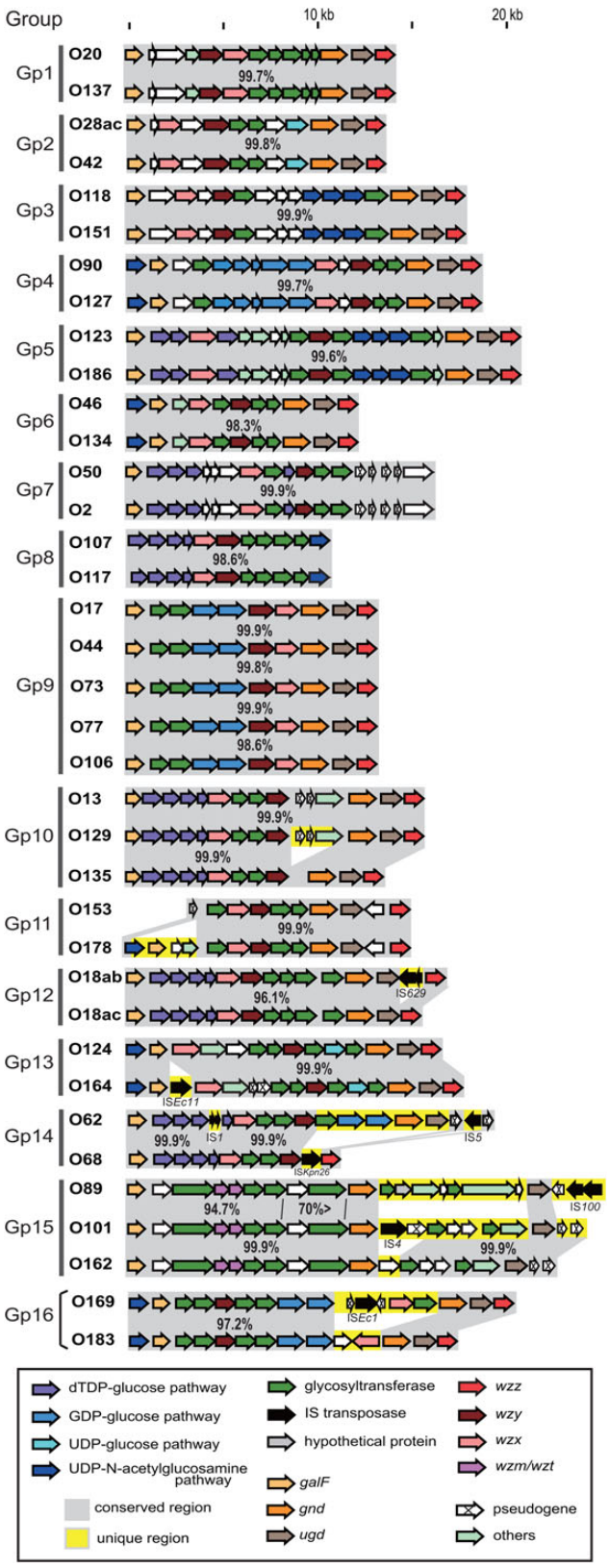

Figure 1. Sixteen *Escherichia coli* O-AGC groups identified in this study. Group members have different O serogroups in each group, but these share nearly identical or highly similar genetic organizations. Group names (Gp) are indicated at the left side. DNA sequence identities (%) between group members are indicated in each group.

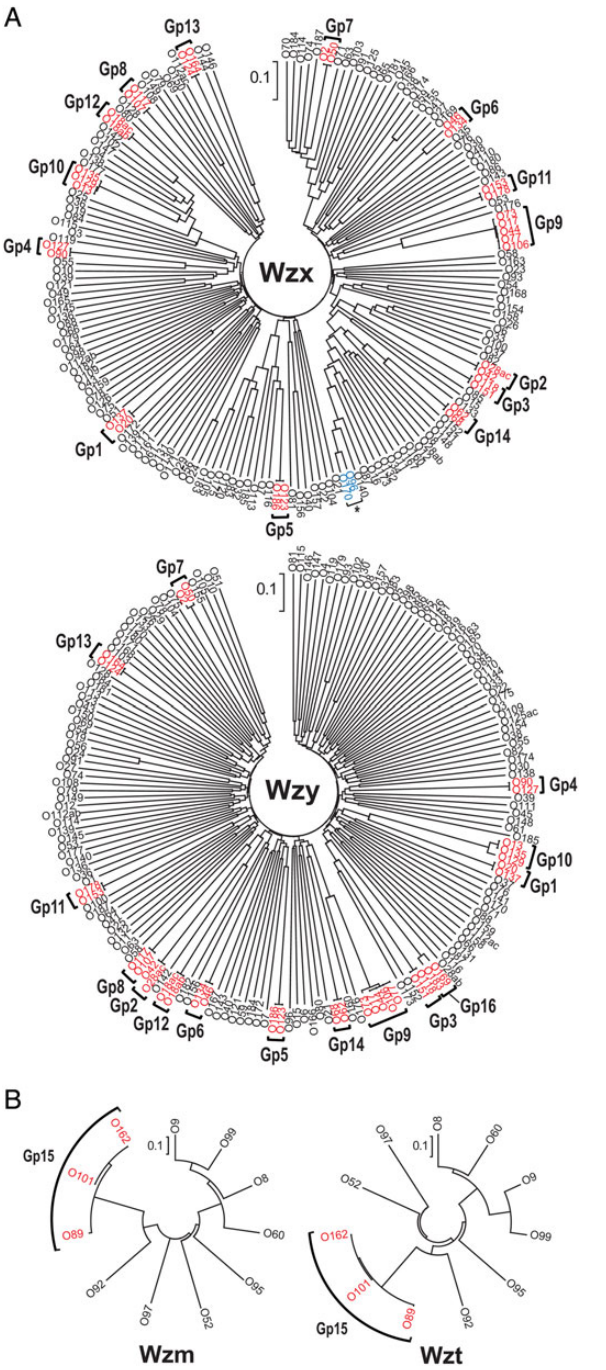

Figure 2. Phylogenetic analysis of homologues of (A) Wzx and Wzy and (B) Wzm and Wzt from *Escherichia coli* O-serogroup reference strains based on the amino acid sequences. The group names are indicated outside of trees. The pair or groups of homologues with high DNA sequence identity (≥95%, mostly ≥97%) are indicated in red. The Wzx homologues of O96 and O170, which are indicated in blue and by an asterisk, showed 86% DNA sequence identity, but in all other proteins showed low-sequence homologies to each other (<70% identity). Note that while the DNA sequence identity between the *wzx*_O46 and *wzx*_O134 in Gp6 is 99.7%, the *wzx*_O46 has a 2-bp deletion at the 3′-region, causing a frame shift.

strains were also identified in phylogroups A (ST34 and ST57) and B1 (ST300) (Fig. 3). The phylogenetic analysis also showed that the types of sugar synthesis gene sets and processing gene sets (*wzx*/*wzy* and *wzm*/*wzt*) were not limited to a specific lineage (Fig. 3).

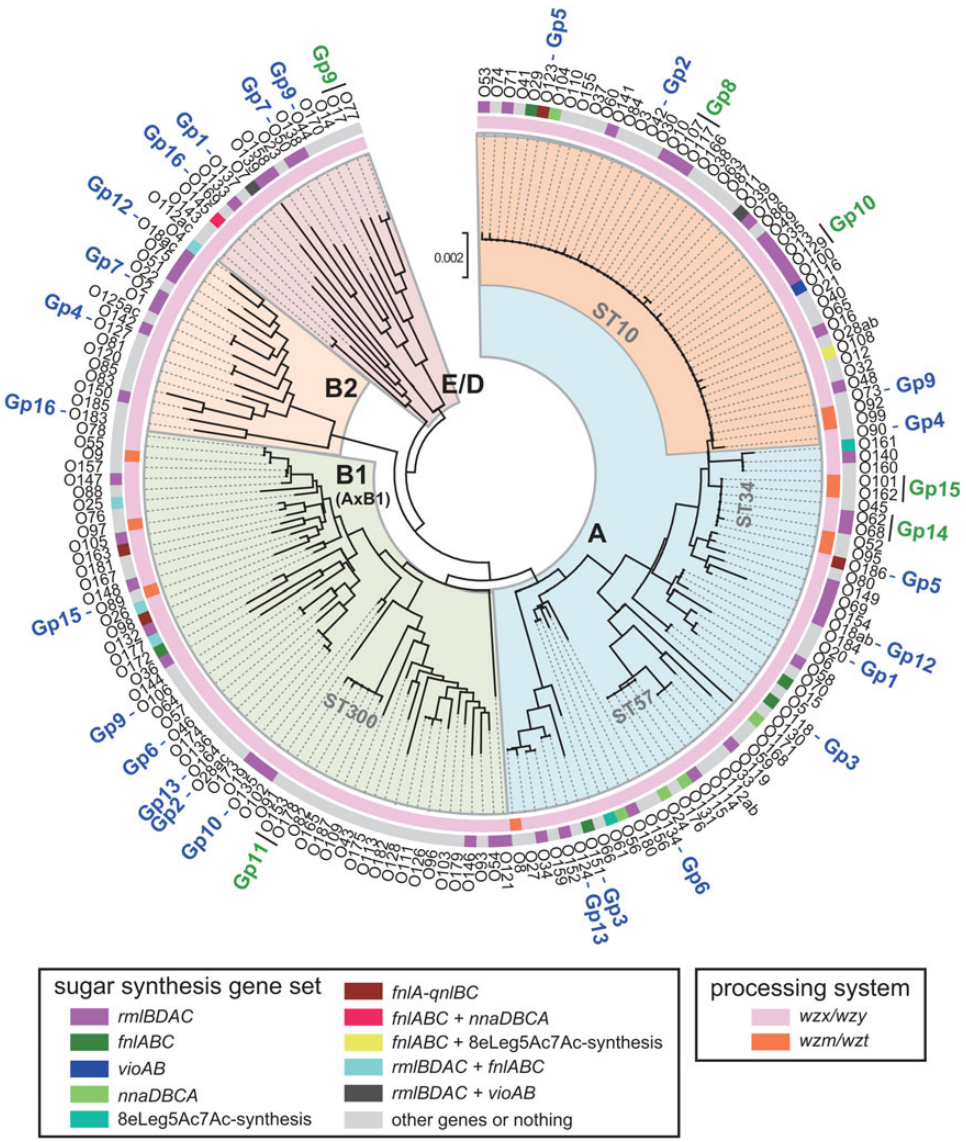

Figure 3. Correlation between the *Escherichia coli* evolutionary lineages and the distribution of O-AGCs. The phylogenetic tree was constructed based on the concatenated sequences of seven housekeeping genes from all 184 *E. coli* O-serogroup reference strains. The group names of O-AGCs (Gp1–Gp16) are indicated in the outermost region. Members in groups indicated in green were found to belong to the same or very closely related lineage, whereas members of the groups indicated in blue were found in distinct lineages. The outer circle next to the O serogroup names indicates the distribution of sugar synthesis gene sets identified in each O-AGC. The inner circle indicates the type of O-antigen processing system (*wzx*/*wzy* or *wzm*/*wzt*). Phylogenetic groups (A, B1, B2, D, and E) were determined by comparing the sequences of the strains tested with the known sequences from the ECOR collection (http://mlst.warwick.ac.uk/mlst/dbs/Ecoli).

3.5. Relationships of the *E. coli* and *Shigella* O-AGCs

Shigella and *E. coli* belong to the same species complex[24] and many *Shigella* O antigens are known to be serologically and genetically identical or very similar to some *E. coli* O antigens, as summarized by Liu *et al.*[25] In addition to the 21 previously shown relationships, we found two additional O-AGC groups shared by *E. coli* and *Shigella*; O38 and *Shigella dysenteria* type 8 (SD8), and O169/O183 and *Shigella boydii* type 6/10 (SB6/SB10) (Supplementary Fig. S4). The O183-AGC was highly similar to the *S. boydii* types 10 cluster (sharing 98.2% DNA sequence identity). In our previous study,[26] we provisionally named a novel O serogroup for a group of Shiga toxin-producing *E. coli* strains as OSB10, which cross-reacted with *S. boydii* type 10. Sequence comparisons in this study revealed that OSB10 is not only serologically but also genetically identical to the new serogroup O183 of Gp16.

4. Discussion

Much of what we know about *E. coli* is defined at some level by O serogroups. To link genomic information to the wealth of data held in public databases, in our collective knowledge, outbreak, and disease reports and elsewhere, we endeavoured to determine whether molecular O-serogroup identification, targeting O-serogroup-specific genes (or unique sequences), was a valuable method to capture this information and maintain this important link. Not only do we show evidence supporting the effectiveness of molecular O-typing, but also we open

up the possibility of generating a molecular O-typing scheme and relate O serogroups to the underlying phylogeny of this bacterium.

By determining and comparing the sequences of O-AGCs from all known *E. coli* O serogroups, we newly defined the sequence and gene content of 145 unique O-AGCs and showed that O-AGCs from 37 O serogroups could be placed into 16 groups based on members in each group sharing nearly identical or highly similar O-AGCs. It is clear from these data that many of the grouped O-AGCs (Gp1-16) were found in distinct phylogenetic lineages indicating that these O-AGCs have been spread across this species by horizontal gene transfer. Moreover, several lineages that contained multiple O serogroups, ST10, ST34, ST57, and ST300, show that frequent exchange occurs between and within lineages. ST10 and its close relatives are particularly interesting as one-quarter of *E. coli* O-serogroup reference strains fell within this clonal group. ST10 and its clonal complex are clinically very important being recently found to include ESBL-producing *E. coli* from human and animals in Spain,[25] Italy and Denmark,[26] China,[27] and the Netherlands,[28] and in various intra-intestinal pathotypes of *E. coli*, such as enteroaggregative *E. coli*,[27,28] enterotoxigenic *E. coli*,[29,30] and EHEC.[31,32] In most cases, the O serogroups of these ST10 or ST10-related strains are unusual compared with the typical O serogroups that represent that pathotype.

Acquisition of O-antigen modification genes located on the genomes of serotype-converting bacteriophages or plasmids is also an important strategy for diversifying O-antigen structures. This mechanism has been well investigated in *Shigella flexneri*.[33,34] In *E. coli*, the O-serogroup conversion by a prophage-like element has been reported for O17 and O44,[17] which belong to Gp9 defined in this study. Another possible mechanism to generate the variation of O antigens is the mutations in the genes of the O-AGC as observed for O107 and O117,[16] which belong to Gp8. In this case, point mutations in a glycosyltransferase gene are responsible for the alteration of O-antigen structure (and thus that of O serogroup).[16] Five O-AGC groups including Gp2, Gp5, Gp7, Gp12, and Gp13 also contained differences in the amino acid sequence of their glycosyltransferases. O serogroup differences in these groups may be generated by the point mutations in glycosyltransferase genes. On the other hand, all glycosyltransferase genes in Gp1, Gp3, Gp4, Gp6, and Gp11; four strains from Gp9 (O17, O44, O73, and O77) and two from Gp10 (O13 and O129) showed 100% amino acid sequence identity. These results suggest that the serological differences between the members of these seven groups have been generated by acquisition of modification genes outside of the O-AGC as shown for O17 and O44 of Gp9.[17]

We believe that the remarkable sequence diversity observed in the *wzx/wzy* and *wzm/wzt* O-AGC genes of all known *E. coli* O serogroups appears to be sufficiently discriminative from one another to make identification of each of the known O serogroups possible. Therefore, our sequence data will serve as a valuable resource for the development of rationally designed molecular methods for O-typing as well as for detecting novel O serogroups.

In conclusion, our study provides a complete sequence set of O-AGCs of all known *E. coli* O serogroups and thus offers a full view on the genetic diversity of O-AGCs of this bacterium. In addition, the results presented suggest that horizontal gene transfer has been involved in the O serogroup diversification in *E. coli* more frequently and in a more biased or lineage-dependent fashion than previously thought.

Acknowledgements

We thank A. Akiyoshi, Y. Kato, and A. Yoshida for technical assistance.

Supplementary data

Supplementary data are available at www.dnaresearch.oxfordjournals.org.

Funding

This work was supported by Health Labor Sciences Research Grants from the Ministry of Health, Labor, and Welfare, Japan to A.I. (H25-Syokuhin-Wakate-018) and M.O. (H24-Shinkou-Ippan-012); Adaptable and Seamless Technology Transfer Program through Target-driven R&D (AS242Z00217P) from Japan Science and Technology Agency to A.I.; and a Scientific Research Grant on Priority Areas from the University of Miyazaki and the Program to Disseminate Tenure Tracking System from the Japanese Ministry of Education, Culture, Sports, Science, and Technology to A.I. (http://www.miyazaki-u.ac.jp/ir/english/index.html). This work was also supported by Wellcome Trust grant (098051). Funding to pay the Open Access publication charges for this article was provided by the University of Miyazaki, Japan.

References

1. Bazaka, K., Crawford, R.J., Nazarenko, E.L. and Ivanova, E.P. 2011, Bacterial extracellular polysaccharides, *Adv. Exp. Med. Biol.*, **715**, 213–26.
2. Liu, B., Knirel, Y.A., Feng, L., et al. 2013, Structural diversity in *Salmonella* O antigens and its genetic basis, *FEMS Microbiol. Rev.*, **38**, 56–89.
3. Stenutz, R., Weintraub, A. and Widmalm, G. 2006, The structures of *Escherichia coli* O-polysaccharide antigens, *FEMS Microbiol. Rev.*, **30**, 382–403.
4. Lam, J.S., Taylor, V.L., Islam, S.T., Hao, Y. and Kocincova, D. 2011, Genetic and functional diversity of *Pseudomonas aeruginosa* lipopolysaccharide, *Front Microbiol.*, **2**, 118.
5. Penner, J.L. and Aspinall, G.O. 1997, Diversity of lipopolysaccharide structures in *Campylobacter jejuni*, *J. Infect. Dis.*, **176** (Suppl. 2), S135–138.
6. Armstrong, G.L., Hollingsworth, J. and Morris, J.G. Jr. 1996, Emerging foodborne pathogens: *Escherichia coli* O157:H7 as a model of entry of a new pathogen into the food supply of the developed world, *Epidemiol Rev.*, **18**, 29–51.
7. Tarr, P.I., Gordon, C.A. and Chandler, W.L. 2005, Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome, *Lancet*, **365**, 1073–86.
8. Johnson, K.E., Thorpe, C.M. and Sears, C.L. 2006, The emerging clinical importance of non-O157 Shiga toxin-producing *Escherichia coli*, *Clin. Infect. Dis.*, **43**, 1587–95.
9. Buchholz, U., Bernard, H., Werber, D., et al. 2011, German outbreak of *Escherichia coli* O104:H4 associated with sprouts, *N. Engl. J. Med.*, **365**, 1763–70.
10. Peirano, G. and Pitout, J.D. 2010, Molecular epidemiology of *Escherichia coli* producing CTX-M beta-lactamases: the worldwide emergence of clone ST131 O25:H4, *Int. J. Antimicrob. Agents*, **35**, 316–21.
11. Samuel, G. and Reeves, P. 2003, Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly, *Carbohydr. Res.*, **338**, 2503–19.
12. DebRoy, C., Roberts, E. and Fratamico, P.M. 2011, Detection of O antigens in *Escherichia coli*, *Anim. Health Res. Rev.*, **12**, 169–85.
13. Leopold, S.R., Magrini, V., Holt, N.J., et al. 2009, A precise reconstruction of the emergence and constrained radiations of *Escherichia coli* O157 portrayed by backbone concatenomic analysis, *Proc. Natl Acad. Sci. USA*, **106**, 8713–8.
14. Iguchi, A., Shirai, H., Seto, K., et al. 2011, Wide distribution of O157-antigen biosynthesis gene clusters in *Escherichia coli*, *PLoS ONE*, **6**, e23250.
15. Iguchi, A., Iyoda, S. and Ohnishi, M. 2012, Molecular characterization reveals three distinct clonal groups among clinical Shiga toxin-producing *Escherichia coli* strains of serogroup O103, *J. Clin. Microbiol.*, **50**, 2894–900.
16. Wang, Q., Perepelov, A.V., Wen, L., et al. 2012, Identification of the two glycosyltransferase genes responsible for the difference between *Escherichia coli* O107 and O117 O-antigens, *Glycobiology*, **22**, 281–7.

17. Wang, W., Perepelov, A.V., Feng, L., et al. 2007, A group of *Escherichia coli* and *Salmonella* enterica O antigens sharing a common backbone structure, *Microbiology*, **153**, 2159–67.
18. Lacher, D.W., Gangiredla, J., Jackson, S.A., Elkins, C.A. and Feng, P.C. 2014, Novel microarray design for molecular serotyping of Shiga toxin-producing *Escherichia coli* isolated from fresh produce, *Appl. Environ. Microbiol.*, **80**, 4677–82.
19. Tzschoppe, M., Martin, A. and Beutin, L. 2012, A rapid procedure for the detection and isolation of enterohaemorrhagic *Escherichia coli* (EHEC) serogroup O26, O103, O111, O118, O121, O145 and O157 strains and the aggregative EHEC O104:H4 strain from ready-to-eat vegetables, *Int. J. Food Microbiol.*, **152**, 19–30.
20. Wang, Q., Ruan, X., Wei, D., et al. 2010, 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*, **24**, 286–90.
21. Thompson, J.D., Higgins, D.G. and Gibson, T.J. 1994, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.*, **22**, 4673–80.
22. Tamura, K., Dudley, J., Nei, M. and Kumar, S. 2007, MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0, *Mol. Biol. Evol.*, **24**, 1596–9.
23. Jensen, S.O. and Reeves, P.R. 2004, Deletion of the *Escherichia coli* O14:K7 O antigen gene cluster, *Can. J. Microbiol.*, **50**, 299–302.
24. Pupo, G.M., Lan, R. and Reeves, P.R. 2000, Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics, *Proc. Natl Acad. Sci. USA*, **97**, 10567–72.
25. Liu, B., Knirel, Y.A., Feng, L., et al. 2008, Structure and genetics of *Shigella* O antigens, *FEMS Microbiol. Rev.*, **32**, 627–53.
26. Iguchi, A., Iyoda, S., Seto, K. and Ohnishi, M. 2011, Emergence of a novel Shiga toxin-producing *Escherichia coli* O serogroup cross-reacting with *Shigella boydii* type 10, *J. Clin. Microbiol.*, **49**, 3678–80.
27. Olesen, B., Scheutz, F., Andersen, R.L., et al. 2012, Enteroaggregative *Escherichia coli* O78:H10, the cause of an outbreak of urinary tract infection, *J. Clin. Microbiol.*, **50**, 3703–11.
28. Okeke, I.N., Wallace-Gadsden, F., Simons, H.R., et al. 2010, Multi-locus sequence typing of enteroaggregative *Escherichia coli* isolates from Nigerian children uncovers multiple lineages, *PLoS ONE*, **5**, e14093.
29. Turner, S.M., Chaudhuri, R.R., Jiang, Z.D., et al. 2006, Phylogenetic comparisons reveal multiple acquisitions of the toxin genes by enterotoxigenic *Escherichia coli* strains of different evolutionary lineages, *J. Clin. Microbiol.*, **44**, 4528–36.
30. Nada, R.A., Shaheen, H.I., Khalil, S.B., et al. 2011, Discovery and phylogenetic analysis of novel members of class b enterotoxigenic *Escherichia coli* adhesive fimbriae, *J. Clin. Microbiol.*, **49**, 1403–10.
31. Monaghan, A.M., Byrne, B., McDowell, D., Carroll, A.M., McNamara, E. B. and Bolton, D.J. 2012, Characterization of farm, food, and clinical Shiga toxin-producing *Escherichia coli* (STEC) O113, *Foodborne Pathog. Dis.*, **9**, 1088–96.
32. Hauser, E., Mellmann, A., Semmler, T., et al. 2013, Phylogenetic and molecular analysis of food-borne shiga toxin-producing *Escherichia coli*, *Appl. Environ. Microbiol.*, **79**, 2731–40.
33. Allison, G.E. and Verma, N.K. 2000, Serotype-converting bacteriophages and O-antigen modification in *Shigella flexneri*, *Trends Microbiol.*, **8**, 17–23.
34. Sun, Q., Knirel, Y.A., Lan, R., et al. 2012, A novel plasmid-encoded serotype conversion mechanism through addition of phosphoethanolamine to the O-antigen of *Shigella flexneri*, *PLoS ONE*, **7**, e46095.