O-antigen biosynthesis gene clusters of *Escherichia albertii*: their diversity and similarity to *Escherichia coli* gene clusters and the development of an O-genotyping method

Tadasuke Ooka1,*, Kazuko Seto2, Yoshitoshi Ogura3, Keiji Nakamura3, Atsushi Iguchi4, Yasuhiro Gotoh3, Mikiko Honda5, Yoshiki Etoh6, Tetsuya Ikeda7, Wakana Sugitani8, Takayuki Konno9, Kimiko Kawano10, Naoko Imuta1, Kiyotaka Yoshie1, Yukiko Hara-Kudo11, Koichi Murakami12, Tetsuya Hayashi3,* and Junichiro Nishi1

Abstract

*Escherichia albertii* is a recently recognized human enteropathogen that is closely related to *Escherichia coli*. In many Gram-negative bacteria, including *E. coli*, O-antigen variation has long been used for the serotyping of strains. In *E. albertii*, while eight O-serotypes unique to this species have been identified, some strains have been shown to exhibit genetic or serological similarity to known *E. coli/Shigella* O-serotypes. However, the diversity of O-serotypes and O-antigen biosynthesis gene clusters (O-AGCs) of *E. albertii* remains to be systematically investigated. Here, we analysed the O-AGCs of 65 *E. albertii* strains and identified 40 *E. albertii* O-genotypes (EAOgs) (named EAOg1–EAOg40). Analyses of the 40 EAOgs revealed that as many as 20 EAOgs exhibited significant genetic and serological similarity to the O-AGCs of known *E. coli/Shigella* O-serotypes, and provided evidence for the interspecies horizontal gene transfer of O-AGCs between *E. albertii* and *E. coli*. Based on the sequence variation in the *wzx* gene among the 40 EAOgs, we developed a multiplex PCR-based O-genotyping system for *E. albertii* (EAO-genotyping PCR) and verified its usefulness by genotyping 278 *E. albertii* strains from various sources. Although 225 (80.9 %) of the 278 strains could be genotyped, 51 were not assigned to any of the 40 EAOgs, indicating that further analyses are required to better understand the diversity of O-AGCs in *E. albertii* and improve the EAO-genotyping PCR method. A phylogenetic view of *E. albertii* strains sequenced so far is also presented with the distribution of the 40 EAOgs, which provided multiple examples for the intra-species horizontal transfer of O-AGCs in *E. albertii*.

DATA SUMMARY

The assembled and read sequences of 11 *Escherichia albertii* strains obtained in this study have been deposited in GenBank/EMBL/DDBJ under the BioProject accession number PRJDB8401. The annotated sequences of O-antigen biosynthesis gene clusters have been deposited in GenBank/EMBL/DDBJ under accession numbers LC494303–LC494359.
INTRODUCTION

Escherichia albertii is a recently recognized human enteropathogen and an avian pathogen responsible for epidemic mortality [1–3]. Several outbreaks caused by this microorganism have recently been reported [4–6]. E. albertii strains carrying Shiga toxin (Stx) genes (stxa2 or stx2f) have been identified [3, 7, 8], indicating that there is a significant risk of severe infections caused by this species. Another important aspect is that E. albertii infections have been underestimated because they have been frequently misidentified as enteropathogenic or enterohaemorrhagic Escherichia coli (EPEC or EHEC, respectively) due to their similarity in phenotypic and genetic features, such as biochemical properties and the possession of the locus of enterocyte effacement (LEE), encoding a type III secretion system (T3SS) [3]. Although genome sequencing analyses of more than 50 E. albertii strains have been reported [9–13], the genomic diversity of this species is yet to be fully understood.

O-antigen is a lipopolysaccharide component of the outer membrane of Gram-negative bacteria. The chemical compositions and structures are highly variable even in the same species [14, 15], which have long been used for the serotyping of strains. In E. coli, O-antigen biosynthesis gene clusters (O-AGCs) are located between the colonic acid biosynthesis gene cluster (the wca genes) and the his operon on the chromosome, with a few exceptions [16, 17]. O-antigen biosynthesis genes fall into three classes: (i) genes for nucleotide sugar biosynthesis, (ii) genes encoding sugar transferase, and (iii) genes for O-unit translocation and chain synthesis [18]. The variation in the repertoire of these genes in O-AGCs is responsible for the differences in O-antigen structure and, therefore, in O-serotype. Structural and serological changes in O-antigens can occur via point mutations in the genes in O-AGCs, such as glycosyltransferase genes, or by acquisition of O-antigen modification genes [19, 20].

The presence of genetic similarity between some E. albertii and E. coli O-AGCs has been suggested by an analysis of O-AGCs of published E. albertii genomes [21], while eight O-serotypes (named EAO1–EAO8) unique to E. albertii have been identified along with their chemical structures [21–26]. It has also been reported that an environmental E. albertii isolate (strain DM104) serologically cross-reacts with the antiserum of Shigella dysenteriae type 4 [27], and that some E. albertii strains express the O-antigen of Shigella boydii serotype 13 [28]. However, to better understand the diversity of E. albertii O-serotypes and the genetic variation underlying their diversity and similarity to E. coli O-serotypes, a systematic, large-scale analysis of E. albertii O-AGCs is required.

In this study, we performed detailed analyses of the O-AGCs of 65 E. albertii strains, including EAO1–EAO8, and those of 57 genome-sequenced strains, to investigate the diversity of E. albertii O-AGCs. Inter-species comparison of O-AGCs with known E. coli O-AGCs and those of E. coli relatives was also performed to understand the serological and genetic similarity between E. albertii and E. coli O-serotypes. Based on the results, we attempted to develop a multiplex PCR-based O-genotyping system for E. albertii (EAO-genotyping PCR) and to provide a current view on the diversity of O-serotypes in the phylogenetic context of E. albertii strains.

METHODS

Bacterial strains

The 65 E. albertii strains analysed in this study are listed in Table S1, (available with the online version of this article) along with detailed information on each strain. Among these strains, 57 were genome-sequenced strains, including 30 that were sequenced in our previous study [9, 10], 16 for which the genome sequences were obtained from the GenBank genome database [11–13] and 11 that were sequenced in this study (Table S1). The strain information for the 92 E. albertii strains used for the evaluation of EAO-genotyping PCR is provided in Table S2. The E. albertii strains analysed were cultured overnight in lysogeny broth (LB; Nippon Becton Dickinson) at 37 °C with shaking.

Genome sequencing of 11 E. albertii strains

Genomic DNA was purified from a 2 ml overnight culture using the DNeasy blood and tissue kit (Qiagen), according to the manufacturer’s instructions. Sequencing libraries for each strain were prepared using the Nextera XT DNA sample prep kit (Illumina) to obtain paired-end sequences (300 bp ×2) on an Illumina MiSeq instrument. Draft genome sequences were obtained by assembling the read sequences using Platanus v1.1.4.

Impact Statement

Escherichia albertii is a close relative of Escherichia coli and has been frequently misidentified as enteropathogenic E. coli (EPEC) due to their phenotypic and genetic similarity and their possession of a similar type III secretion system (T3SS). The clinical importance of E. albertii is increasingly being recognized, as this species causes outbreaks of gastroenteritis, and some strains produce Shiga toxins, similar to enterohaemorrhagic E. coli (EHEC). Rapid strain typing systems useful for outbreak investigation and surveillance, such as O-serotyping, which is widely used for E. coli, need to be developed. However, the diversity of O-serotypes and O-antigen biosynthesis gene clusters (O-AGCs) of E. albertii is largely unknown. This study identified 40 E. albertii O-genotypes by analysing the O-AGCs of 65 E. albertii strains and provides what is believed to be the first insight into the diversity of E. albertii O-AGCs, as well as evidence for the horizontal transfer of O-AGCs within E. albertii, and between E. albertii and E. coli. Moreover, based on these results, a multiplex PCR-based O-genotyping system for E. albertii was developed, which could contribute to the improvement and promotion of epidemiological studies of this understudied pathogen.
with default parameters [29]. The sequencing status of each strain is shown in Table S3.

**Identification and annotation of O-AGCs of E. albertii strains**

O-AGCs were identified by BLASTN search using the galF and gnd sequences as queries, with an E value threshold of 0.01. Protein encoding sequences and their functions were predicted using the Microbial Genome Annotation Pipeline (MiGAP; http://www.migap.org) which has been recently closed, followed by manual curation using the *in silico* Molecular Cloning Genomics Edition software v7.29L (IMC-GE; In Silico Biology). Domain searches were performed using the Pfam program v32, with an E value threshold of 0.01 [30].

**Intra- and inter-species comparison of O-AGCs**

For this analysis, we obtained the O-AGC sequences of 8 known *E. albertii* serotypes [21–23, 26], 201 known *E. coli* O-serotypes [16, 17, 31] and 19 known *Shigella* serotypes [32] from the NCBI database (Table S4). In addition, we analysed the publicly available genome sequences of three *Escherichia fergusonii* strains and eight strains belonging to cryptic *Escherichia* clades, and identified and annotated their O-AGCs as described above. The accession numbers of the O-AGC sequences and genome sequences used for comparative analysis are listed in Table S4 (*Escherichia/Shigella* species). Sequence comparison of the O-AGCs was performed by BLASTN using the nucleotide sequences of each protein.

**Agglutination test using the antisera for known *E. coli* O-antigens**

O-serotype cross-reactivity was determined by the slide agglutination test and/or the microtitre plate method using the *E. coli* antiserum set 1 (Denka Seiken), the pooled and single antiserum against all known *E. coli* O-serotypes (SSI Diagnostica), and the *S. boydii* type 13 antiserum (Denka Seiken), according to the manufacturers’ instructions. The SSI Diagnostica set included antiserum against *E. coli* O1 to O187, excluding five serotypes (O31, O47, O67, O94 and O122; these serotypes are missing in the current serotyping scheme). When a strain was weakly agglutinated compared with the reference strain or agglutinated with multiple antiserum, we performed titration tests using microtitre plates. Weak agglutination was defined if the titre was four times lower than that of the reference strain.

**Development of a multiplex EAO-genotyping PCR**

Based on the sequence variation in the *wzx/wzy* and *wzm/wzt* genes of 40 *E. albertii* EAOs and those of known *E. coli/Shigella* O-serotypes and other *Escherichia* species (Table S4) were performed using MEGA6 software [36]. In brief, the nucleic acid sequences of each gene were aligned by ClustalW with default parameters. A phylogenetic tree was reconstructed with the neighbour-joining method using the p-distance model to calculate nucleotide distance. Bootstrap analysis with 1000 replicates was performed to evaluate the significance of internal branches. For the reconstruction of a core gene-based phylogenetic tree, the genome assemblies of 57 strains used for detailed analysis of O-AGCs and 186 strains used for *in silico* EAO-genotyping were annotated using Prokka [37], and core genes (*n*=2128) were identified using Roary v3.11.2 [38] with a 90 % amino acid sequence identity cut-off. Core gene SNPs (*n*=94 287) were extracted using the core gene alignment tool in Roary and used as inputs for maximum-likelihood (ML) inference with RAxML v8 [39]. The ML tree was displayed and annotated using iTOL v4 [https://itol.embl.de] [40]. The tree was rooted by the mid-point rooting method and the confidence of each branch was estimated by bootstrap with 200 replications. Identical *E. albertii* genomes showing no SNPs were deduplicated (excluded strains are indicated in Table S6).

**In silico EAO-genotyping**

*In silico* EAO-genotyping was performed for the strains registered in the EnteroBase website v1.1.2. (https://enterobase.warwick.ac.uk [34]). From the 274 strains registered as *E. albertii* (accessed on 9 September 2019), 87 strains were excluded due to the overlap with the strains used for detailed analysis in this study or the lack of genome sequence information. In addition, one strain was excluded due to a low completeness (<95 %) and a high contamination (>5 %) as estimated by CheckM [35], and a high level of fragmentation (longest contig<100 kb). The final set of *E. albertii* strains used for *in silico* EAO-genotyping (*n*=186) is listed in Table S6. *In silico* EAO-genotyping was performed by BLASTN search using the nucleotide sequences of the 42 primer pairs (including 2 primer pairs targeting a genetic marker of *E. albertii*) designed for EAO-genotyping PCR as queries (Table S5). Only a perfect match was considered.

**Phylogenetic analyses**

Sequence comparisons and phylogenetic analyses of the *wzx/wzy* and *wzm/wzt* genes of 40 *E. albertii* EAOs and those of known *E. coli/Shigella* O-serotypes and other *Escherichia* species (Table S4) were performed using MEGA6 software [36]. In brief, the nucleic acid sequences of each gene were aligned by ClustalW with default parameters. A phylogenetic tree was reconstructed with the neighbour-joining method using the p-distance model to calculate nucleotide distance. Bootstrap analysis with 1000 replicates was performed to evaluate the significance of internal branches. For the reconstruction of a core gene-based phylogenetic tree, the genome assemblies of 57 strains used for detailed analysis of O-AGCs and 186 strains used for *in silico* EAO-genotyping were annotated using Prokka [37], and core genes (*n*=2128) were identified using Roary v3.11.2 [38] with a 90 % amino acid sequence identity cut-off. Core gene SNPs (*n*=94 287) were extracted using the core gene alignment tool in Roary and used as inputs for maximum-likelihood (ML) inference with RAxML v8 [39]. The ML tree was displayed and annotated using iTOL v4 [https://itol.embl.de] [40]. The tree was rooted by the mid-point rooting method and the confidence of each branch was estimated by bootstrap with 200 replications. Identical *E. albertii* genomes showing no SNPs were deduplicated (excluded strains are indicated in Table S6).
Fig. 1. Genetic structures of the 40 *E. albertii* O-genotypes identified. Genotypes, strain names and results of the agglutination test using antisera against known *E. coli/Shigella* O-serotypes (N, no agglutination; NA, not applicable) are indicated on the left side. The structures of O-AGCs were deduced from the genome sequences of each strain except for the eight strains indicated by asterisks. Only the sequences of O-AGCs are available for the eight strains. For putative glycosyltransferases (indicated by green), types of domain families are indicated instead of gene names (UT, untypable).
Fig. 2. Comparison of the O-AGCs of EAOgs with homologous O-AGCs of known E. coli/Shigella O-serotypes. Shading and numbers between O-AGCs indicate the nucleotide sequence identities (%) of each gene. EAOgs indicated by asterisks (n=20) were found to be serologically cross-reactive with each E. coli/Shigella counterpart. Although EAOg30 cross-reacts with E. coli O69 and weakly with O124, its O-AGC shows similarity only to that of E. coli O69. EAOg8 exhibits significant similarity to E. coli O55 and O128 and cross-reacts with both E. coli O-serotypes. In the five EAOgs without asterisks, the cross-reactivity to their counterparts could not be examined due to a lack of E. albertii strains in our laboratory.
RESULTS
Sequence analysis of the O-AGCs from 65 E. albertii strains

To enrich the genome sequence and strain resources of E. albertii that can be used for the detailed analysis of O-AGCs and serotypes, we newly sequenced 11 E. albertii strains in this study. By adding these sequences to the 46 previously sequenced genomes and the 8 O-AGC sequences previously reported as EAO1–EAO8, we analysed the O-AGCs of 65 E. albertii strains from various sources, including humans, birds, cats and the environment (Table S1).

General features and genotypes of the E. albertii O-AGCs

In 47 of the 65 strains analysed, O-AGCs were located between the truncated wcaM gene and the hisI gene. In 10 strains, either the wcaM or the hisI gene was identified, but the opposite boundary of O-AGC was not identified due to the interruption of corresponding contig sequences in their draft genomes (Fig. 1). The remaining eight strains were the previously reported strains of EAO1–EAO8, for which only the sequences between the galF and ugd genes were available. The sizes of E. albertii O-AGCs ranged from 13 to 30 kb. The mean G+C content was 37.8 mol% (ranging from 35.1 to 42.5 mol%). Based on gene organization, the 65 O-AGCs were grouped into 40 clearly distinguishable genotypes, named EAOg1–EAOg40 (Fig. 1, Table S1). EAO1–EAO8 are referred to as EAOg1–EAOg8, respectively, in this manuscript. The O-AGC corresponding to EAO6 (now referred to as EAOg6) was not found in the 57 genome-sequenced strains.

Nucleotide sugar biosynthesis genes

The rmlBDAC genes, required for the synthesis of deoxythymidine diphosphate (dTDP)-l-rhamnose (the precursor of l-rhamnose) by the dTDP-sugar biosynthesis pathway, were present in 12 O-gengotypes (Fig. 1). Three of the
Fig. 3. Neighbour-joining trees of the genes for O-antigen subunit translocation and chain synthesis. Neighbour-joining trees were reconstructed based on the sequences of the \textit{wzx}, \textit{wzy}, \textit{wzm} and \textit{wzt} genes from 40 EAOGs (red), 184 known O-genotypes of \textit{E. coli} (black), 34 known serogroups of \textit{Shigella} species (purple), 3 genome-sequenced \textit{E. fergusonii} strains (green) and 8 genome-sequenced strains belonging to cryptic \textit{Escherichia} clades (blue). Bar, the nucleotide substitutions (%) per site.
Fig. 4. Electrophoresis patterns obtained by EAO-genotyping PCR. A total of 33 strains representing 33 EAOgs were analysed using three sets of PCR primers designed in this study. Strain names are indicated in parentheses. Strains belonging to the remaining seven EAOgs were not available in our laboratory. Arrowheads 1 and 2 indicate the bands derived from the E. albertii-specific primer pairs E_al_1_OF/OR and E_al_1_NF/NR, respectively. Lane M, 100 bp DNA ladder.

genotypes (EAOg12, EAOg13 and EAOg16) contained only the rmlBA genes. Similar to the O-AGCs of E. coli O63, O184 and O65, which are homologous to each of the three EAOgs described below, it is likely that the dTDP-6-deoxy-δ-xylo-4-hexulose 3,5-epimerase (encoded by rmlC) and dTDP-4-dehydrorhamnose reductase (rmlD) genes are located outside the O-AGCs. The nnaDBCA genes for the synthesis of cytidine monophosphate-N-acetylneuraminate (CMP-NeuNAc) were found in three O-genotypes (EAOg7, EAOg33 and EAOg39). The fnlA and qnlBC genes for the synthesis of UDP-N-acetyl-l-quinovosamine (UDP-l-QuiNAc) were found in EAOg9. The remaining 24 O-genotypes contained no nucleotide sugar biosynthesis genes or only a limited number of genes for the biosynthesis of certain nucleotide sugars, suggesting that, in these O-genotypes, nucleotide sugars or their precursors to be incorporated into O-antigen are supplied via the functions of genes located outside the O-AGCs.

Glycosyltransferase genes
All the genotypes except EAOg34 contained two to six genes encoding putative glycosyltransferases (132 glycosyltransferase genes in total). A Pfam search of the 132 gene products identified 9 types of glycosyltransferase-related domains (Table S7; 18 were untypable), with families 2 (PF00535), 1 (PF00534) and 1_4 (PF13692) being the most widely distributed (51, 37 and 13 genes, respectively).

Genes for O-antigen processing
For O-unit processing and conversion of the O-unit to O-antigen, all the O-genotypes contained the wzx/wzy gene set. EAOg29 additionally contained the wzml/wzmt gene set.

Comparison of E. albertii O-AGCs with known Escherichia/Shigella O-AGCs
We analysed the structural similarity of the O-AGCs of the 40 EAOgs to those of known Escherichia/Shigella O-serotypes. The O-AGCs of 20 EAOgs (EAOg3, EAOg6 and EAOg9–26) were significantly similar to those of some Escherichia/Shigella serotypes in terms of gene organization and nucleotide sequence identity [most genes showed >80 % nucleotide sequence identity (83–99 % on average) to the gene of each counterpart] (Table S1, Fig. 2). Among the 20 EAOgs, eight (EAOg6 and EAOg9–EAOg15) exhibited particularly high sequence similarity [most genes showed >98 % nucleotide sequence identity (>96 % on average)] to the counterparts of E. coli or S. boydii (Table S1, Fig. 2).

In contrast to the 20 EAOgs, most genes in 3 EAOgs (EAOg27–EAOg29) contained the same set of genes as counterparts in E. coli, but the nucleotide sequence identities were 83–89 % on average (Table S1, Fig. 2). Two EAOgs (EAOg8 and EAOg30) showed sequence similarity to only parts of the O-AGCs of E. coli O128/O55 and O69, respectively (Fig. 2). The remaining 15 EAOgs (EAOg1, EAOg2, EAOg4, EAOg5, EAOg7 and EAOg31–EAOg40) showed no significant similarity to the O-AGCs of known E. coli/Shigella serotypes, except for local similarities (e.g. between the genes encoding proteins of the same family). No significant similarity was observed between the O-AGCs of the 40 EAOgs and those of genome-sequenced strains belonging to E. fergusonii and the cryptic clades of Escherichia species.
Serological cross-reactivity of *E. albertii* O-antigens to known *E. coli* O-serotypes and *S. boydii* type 13

As many of the 40 EAOgs showed high similarity to the O-AGCs of known *E. coli/Shigella* serotypes, we analysed the serological cross-reactivity of the O-antigens of *E. albertii* to *E. coli/Shigella* O-antigens using antisera against 182 known *E. coli* O-serotypes and those of *Shigella* species. In this analysis, only 42 *E. albertii* strains (belonging to 33 EAOgs) were used, because the remaining 14 strains (belonging to EAOg6, EAOg10, EAOg15, EAOg23, EAOg28, EAOg37 and
Correlation of the genetic similarity of O-genotypes and their serological cross-reactivity to known O-serotypes of Escherichia/Shigella species

We examined the correlation between the structural similarity of O-AGCs and the serological cross-reactivity to known Escherichia/Shigella O-serotypes. Among 25 EAOgs (EAOg3, EAOg6 and EAOg8–30) that showed notable similarities to known E. coli/Shigella serotypes, 5 (EAOg6, EAOg10, EAOg15, EAOg23 and EAOg28) were unable to undergo serological examination. However, all of the remaining 20 EAOgs tested showed serological cross-reactivity to the E. coli/Shigella counterparts (the serotypes that showed structural similarities). Thus, the structural similarity of O-AGCs between E. albertii and Escherichia/Shigella is well correlated with the serological cross-reactivity between these O-serotypes, suggesting that the five EAOgs that could not be tested probably cross-reacted with their E. coli counterparts. The exceptions observed were EAOg1 and EAOg30, which showed cross-reactivity with E. coli O180 and O175, respectively, but were not genetically related to these E. coli O-serotypes. This finding suggested that some common chemical structures that could not be predicted based on genetic structures are responsible for these cross-reactivities.

Intra- and inter-species sequence comparison of 0-antigen processing-related genes (wzx/wzy and wzm/wzt)

A pair of O-antigen processing-related genes, wzx/wzy, which are essential for O-antigen biosynthesis, were identified in the O-AGCs of all E. albertii strains analysed. One strain also contained other types of genes (wzm/wzt). As these genes were used for developing a PCR-based O-genotyping system of E. coli [41], these E. albertii genes may also be usable as targets to develop a similar PCR-based O-genotyping system for E. albertii. Therefore, we performed a fine sequence comparison of the wzx/wzy and wzm/wzt genes of E. albertii with those of known E. coli/Shigella O-serotypes and other Escherichia species (Fig. 3, Table S1). This analysis revealed that the wzx/wzy genes of EAOg6, EAOg9–18 and EAOg30 and the wzm/wzt genes of EAOg2 are highly similar (>95 % nucleotide sequence identity) to the genes of E. coli/Shigella counterparts, as expected from the results of whole-O-AGC comparison (Fig. 3, Table S1). Among the remaining 28 EAOgs, although 12 contained the wzx/wzy genes showing moderate similarity to the genes of E. coli/Shigella or other Escherichia species (80–94 % nucleotide sequence identity), the genes of 16 EAOgs shared less than 70 % nucleotide sequence identity with any of the wzx/wzy genes from E. coli and other Escherichia species. Interestingly, while the wzm/wzt genes of EAOg29 shared more than 98 % nucleotide sequence identity with the genes of E. coli O8 as mentioned above, the wzx/wzy of EAOg29 showed no significant similarity to the genes of E. coli O8, suggesting the exchange of wzx/wzy in either the E. albertii or E. coli O-AGCs.

Development and evaluation of an O-genotyping PCR system for E. albertii

Although the sequences of some wzx/wzy genes of E. albertii were highly similar to those of E. coli, the wzx/wzy genes of the 40 EAOgs were significantly divergent in sequence, indicating that the wzx/wzy genes can be used for developing an O-genotyping PCR system for E. albertii (EAO-genotyping PCR). Therefore, we designed three multiplex PCR primer sets (Table S5) based on the sequence variation in the wzx genes of the 40 EAOgs (Fig. 3, Table S1). Considering the high similarities of several wzx genes to E. coli genes, we included one E. albertii–specific primer pair (E_al_1_OF/E_al_1_OR in the first set, and E_al_1_NF/E_al_1_NR in the second and third sets) [9] in each primer set as a marker to distinguish E. albertii from E. coli and other species. Using the three sets of PCR primers, we examined the strains representing 33 EAOgs (Fig. 4). The result indicated that the three primer sets yielded PCR products of the expected sizes for each strain. Strains belonging to the remaining seven EAOgs were not available in our laboratory.

To evaluate the performance of the system developed, we determined the O-genotypes of 92 E. albertii strains using this system. These strains were isolated from diarrhoeal patients or birds in various regions in Japan. In this analysis, we were able to genotype 76 strains (29 EAOgs; Table S2). Notably, 17 (26.6 %) of the 64 strains isolated from pigeon faeces belonged to EAOg30, although these samples were obtained in five different geographical regions. We were unable to assign 16 strains (17.4 %) to any of the 40 EAOgs. Using the primer sequences, we further performed in silico E. albertii O-genotyping of an additional 186 genome-sequenced E. albertii strains registered in the Enterobase website. In this analysis, 151 strains were genotyped (25 EAOgs), while 35 strains (18.8 %) were not. Although the ratio of untypeable strains was similar to that of the 92 Japanese strains, EAOg4 (37 strains; 19.9 %) and EAOg9 (25 strains; 13.4 %) were predominant in the Enterobase strains (Table S3). These results indicate that while the system that we developed is useful for O-genotyping of E. albertii, the diversity of...
O-genotypes among *E. albertii* strains is much larger than that captured in this study.

**Distribution of the 40 EAOgs in the *E. albertii* strains genome-sequenced so far and a current phylogenetic view of *E. albertii***

Finally, we reconstructed an ML phylogenetic tree based on the core gene sequences to investigate the phylogenetic relationship of the genome-sequenced strains used in this study and the distribution of the 40 EAOgs defined in this study (Fig. 5; note that strains having identical core gene sequences were deduplicated; therefore, 225 strains were included in this analysis). This analysis provided a current phylogenetic view of *E. albertii* and the distribution of the 40 EAOgs in it. Although *E. albertii* strains were divided into two clades (clades 1 and 2) and both clades, particularly clade 2, contained multiple deep branching lineages, clear correlation was not observed between specific O-genotypes and strain sources (isolation sources and geographical locations). Notably, seven O-genotypes (EAOg8, EAOg10, EAOg16, EAOg18, EAOg32, EAOg35 and EAOg38) were distributed in both clades, providing evidence for the intra-species transfer of O-AGCs in *E. albertii* (Fig. 5).

**DISCUSSION**

In this study, we analysed the O-AGCs of 65 *E. albertii* strains and defined 40 O-genotypes, named EAOg1–EAOg40. Detailed analyses of the genetic structures and sequences of these genotypes revealed that as many as 25 EAOgs showed notable similarities to the O-AGCs of known *E. coli/Shigella* serotypes. In accordance with this finding, these EAOgs, except for five EAOgs that could not be examined due to unavailability of strains, showed serological cross-reactivity with corresponding *E. coli* and *Shigella* serotypes. Presumably, the five untested EAOgs also cross-react with the *E. coli/Shigella* serotypes, to which they showed high similarity. Among the 25 EAOgs, 8 (EAOg6 and EAOg9–15) are particularly interesting because their entire O-AGCs were highly similar, even at the level of nucleotide sequence (most showing >98 % nucleotide sequence identity), to their *E. coli/Shigella* counterparts. The observed high similarity suggests the recent inter-species transfer of these O-AGCs. A similar finding was also obtained between *E. coli* and *E. fergusonii* [42].

Another notable finding was the cross-reactivity of EAOg8 to *E. coli* O55 and O128. This result is consistent with previous findings that the main chain of O-poly saccharides of EAO8 showed a marked similarity to that of *E. coli* O128, and that the terminal monosaccharide of O-poly saccharide is α-colitose in both EAO8 and O55 [26]. EAOg29 is also interesting in that it contains both the wzx/wzy and wzr/wzt gene sets for O-antigen processing. The latter gene set is very similar to that of *E. coli* O8, and EAOg29 was genetically and serologically similar to *E. coli* O8 (Fig. 3). These findings suggest that the wzr/wzt gene set is mainly involved in the synthesis of EAOg29 O-antigen. It would be interesting to know whether or how the wzx/wzy genes are involved in or affect O-antigen synthesis in EAOg29.

We developed an O-genotyping PCR system for *E. albertii* (EAO-genotyping PCR) based on the finding that the wzx/wzy genes are highly divergent in sequence among the 40 EAOgs. The system is similar to those for *E. coli* [41, 43], but differs in the inclusion of primers to detect *E. albertii*-specific gene [9] in each primer set. These primers were included because the wzx sequences of several EAOgs showed >95 % identity to those of *E. coli/Shigella* serotypes. The PCR products derived from these primers would work as markers to distinguish *E. albertii* from *E. coli* and other *Escherichia* species. The utility of the system was verified by genotyping 92 *E. albertii* strains (PCR-based analysis) and 186 strains (in silico analysis) obtained from various sources. However, 51 (18.3 %) of the 278 strains tested could not be assigned to any of the 40 EAOgs. This observation indicates that further analyses, particularly of O-AGCs of strains untyped in this study, are required to understand the diversity of O-AGCs in *E. albertii* and improve the EAO-genotyping PCR method.

Our analysis of the phylogenetic relationship of more than 200 *E. albertii* strains genome-sequenced so far and the distribution of the 40 EAOgs in these strains revealed that there was no clear correlation between the distribution of EAOgs and the geographical locations or sources of strain isolation in the strain set analysed. An important finding of this analysis was that as many as seven EAOgs were distributed in two phylogenetically distinct clades (Fig. 5). This finding suggests that intra-species transfer of O-AGCs may have occurred more frequently in *E. albertii* than in *E. coli* [16, 44, 45].

In conclusion, this study provides what is believed to be the first insight into the diversity of O-AGCs in *E. albertii* and defines 40 EAOgs. We showed the genetic and serological similarity of many EAOgs to O-AGCs of known *E. coli/Shigella* serotypes, and obtained evidence for the intra- and inter-species transfer of O-AGCs within *E. albertii* and between *E. albertii* and *E. coli*. In addition, based on the nucleotide sequence diversity of the wzx genes among the 40 EAOgs, we developed an O-genotyping PCR system for *E. albertii*. Results of the evaluation of this system by PCR or in silico genotyping indicate that this O-genotyping PCR system is useful, but further analyses are required to understand the diversity of O-AGCs in *E. albertii*.

**Funding information**

This work was supported by the Japan Society for the Promotion of Science KAKENHI (grant numbers 23790480, 25460539 and 16K08781 to T. O., and 20310116 and 22150002 to T. H.), and a Health Labour Sciences Research Grant [H30-Shokuhin-Ippan-001 to Y. H.-K.].

**Acknowledgements**

The authors thank K. Katsura, A. Yoshida, N. Kawano, Y. Inoue and F. Funakura for providing technical assistance.

**Author contributions**

Conceptualization: T. O., T. H. Methodology: T. O., T. H. Investigation: T. O., K. S. Formal Analysis: T. O., K. S., Y. O., A. I., K. N., Y. G., T. H. Resources: M. H., Y. E., T. I., W. S., T. K., K. K., N. I., K. Y., Y. H.-K., K. M. Funding: T. O., Y. H.-K., T. H. Writing
Conflicts of interest
The authors declare that there are no conflicts of interest.

Data bibliography
Ooka T. GenBank, BioProject number PRJDB8401, accession numbers LC94303−LC94359 (2019).

REFERENCES
1. Huys G, Cnockaert M, Janda J, Swings J. Escherichia albertii sp. nov., a diarrhoeagenic species isolated from stool specimens of Bangladeshi children. Int J Syst Evol Microbiol 2003;53:807−810.

2. Ooka TL, Besser TE, Walk ST, Gordon DM, Beckmen KB et al. Escherichia albertii in wild and domestic birds. Emerg Infect Dis 2010;16:638−646.

3. Ooka T, Seto K, Kawano K, Kobayashi H, Etoh Y et al. Clinical significance of Escherichia albertii. Emerg Infect Dis 2012;18:488−492.

4. Ooka T, Tokuoka E, Furukawa M, Nagamura T, Ogura Y et al. Human gastroenteritis outbreak associated with Escherichia albertii. Japan. Emerg Infect Dis 2013;19:144−146.

5. Inglis TJ, Merritt AJ, Bzdyl N, Lansley S, Urosevic MN. First bacte- riaemic human infection with Escherichia albertii. New Microbes New Infect 2015;8:171−173.

6. National Institute of Infectious Diseases. Enterohemorrhagic Escherichia coli (EHEC) infection, as of April 2016. Japan. Infect Agents Survell Rep 2016;37:435.

7. Murakami K, Etoh Y, Tanaka E, Ichihara S, Horikawa K et al. Shiga toxin 2-producing Escherichia albertii from a symptomatic human. Jpn J Infect Dis 2014;67:204−208.

8. Brandal LT, Tunsjo HS, Ranheim TE, Lebersli I, Lange H et al. Shiga toxin 2A in Escherichia albertii. J Clin Microbiol 2015;53:1454−1455.

9. Ooka T, Ogura Y, Katsura K, Seto K, Kobayashi H et al. Defining the genome features of Escherichia albertii, an emerging enteropathogen closely related to Escherichia coli. Genome Biol Evol 2015;7:3170−3179.

10. Romão FT, Hernandes RT, Ooka T, Hayashi T, Sperandio V et al. Complete genome sequence of Escherichia albertii strain 1551-2, a potential extracellular and intracellular pathogen. Genome Announc 2018;6:e00075-18.

11. Fiedoruk K, Daniulit T, Swiecicka I, Murawska E, Sciepuk M et al. First complete genome sequence of Escherichia albertii strain KF1, a new potential human enteric pathogen. Genome Announc 2014;2:e00004-14.

12. Schroeder MR, Juengp B, Batra D, Kniep K, Rowe LA et al. High-quality complete and draft genome sequences for three Escherichia spp. and three Shigella spp. generated with Pacific Biosciences and Illumina sequencing and optical mapping. Genome Announc 2018;6:e01384-17.

13. Lindsley RL, Rowe LA, Batra D, Smith P, Strockbine NA. PacBio genome sequences of eight Escherichia albertii strains isolated from humans in the United States. Microbiol Resour Announc 2019;8:e01663-18.

14. Stenutz R, Weintraub L, Widmalm G. The structures of Escherichia coli O-polsaccharide antigens. FEMS Microbiol Rev 2006;30:382−403.

15. Liu B, Knirel YA, Feng L, Perepelov AV, Senchenkova SN et al. Structural diversity in Salmonella O antigens and its genetic basis. FEMS Microbiol Rev 2014;38:56−89.

16. Iguchi A, Iyoza S, Kikuchi T, Ogura Y, Katsura K et al. A complete view of the genetic diversity of the Escherichia coli O-antigen biosynthesis gene cluster. DNA Research 2015;22:101−107.

17. Iguchi A, Iyoza S, Seto K, Nishi H, Ohnishi M et al. Six novel O genotypes from Shiga toxin-producing Escherichia coli. Front Microbiol 2016;7:765.

18. Samuel G, Reeves P. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. Carbohydr Res 2003;338:2503−2519.

19. Wang W, Perepelov AV, Feng L, Shevelev SD, Wang Q et al. A group of Escherichia coli and Salmonella enterica O antigens sharing a common backbone structure. Microbiology 2007;153:2159−2167.

20. Wang Q, Perepelov AV, Wen L, Shashkov AS, Wang X et al. Identification of the two glycosyltransferase genes responsible for the difference between Escherichia coli O107 and O117 O-antigens. Glycobiology 2012;22:281−287.

21. Wang H, Zheng H, Li Q, Xu Y, Wang J et al. Defining the genetic features of O-antigen biosynthesis gene cluster and performance of an O-antigen serotyping scheme for Escherichia albertii. Front Microbiol 2017;8:1857.

22. Zheng H, Shashkov AS, Xiong Y, Naumenko OI, Wang H et al. Structure and gene cluster of the O-antigen of Escherichia albertii O1 resembling the O-antigen of Pseudomonas aeruginosa O5. Carbohydr Res 2017;446−447:28−31.

23. Naumenko OI, Zheng H, Senchenkova SN, Wang H, Li Q et al. Structures and gene clusters of the O-antigens of Escherichia albertii O3, O4, O6, and O7. Carbohydr Res 2017;449:17−22.

24. Naumenko OI, Zheng H, Wang J, Senchenkova SN, Wang H et al. Structure elucidation of the O-specific polysaccharide by NMR spectroscopy and selective cleavage and genetic characterization of the O-antigen of Escherichia albertii O5. Carbohydr Res 2018;457:25−31.

25. Naumenko OI, Zheng H, Xiong Y, Senchenkova SN, Wang H et al. Studies on the O-polsaccharide of Escherichia albertii O2 character- ized by non-stoichiometric O-acetylation and non-stoichiometric side-chain L-fucosylation. Carbohydr Res 2018;461:80−84.

26. Zheng H, Naumenko OI, Wang H, Xiong Y, Wang J et al. Colitose- containing O-polsaccharide structure and O-antigen gene cluster of Escherichia albertii HK18069 related to those of Escherichia coli O55 and E. coli O128. Carbohydr Res 2019;480:73−79.

27. Rahman MZ, Akter S, Azmuda N, Sultana M, Weil FX et al. Serological cross-reaction between O-antigens of Shigella dysenteriae type 4 and an environmental Escherichia albertii isolate. Curr Microbiol 2013;67:590−596.

28. Hyma KE, Lacher DW, Nelson AM, Bumbaugh AC, Janda J et al. Evolutionary genetics of a new pathogenic Escherichia species: Escherichia albertii and related Shigella boydii strains. J Bacteriol 2005;187:619−628.

29. Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y et al. Efficient de novo assembly of highly heterozygous genomes from whole- genome shotgun short reads. Genome Res 2014;24:1384−1395.

30. Finn RD, Mistry J, Tate J, Coggill P, Heger A et al. The Pfam protein families database. Nucleic Acids Res 2010;38:D211−D222.

31. DebRoy C, Fratamico PM, Yan X, Baranzenzi G, Liu Y et al. Comparison of O-antigen gene clusters of all O-serogroups of Escherichia coli and proposal for adopting a new nomenclature for O-typing. PLoS One 2011;6:e1047434.

32. Liu B, Knirel YA, Feng L, Perepelov AV, Senchenkova SN et al. Structure and genetics of Shigella O antigens. FEMS Microbiol Rev 2008;32:627−653.

33. Ooka T, Terajima J, Kusumoto M, Iguchi A, Kurokawa K et al. Development of a multiplex PCR-based rapid typing method for enterohemorrhagic Escherichia coli O157 strains. J Clin Microbiol 2009;47:2888−2894.

34. Zhou Z, Alikhan NF, Mohamed K, Achtman M, The Agama Study Group. The user’s guide to comparative genomics with Enter- oBase. Three case studies: micro-clades within Salmonella enterica serovar Agama, ancient and modern populations of Yersinia pestis, and core genomic diversity of all Escherichia. bioRxiv 2019.

35. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells and metagenomes. Genome Res 2015;25:1043−1055.
Ooka et al., Microbial Genomics 2019:5

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as ‘excellent’ or ‘very good’.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.