



# Evidence for Contemporary Switching of the O-Antigen Gene Cluster between Shiga Toxin-Producing *Escherichia coli* Strains Colonizing Cattle

Lutz Geue<sup>1</sup>, Christian Menge<sup>1</sup>, Inga Eichhorn<sup>2</sup>, Torsten Semmler<sup>2,3</sup>, Lothar H. Wieler<sup>2,3</sup>, Derek Pickard<sup>4</sup>, Christian Berens<sup>1</sup> and Stefanie A. Barth<sup>1\*</sup>

<sup>1</sup> Friedrich-Loeffler-Institut/Federal Research Institute for Animal Health, Institute of Molecular Pathogenesis, Jena, Germany,

<sup>2</sup> Institute of Microbiology and Epizootics, Free University Berlin, Berlin, Germany, <sup>3</sup> Robert Koch Institute, Berlin, Germany,

<sup>4</sup> Wellcome Trust Sanger Institute, Pathogen Genomics, Cambridge, UK

## OPEN ACCESS

### Edited by:

Sergey M. Stolyar,  
University of Idaho, USA

### Reviewed by:

Ruth Serra-Moreno,  
Texas Tech University, USA

Patrick Fach,

French Agency for Food,  
Environmental and Occupational  
Health & Safety (ANSES), France

### \*Correspondence:

Stefanie A. Barth  
stefanie.barth@fli.de

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 03 February 2017

**Accepted:** 28 February 2017

**Published:** 21 March 2017

### Citation:

Geue L, Menge C, Eichhorn I,  
Semmler T, Wieler LH, Pickard D,  
Berens C and Barth SA (2017)  
Evidence for Contemporary Switching  
of the O-Antigen Gene Cluster  
between Shiga Toxin-Producing  
*Escherichia coli* Strains Colonizing  
Cattle. *Front. Microbiol.* 8:424.  
doi: 10.3389/fmicb.2017.00424

Shiga toxin-producing *Escherichia coli* (STEC) comprise a group of zoonotic enteric pathogens with ruminants, especially cattle, as the main reservoir. O-antigens are instrumental for host colonization and bacterial niche adaptation. They are highly immunogenic and, therefore, targeted by the adaptive immune system. The O-antigen is one of the most diverse bacterial cell constituents and variation not only exists between different bacterial species, but also between individual isolates/strains within a single species. We recently identified STEC persistently infecting cattle and belonging to the different serotypes O156:H25 ( $n = 21$ ) and O182:H25 ( $n = 15$ ) that were of the MLST sequence types ST300 or ST688. These STs differ by a single nucleotide in *purA* only. Fitness-, virulence-associated genome regions, and CRISPR/CAS (clustered regularly interspaced short palindromic repeats/CRISPR associated sequence) arrays of these STEC O156:H25 and O182:H25 isolates were highly similar, and identical genomic integration sites for the *stx* converting bacteriophages and the core LEE, identical Shiga toxin converting bacteriophage genes for *stx1a*, identical complete LEE loci, and identical sets of chemotaxis and flagellar genes were identified. In contrast to this genomic similarity, the nucleotide sequences of the O-antigen gene cluster (O-AGC) regions between *galF* and *gnd* and very few flanking genes differed fundamentally and were specific for the respective serotype. Sporadic aEPEC O156:H8 isolates ( $n = 5$ ) were isolated in temporal and spatial proximity. While the O-AGC and the corresponding 5' and 3' flanking regions of these aEPEC isolates were identical to the respective region in the STEC O156:H25 isolates, the core genome, the virulence associated genome regions and the CRISPR/CAS elements differed profoundly. Our cumulative epidemiological and molecular data suggests a recent switch of the O-AGC between isolates with O156:H8 strains having served as DNA donors. Such O-antigen switches can affect the evaluation of a strain's pathogenic and virulence potential, suggesting that NGS methods might lead to a more reliable risk assessment.

**Keywords:** *E. coli*, STEC, O-antigen gene cluster, recombination, genome sequencing

## INTRODUCTION

Shiga toxin-producing *Escherichia coli* (STEC) comprise a group of zoonotic enteric pathogens (Nataro and Kaper, 1998). The main reservoirs for STEC strains are ruminants, with cattle in particular. In humans, STEC infection may result in diarrhea, frequently complicated by the onset of hemorrhagic colitis (HC), or several renal and neurological sequelae, including the hemolytic uremic syndrome (HUS; Griffin and Tauxe, 1991; Su and Brandt, 1995; Paton and Paton, 1998; Remuzzi and Ruggenti, 1998). Factors contributing to the virulence of STEC strains causing human disease, also referred to as enterohemorrhagic *E. coli* (EHEC), include two major phage-encoded toxins, Shiga toxin 1 (Stx1) and 2 (Stx2), which can be produced and secreted by different strains individually or in combination. STEC may additionally possess virulence characteristics such as the ability to cause attaching-and-effacing (AE) lesions in the large intestine (McKee et al., 1995), and a large plasmid encoding for an enterohemolysin (*hlyA/ehxA*), a catalase-peroxidase (*katP*), and an extracellular serine protease (*espP*; Schmidt et al., 1995; Brunder et al., 1996, 1997, 1999).

Lipopolysaccharide (LPS), a major component of the outer membrane, represents the principal virulence factor of gram-negative bacteria (reviewed in Lerouge and Vanderleyden, 2002). It consists of three distinct regions: lipid A, core oligosaccharide, and O-specific polysaccharides (O-antigens; Lerouge and Vanderleyden, 2002; Samuel and Reeves, 2003). While lipid A is the main driver of inflammatory responses, O-antigens are instrumental for host colonization, and bacterial niche adaptation (Reeves, 1995). O-antigens are highly immunogenic and, therefore, targeted by the adaptive immune system. Recognition by antibodies, e.g., initiates the classical complement pathway resulting in bacterial cell death or increased phagocytosis by cells of the host defense (Reeves, 1995). Due to this strong selective pressure, the O-antigen is one of the most variable bacterial cell constituents, with variation in the types of sugars present, their arrangement within the O-unit, and the linkages between O-units (Wang et al., 2001; Bazaka et al., 2011). Variation not only exists between different bacterial species, but also between individual clones within a single species (Penner and Aspinall, 1997; Stenutz et al., 2006; Lam et al., 2011; Liu et al., 2014). The existence of more than 180 O-antigens has been proposed so far for *E. coli* (Wang et al., 2001). This high variability is made use of in clinical and food microbiology and in epidemiology by applying serotyping for infection chain tracing and risk assessment of STEC strains isolated from patients or food for human consumption.

Three mechanisms for biosynthesis of the O-antigen seem to exist (Samuel and Reeves, 2003). The one most frequently employed by *E. coli* is the Wzy/Wzx-dependent pathway. The genes involved in O-antigen biosynthesis generally form an O-antigen gene cluster (O-AGC) in the chromosome. In *E. coli*, this cluster is flanked by the colanic acid biosynthesis gene cluster (*wca* genes) and the histidine biosynthesis (*his*) operon (Iguchi et al., 2015). The sequences of many O-AGC were used to determine the genetic basis of O-antigen evolution (Samuel and Reeves, 2003; Samuel et al., 2004). The results

show that the O-AGC sections located between the *gnd* and *galF* genes have G+C contents lower (usually <40% in *E. coli* and *Salmonella enterica*) compared to the usual 51% genome average G+C content (Samuel et al., 2004). This atypical G+C content indicates that the O-AGC was acquired by interspecies horizontal gene transfer (HGT). Such an interspecies exchange was described for the O8 and O9 O-antigens of *E. coli*, as they are identical to O5 and O3 of *Klebsiella pneumoniae*, respectively (Sugiyama et al., 1997). An intraspecies switch in *Vibrio cholerae* from serogroup O1 to the novel serotype O139 was proposed to have been the initial event at the advent of a cholera epidemic in Asia (Mekalanos et al., 1997). This strain arose by HGT from a strain closely related to the pandemic *V. cholerae* O1 El Tor (Bik et al., 1995). Likewise, the acquisition in *E. coli* of the O157 O-AGC by an O55:H7 strain to generate the O157:H7 clone is considered critical for the evolution of this pandemic food-borne pathogen (Tarr et al., 2000; Wang et al., 2002). Besides having potential implications for the host range and virulence of clones of gram-negative bacteria, serotype switches interfere with serotype-based epidemiologic approaches to unveil infection chains and may even impact on the reliability of diagnostic workflows when these are shaped by serotype-based risk assessments. Here, we present evidence by integrating epidemiological and whole genome sequence data that an O-AGC cluster switch recently occurred between STEC and atypical enteropathogenic *E. coli* (aEPEC) clones of the serotypes O182:H25, O156:H25, and O156:H8 in a cattle herd (Geue et al., 2002; Barth et al., 2016).

## MATERIALS AND METHODS

### Bacterial Isolates

During a longitudinal study investigating the prevalence of Shiga toxin-producing *E. coli* (STEC) in cattle (Geue et al., 2002), STEC isolates belonging to the serotypes O156:H25 ( $n = 21$ ) and O182:H25 ( $n = 15$ ) were isolated (Table 1). Isolates of the serotype O156:H25 were present in three of four different herds tested (groups from farms B, C, and D1/D2; Geue et al., 2010), similar to isolates of serotype O182:H25 (groups from farms A, B, and D1/D2). Time periods in which STEC of both serotypes were isolated overlapped in farms B and D1/D2. Additionally, STEC isolates of serotype O156:H8 ( $n = 5$ ) were found in farms A, B and D2 (Table 1). Other STEC isolates with H25 chemotaxis and flagella genes (O51:H25 [ $n = 1$ ], O153:H25 [ $n = 3$ ], O165:H25 [ $n = 22$ ], O172:H25 [ $n = 4$ ], O177:H25 [ $n = 6$ ], ONT:H25 [ $n = 1$ ]), that had been isolated during the same longitudinal study, were included in the current study.

### Whole Genome Sequencing

Genomic DNA of the *E. coli* isolates was prepared using the ZR fungal/bacterial DNA kit (Zymo Research Europe GmbH, Freiburg, Germany) from overnight cultures in Luria Bertani broth following the instructions of the manufacturer. The DNA concentration was determined spectrophotometrically at 260 nm and analyzed for fragmentation by 1% TBE agarose gel electrophoresis. All isolates were whole genome sequenced using Illumina MiSeq 300 bp paired-end sequencing and a coverage

**TABLE 1 | Overview of strains included in this study, their origin, time of isolation and selected genetic properties.**

Isolate	Serotype	Sequence type	Farm	No. of animal [DD.MM.YYYY]	Sampling day	stx phage	Phage integration site	eae subtype	LEE integration site	T2SS	espZ-9 nuc-insertion	efa1/lifA-like	AidA-I adhesin like	bcsA gene insertion of 84 nt	10kb island phage like	38.4 kb phage (for ImpK, OmpA)	cad operon (10 kb)	Same locus	
																		<i>pgaABCD</i> operon (biofilm adhesin)	<i>efeJOB</i> operon (low pH iron transporter)
13E0602	O182:H25	300	A	9	23.02.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	-	-	-	-	+	+
13E0644	O182:H25	300	A	26	23.03.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	-	-	-	-	+	+
13E0646	O182:H25	300	A	26	23.03.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	-	-	-	-	+	+
13E0647	O182:H25	300	A	26	23.03.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	-	-	-	-	+	+
13E0648	O182:H25	300	A	26	23.03.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	-	-	-	-	+	+
13E0701	O182:H25	300	B	4	16.03.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	-	-	-	-	+	+
13E0725	O182:H25	300	B	20	07.12.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	-	+	+	+	-	-	-	-	+	+
13E0726	O182:H25	300	B	20	07.12.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	-	+	+	+	-	-	-	-	+	+
13E0728	O182:H25	300	B	20	07.12.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	-	+	+	+	-	-	-	-	+	+
13E0729	O182:H25	300	B	23	17.08.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	-	+	+	+	-	-	-	-	+	+
13E0798	O182:H25	300	D1	11	20.07.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	-	+	-	-	+	+
13E0799	O182:H25	300	D1	11	20.07.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	-	+	+	+	-	-	-	-	+	+
13E0820	O182:H25	300	D1	17	30.03.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	-	+	-	-	+	+
13E0862	O182:H25	300	D2	6	18.01.1999	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	-	-	-	-	+	+
13E0879	O182:H25	300	D2	12	15.02.1999	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	-	-	-	-	+	+
13E0702	O156:H25	300	B	4	16.03.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	+	+	+	+	-	-
13E0730	O156:H25	300	B	23	17.08.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	+	+	-	+	-	-
13E0732	O156:H25	300	B	23	17.08.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	+	+	-	+	-	-
13E0733	O156:H25	300	B	23	17.08.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	-	+	-	+	-	-
13E0753	O156:H25	688	B	25	10.11.1997	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	-	+	-	+	+	+
13E0755	O156:H25	688	B	25	10.11.1997	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	-	-	-	-	-	+	+	+	+	+
13E0757	O156:H25	688	B	25	10.11.1997	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	-	+	+	+	+	+
13E0758	O156:H25	300	B	25	27.07.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	+	+	-	+	-	-
13E0759	O156:H25	688	B	25	27.07.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	-	+	+	+	+	+
13E0764	O156:H25	300	B	25	09.11.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	+	+	-	+	-	-
13E0776	O156:H25	688	B	26	08.09.1997	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	-	+	-	+	+	+
13E0778	O156:H25	688	B	28	06.07.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	+	+	+	+	-	-
13E0780	O156:H25	300	B	28	06.07.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	+	+	+	+	-	-
13E0781	O156:H25	300	B	28	06.07.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	+	+	+	+	-	-
13E0801	O156:H25	300	D1	12	10.08.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	+	+	+	+	-	-
13E0802	O156:H25	300	D1	12	10.08.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	+	+	+	+	-	-
13E0803	O156:H25	300	D1	12	10.08.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	+	-	+	-	-	-
13E0845	O156:H25	300	C	25	09.02.1998	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	+	+	+	-	+	+
13E0848	O156:H25	300	D2	2	18.01.1999	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	+	+	-	+	-	-
13E0850	O156:H25	300	D2	2	18.01.1999	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	+	-	-	-	+	+	-	+	-	-
13E0902	O156:H25	300	D2	25	12.04.1999	1a	<i>mlrA/lyeU</i>	$\zeta$	<i>pheU/pheV</i>	-	-	-	-	+	+	-	+	-	-
13E0685	O156:H8	327	A	29	23.03.1998	-	-	$\vartheta$	<i>ileX</i>	+	-	-	-	+	-	-	+	+	+
13E0767	O156:H8	327	B	26	07.12.1998	-	-	$\vartheta$	<i>ileX</i>	+	-	-	+	+	-	-	+	+	+
13E0775	O156:H8	327	B	26	07.12.1998	-	-	$\vartheta$	<i>ileX</i>	+	-	-	+	+	-	-	+	+	+
13E0883	O156:H8	327	D2	12	31.05.1999	-	-	$\vartheta$	<i>ileX</i>	+	-	-	+	+	-	-	+	+	+
13E0890	O156:H8	327	D2	23	15.03.1999	-	-	$\vartheta$	<i>ileX</i>	+	-	-	+	+	-	-	+	+	+

>40× was obtained. The sequence read data was first subjected to quality control using the NGS toolkit (Patel and Jain, 2012). Reads with a minimum of 70% of bases having a phred score of >20 were defined as high quality reads. *De novo* assembly of resulting high quality filtered reads into contiguous sequences (contigs and scaffolds) was achieved using CLC Genomics Workbench 8.0 (CLC bio, Aarhus, Denmark).

One strain per serotype (O182:H25 [13E0725], O156:H25 [13E0780], O156:H8 [13E0767]) was additionally whole genome sequenced on a PacBio RSII system (Pacific Biosciences, USA) by a commercial service provider (GATC Biotech, Konstanz, Germany) utilizing PacBio single-molecule real-time (SMRT) technology. Subsequent *de novo* assembly utilizing the HGAP3 protocol yielded a single polished contig with 200-fold average reference coverage. In order to ensure closed circle conformation of the bacterial chromosome, mapping, sequence analyses, and annotation were carried out using the commercial software package Geneious (version 9.1.6, Biomatters Ltd., Auckland, New Zealand). The whole genome alignments were performed by MAUVE analysis (version 2.3.1; Darling et al., 2004) as plugin in the Geneious software package.

A Maximum Likelihood tree (GTR+G+I substitution model, 1000 bootstraps) was calculated from 53 previously described genes associated with chemotaxis and flagella of flagellar serotype H25 (Sperandio, 2001; Niba et al., 2007; Table S1 in Supplemental Material) of eight different O-antigen serotypes (O51, O153, O156, O165, O172, O177, O182, ONT) in PAUP\* (version 4.0b10). The same genes of the flagellar serotype H8 were used as outgroup.

## Ethics Statement

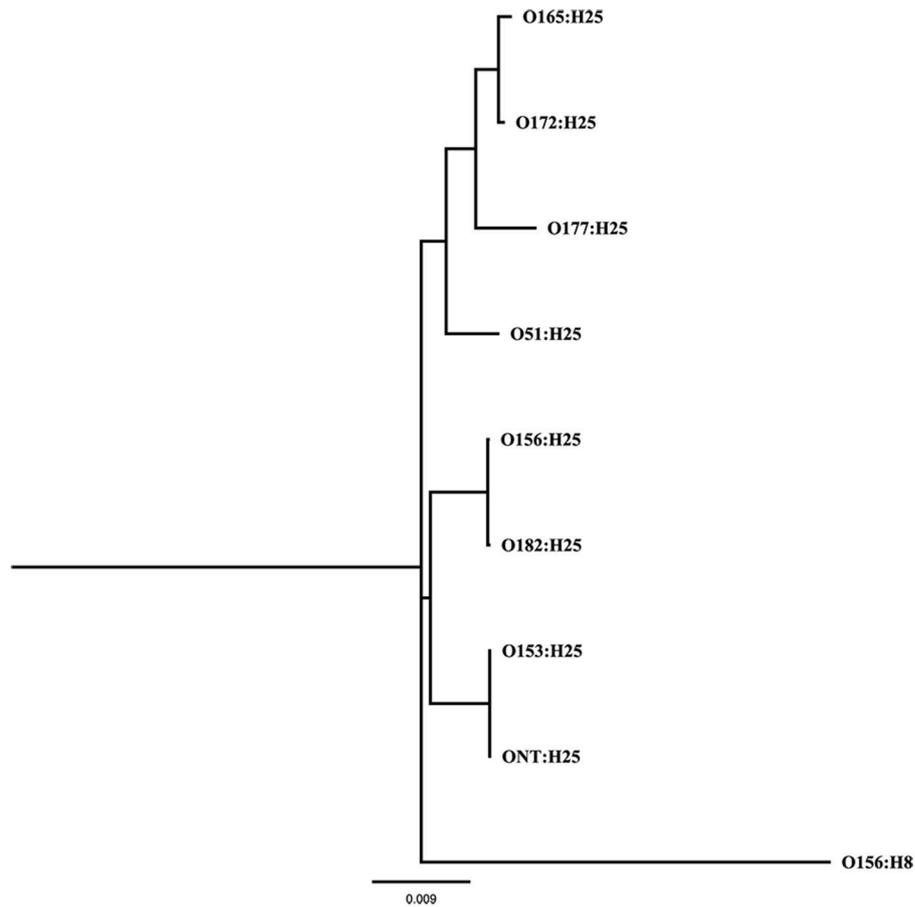
An Ethics Statement is not necessary. The isolates were obtained by non-invasive rectal swabs during a longitudinal study already published (Geue et al., 2002). No animal experiments were carried out for this study.

## RESULTS AND DISCUSSION

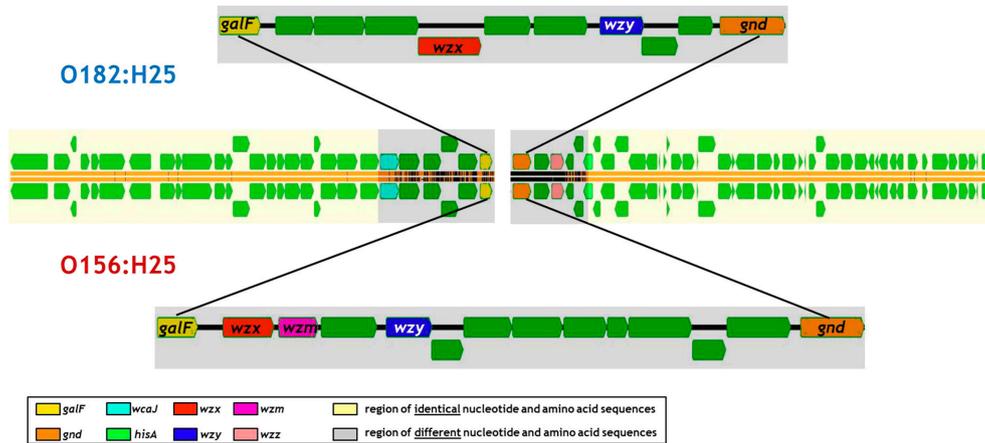
A previous study on STEC colonization in cattle herds identified specific STEC clones, which could be isolated from herds over extended periods of time and were therefore considered as persistently colonizing this animal reservoir (Geue et al., 2002; Barth et al., 2016). These strains expressed the flagellar serotype H25 but differed in O-antigen serotypes (O156, O165, O182). To assess the underlying genetic basis, we performed whole genome sequence analyses of these *E. coli* isolates. By MLST (Wirth et al., 2006), all O182:H25 and 15 of the O156:H25 isolates were assigned to ST300. The remaining 6 O156:H25 isolates were allocated to ST688. These STs differ from each other by a single nucleotide in *purA* only (Barth et al., 2016). By contrast, the O165:H25 isolates were classified as ST119 which is widely separated from ST300/ST688 (Barth et al., 2016).

To further analyze that the genomic similarity of O156:H25 and O182:H25 isolates, we assessed the presence and relatedness of selected fitness- and virulence-associated genes. First, we compared 53 previously described genes associated with

chemotaxis and flagella of H25 (Sperandio, 2001; Niba et al., 2007; Table S1 in Supplemental Material). A 100 % identity was detected for the nucleotide and amino acid sequences. Of the 45,429 nucleotides studied, only seven nucleotides differed between the strains of the two serotypes. They were located in the *flgD*, *flgI*, *flhA*, *fliG*, *fliZ*, *motA*, and *tar* genes and resulted in one amino acid exchange each in FlhA, FliG, and MotA, respectively. In contrast, H25 chemotaxis and flagellar genes from isolates belonging to other O serogroups (O51:H25, O153:H25, O165:H25, O172:H25, O177:H25, ONT:H25) exhibited larger genetic distances (between 97.2 and 99% identity, **Figure 1**). Additionally, a region spanning ~60 kb, which includes the complete core regions of the locus of enterocyte effacement (LEE) and the LEE insertions sites, was compared in all O156:H25 and O182:H25 isolates. The core regions were nearly identical (99.9–100%) in all isolates probed. Less than 40 nucleotides differed in the ~33,000 nucleotides considered. The main difference detected was a nine nucleotide insertion in the *espZ* gene of 5 of the 15 O182:H25, but in none of the O156:H25 isolates. Identical  $\zeta$  *eae* genes were found in all O156:H25 and O182:H25 isolates and the LEE was inserted at the same *pheU/pheV* tRNA site in all isolates. The sequences of the 5' and 3' regions flanking the integration sites were identical in all O156:H25 and O182:H25 isolates. The *stx*-converting bacteriophages found in all O156:H25 and O182:H25 isolates encode the Stx subtype 1a and possess identical gene sequences for the A and B subunits. Phage genomes were all integrated between the *mlrA* (*yheY*) and *yheU* genes. In comparison to *E. coli* MG1655 (accession no. U00096.2), IAI1 (accession no. NC\_011741.1), and HUSEC2011 (accession no. HF572917.2), the first 160 nucleotides of the coding sequence of the *mlrA* gene, associated with the regulation of curli synthesis in *E. coli* and *Salmonella enterica* (Brown et al., 2001), were lacking in all O156:H25 and O182:H25 isolates. Regarding the clustered, regularly interspaced, short palindromic repeat (CRISPR) acquired immune system, which is being used for determining the evolutionary divergence of *E. coli* isolates, especially for closely related strains (Touchon et al., 2011; Yin et al., 2013), we identified a set of identical CRISPR associated sequence type E (CAS-E) genes adjacent to the CRISPR2.1 locus (nomenclature as described by Diez-Villasenor et al., 2010) between the *cysH* and *iap* genes in all 21 O156:H25 isolates. The STEC O182:H25 isolate 13E0725, which had been sequenced by PacBio RSII, also contained a 100% identical CAS-E region. These CAS genes were also found in the other 14 O182:H25 isolates, but the quality of the Illumina sequence data was not sufficient to allow an unambiguous sequence assignment for the entire region. The 10 repeats and the 9 spacers of the CRISPR2.1 loci were 100% identical in all O156:H25 and O182:H25 isolates. An additional CRIPR2.2-3 array, lacking CAS genes, was detected between *queE* (*ygcF*) and *ygcE* in all O156:H25 and O182:H25 isolates. Here, too, the identity of the 7 repeats and the 6 spacers was 100% in all isolates. Besides the CRISPR2 loci, we also detected a CRISPR4.1-2 array situated between the genes encoding *clpA* and *infA*. This array also lacked CAS genes, but the entire 684 bp sequence between the stop codons of *infA* and *clpA* contained two repeats flanking a single spacer element.



**FIGURE 1 |** Maximum Likelihood tree (GTR+G+I substitution model, 1,000 bootstraps) calculated from 53 genes associated with chemotaxis and flagella of flagellar serotype H25 of bovine STEC strains belonging to eight different O-antigen serotypes (O51, O153, O156, O165, O172, O177, O182, ONT); the same genes of the flagellar serotype H8 were used as outgroup.



Geneious, version 9.1.6

**FIGURE 2 |** Comparison of the architecture of the O-antigen gene clusters in bovine STEC strains of the O156:H25 (isolate 13E0780) and the O182:H25 (13E0725) serotype consecutively isolated on the same farm.

Its sequences were identical in the 21 O156:H25 and the 15 O182:H25 isolates and also in the *E. coli* reference strain MG1655 (accession no. U00096.2).

When analyzing the O-AGC and the corresponding 5' and 3' flanking regions of the O156:H25 and O182:H25 isolates, it became apparent that both O-AGC serotypes deploy the Wzy/Wzx-dependent pathway. The G+C content of the O-AGC part located between the *galF* and *gnd* genes was <40% (35.8% for O156:H25 and 34.0% for O182:H25). Similar low G+C-values have been described previously by Samuel and coworkers for other *E. coli* and *Salmonella enterica* strains (Samuel et al., 2004). However, the nucleotide sequences of the regions between the *galF* and *gnd* genes were very different between the two serotypes (45.7% identity). The size of the O-AGC of the O156:H25 isolates amounted to 13,260 bp. In contrast, the region between *galF* and *gnd* of the O182:H25 isolates was only 9,861 bp long. Both the number and the order of genes differed substantially between the O-AGCs (Figure 2). Furthermore, the *galF* genes and the next five genes upstream of *galF* varied noticeably in their nucleotide sequences (between 9 and 59 different nucleotides) although the differences at the amino acid sequence level were considerably lower (only 1–5 amino acid exchanges, Figure 2, Table 2). Further upstream, all corresponding genes of the O156:H25 and O182:H25 isolates were identical. Likewise, the *gnd* genes and the next five genes downstream of *gnd* differed significantly in their nucleotide and amino acid sequences (between 7 and 182 nucleotides, 2–51 amino acids). Further downstream, all corresponding genes were again identical in the O156:H25 and O182:H25 isolates.

Taken together, the results of both core and accessory genome analyses based on Illumina sequence data proved a great degree of similarity between the genomes of the O156:H25 and the O182:H25 isolates. Only the genomic regions encoding the O-AGC between the *galF* and *gnd* genes and very few genes flanking these regions, but still part of the O-AGC, varied substantially and were specific for the respective serotype. This could suggest a possible switch of the O-AGC between isolates, and the epidemiological data substantiated this hypothesis, despite the O-AGC not having been specifically selected for in the original strain isolation (Geue et al., 2002). Isolates of both serotypes were isolated on the same farms and in identical sampling periods. On farm B, O156:H25 and O182:H25 isolates were even detected on the same day in the same group (March of the 2nd study year) and, on one occasion, also on the same day in the same animal (August of the 2nd study year; Table 1).

To further substantiate these findings, one isolate per serotype was randomly picked for 3rd generation whole-genome sequencing on a PacBio RSII system. Closed circle conformations of the bacterial chromosomes and their annotation were performed. By MAUVE analysis the arrangement of homologous sequence blocks was found to be very similar in both isolates (Figure 3; Darling et al., 2004). A difference was observed in the size of the genomes. The genome of O156:H25 isolate 13E0780 had a size of 5,371,291 nucleotides, whereas the one of O182:H25 isolate 13E0725 comprised of only 5,112,484 nucleotides. The difference was mainly due to the presence of an additional pathogenic island with a type II secretion system and an additional phage-like sequence in the STEC isolate 13E0780.

**TABLE 2 | Comparison of nucleotide and amino acid sequences as well as the GC content of genes flanking the O-antigen cluster in bovine STEC strains belonging to the O156:H25 ( $n = 21$  strains) and the O182:H25 serotype ( $n = 15$  strains).**

No.	Gene	% Identity nucleotide sequence (no. of mismatches)	% GC content	% Identity amino acid sequence (no. of exchanges)
1	<i>yegE</i>	100	53.8	100
2	<i>udk</i>	100	48.0	100
3	<i>dcd</i>	100	60.3	100
4	<i>asmA</i>	99.9 (1 in all isolates)	52.8–52.9	100
5	<i>yegH</i>	100	53.1	100
6	<i>wza</i>	99.9 (1 in all isolates)	53.8	100
7	<i>wzb</i>	99.8–100 (1 in 4 of 21 O156:H25 isolates)	54.2–54.3	99.3–100 (1 in 4 of 21 O156:H25 isolates)
8	<i>wzc</i>	100	52.5	100
9	<i>wcaA</i>	100	51.0	100
10	<i>wcaB</i>	99.8–100 (1 in 15 of 21 O156:H25 isolates)	53.1	100
11	<i>wcaC</i>	100	54.4	100
12	<i>wcaD</i>	100	42.4	100
13	<i>wcaE</i>	100	45.9	100
14	<i>wcaF</i>	100	47.5	100
15	<i>gmd</i>	100	55.1	100
16	<i>fcl</i>	100	56.3	100
17	<i>gmm</i>	100	55.7	100
18	<i>wcaI</i>	100	55.7	100
19	<i>cpsB</i>	99.9–100 (1 in 5 of 15 O182:H25 isolates)	55.3–55.4	99.8–100 (1 in 5 of 15 O182:H25 isolates)
20	<i>cpsG</i>	100	55.2	100
21	<i>wcaJ</i>	99.4 (9 in all isolates)	53.8	100
22	<i>wzxC</i>	97.3 (40 in all isolates)	54.1	99.2 (2 in all isolates)
23	<i>wcaK</i>	96.2 (49 in all isolates)	54.5	98.8 (4 in all isolates)
24	<i>wcaL</i>	95.2 (59 in all isolates)	54.5	99.0 (4 in all isolates)
25	<i>wcaM</i>	96.0 (56 in all isolates)	47.6	98.3 (5 in all isolates)
26	<i>galF</i>	96.5 (31 in all isolates)	50.7	99.7 (1 in all isolates)
<b>O-antigen gene cluster</b>				
27	<i>gnd</i>	95.4 (65 in all isolates)	50.6	99.6 (2 in all isolates)
28	UDP-glucose 6-dehydrogenase gene	86.7 (182 + gaps in all isolates)	43.6	89.2 (51 + gaps in all isolates)
29	<i>wzz</i>	94.1 (112 + gaps in all isolates)	47.7	97.5 (26 + gaps in all isolates)
30	<i>hisE</i>	96.6 (21 in all isolates)	53.3	98.5 (3 in all isolates)
31	<i>hisF</i>	97.2 (22 in all isolates)	52.2	98.8 (3 in all isolates)
32	<i>hisA</i>	99.1 (7 in all isolates)	56.6	99.6 (1 in all isolates)

(Continued)

TABLE 2 | Continued

No.	Gene	% Identity nucleotide sequence (no. of mismatches)	% GC content	% Identity amino acid sequence (no. of exchanges)
33	<i>hisH</i>	100	56.3	100
34	<i>hisB</i>	100	53.7	100
35	<i>hisC</i>	100	54.7	100
36	<i>hisD</i>	100	57.1	100
37	<i>hisG</i>	100	54.6	100
38	<i>hisL</i>	100	43.1	100
39	<i>yefM</i>	100	47.6	100
40	<i>yoeB</i>	100	44.3	100
41	<i>yeeZ</i>	100	51.5	100
42	<i>yeeY</i>	100	50.6	100
43	<i>yeeF</i>	100	52.7	100
44	<i>yeeE</i>	100	50.5	100
45	Hypothetical protein gene (1)	100	51.3	100
46	<i>sbcB</i>	100	50.7	100
47	<i>dacD</i>	99.9–100 (1 in 7 of 15 O182:H25 isolates)	49.3	99.9–100 (1 in 7 of 15 O182:H25 isolates)
48	<i>sbmC</i>	100	46.0	100
49	<i>yeeA</i>	100	50.1	100
50	Hypothetical protein gene (2)	100	47.6	100
51	<i>yoeF</i>	100	45.0	100
52	<i>cobU</i>	100	50.8	100
53	<i>cobS</i>	100	53.5	100
54	<i>cobT</i>	100	52.6	100
55	<i>erfK</i>	100	49.9	100
56	Transcriptional regulator, DeoR family gene	100	44.7	100
57	Putative PTS IIA-like nitrogen-regulator protein PtsN gene	100	43.6	100
58	PTS system, galactitol-specific IIB component gene	100	43.2	100
59	Sugar-phosphate isomerase, RpiB/LacA/LacB family gene	100	45.6	100
60	Transcriptional regulator, DeoR family gene 2	99.9–100 (1 in 14 of the 21 O156:H25 isolates), 99.7 (2 in 1 of O156:H25 isolates)	43.9–44.0	99.9–100 (1 in 14 of the 21 O156:H25 isolates), 99.7 (2 in 1 of O156:H25 isolates)
61	Putative PTS IIA-like nitrogen-regulatory protein PtsN gene 2	99.8 (1 in all isolates)	42.2	100

(Continued)

TABLE 2 | Continued

No.	Gene	% Identity nucleotide sequence (no. of mismatches)	% GC content	% Identity amino acid sequence (no. of exchanges)
62	PTS system, galactitol-specific IIC component gene	100	43.5	100
63	Ribulose-phosphate3-epimerase gene	100	38	100
64	Short-chain dehydrogenase/reductase SDR gene	100	42.6	100
65	Sugar isomerase gene	100	43.3	100
66	<i>nac</i>	100	48.8	100

A peculiarity of STEC isolate 13E0725 was the presence of a pathogenic island with an *efa1/lifA*-like gene and a gene encoding an AidA-I adhesion like protein. Upon inspection of the Illumina whole genome data of all STEC O182:H25 and STEC O156:H25 strains, the type II secretion system was missing in strain 13E0725 and in three other STEC O182:H25 strains, but it was also not present in 3 of the 21 STEC O156:H25 strains. In contrast, the pathogenicity island harboring an *efa1/lifA*-like gene and a gene for an AidA-I adhesion like protein was detected in only 5 of the 15 STEC O182:H25 strains and in none of the STEC O156:H25 strains (Table 1). From integrating molecular and epidemiological data it is tempting to conclude that a switch of the O-AGC might have occurred between the STEC O182:H25 and the O156:H25 clone present in the sampled cattle herd.

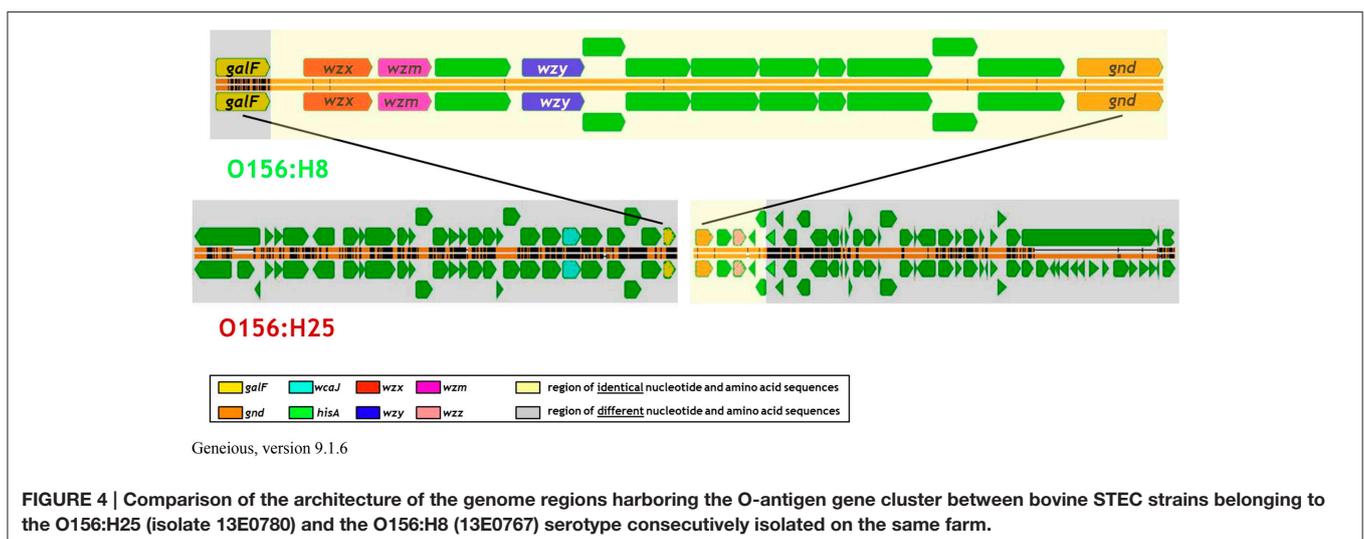
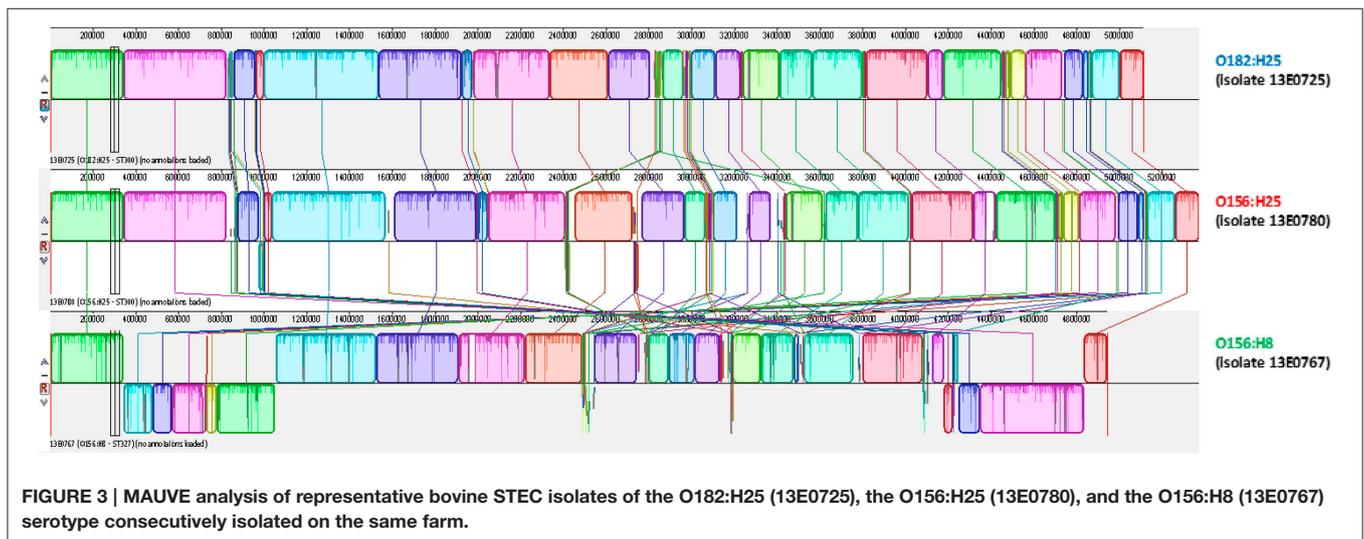
Noteworthy, five O156:H8 isolates were isolated during the same investigation period and on the same farms B and D. In contrast to the O156:H25 isolates, these isolates were typed as ST327 by MLST analysis. This sequence type is very different from ST300/688 (six of seven alleles different). Also, the virulence associated genome regions were significantly different from the O156:H25/O182:H25 isolates. The O156:H8 isolates are atypical enteropathogenic *E. coli* (aEPEC; Hernandez et al., 2009), lacking both *stx*-bacteriophages as well as *bfpA*. Also the LEE locus differed from the one present in the O156:25/O182:H25 isolates. For example, *eae* genes for a  $\vartheta$  intimin were detected and the LEE of all O156:H8 isolates was inserted in the *ileX* tRNA site. As in the case of STEC O156:H25 and O182:H25, one random aEPEC O156:H8 isolate (13E0767) was whole-genome sequenced using 3rd generation sequencing. The sequences were aligned in a MAUVE analysis together with both STEC isolates. The orientation and the sequence of genome blocks varied distinctly in comparison to O156:H25/O182:H25 (Figure 3). With respect to the CRISPR/CAS systems in the five aEPEC O156:H8 isolates, most of the CAS gene array is missing in the CRISPR2.1 locus. These isolates have a deletion spanning from the second nucleotide of the codon encoding for Gly683 in the

*cas3* gene to the last nucleotide (nt 29) of a CRISPR2 repeat element. The truncated CAS3 protein lacks 169 residues at its C-terminus and another C-terminal 48 residues are mutated due to the frameshift. The number of repeat and spacer elements also differed in contrast to the O156:H25/O182:H25 isolates. Only seven repeats and six spacers were found. Compared to the O156:H25/O182:H25 isolates, all O156:H8 isolates carried a deletion of ca. 12,000 nt ranging from *queD* right up to the CRISPR2.2-3 array. A smaller number of six repeats and five spacers was found in this locus. We also detected a CRISPR4.1-2 array in the O156:H8 isolates. Its two repeats are sequence-identical with the O156:H25/O182:H25 repeats, but the spacer's sequence differs.

Analysis of the O-AGC and the corresponding 5' and 3' flanking regions of the O156:H8 isolates revealed that the region between *galF* and *gnd* of the O156:H8 isolates was identical to the respective region in the STEC O156:H25 isolates (Figure 4). Only 7 of the 14,154 nucleotides differed between the two serotypes in

five different genes. Thereof, two nucleotide exchanges resulting in two amino acid exchanges were found in the *wzx* gene. One nucleotide exchange each was detected in the glycosyl transferase genes *wfeX*, *wfeY*, and in the *manB* gene. These nucleotide exchanges caused one amino acid exchange in *WfeY*. In contrast, the *galF* genes and all genes upstream of *galF* varied distinctly in their nucleotide and amino acid sequences (Figure 4, Table 3). The *gnd* genes and the next four genes downstream of *gnd* were identical in the O156:H8 and O156:H25 isolates. Further downstream in the genome, all genes differed substantially.

The results presented herein imply that specific persistent STEC isolates can replace their O-AGC to change their phenotype. We postulate that STEC isolates that originally had the serotype O182:H25 changed their O-AGC to become STEC O156:H25 with sporadic O156:H8 isolates having served as potential DNA donors. An intraspecies switch in *E. coli* from O55:H7 to O157:H7 was previously described (Tarr et al., 2000; Wang et al., 2002). Another intraspecies gene exchange



**TABLE 3 | Comparison of nucleotide and amino acid sequences as well as the GC content of genes flanking the O-antigen cluster in bovine STEC strains belonging to the O156:H25 (*n* = 21 strains) and the O156:H8 serotype (*n* = 5 strains).**

No.	Gene	% Identity nucleotide sequence (no. of mismatches)	% GC content	% Identity amino acid sequence (no. of exchanges)
1	<i>yegE</i>	Much smaller in H25 (probably two genes)	n.t.	n.t.
2	<i>udk</i>	99.2 (5 in all isolates)	48.4	99.5 (1 in all isolates)
3	<i>dcd</i>	99.5 (3 in all isolates)	60.2	100
4	<i>asmA</i>	98.7 (24 in all isolates)	52.6	99.8 (1 in all isolates)
5	<i>yegH</i>	98.8–99.0 (17 in 1 O156:H8 isolate, 16 in 4 other O156:H8 isolates)	58.2	99.8 (1 in all isolates)
6	<i>wza</i>	98.8 (14 in all isolates)	53.5	99.7 (1 in all isolates)
7	<i>wzb</i>	99.8 (1 in all O156:H8 isolates, 1 additional nucleotide in 4 of 21 O156:H25 isolates)	54.2	99.9–100 (1 in 4 of 21 O156:H25 isolates)
8	<i>wzc</i>	97.8 (47 in all isolates)	52.7	99.0 (7 in all isolates)
9	<i>wcaA</i>	98.0 (18 in 1 O156:H8 isolate, 17 in 4 other O156:H8 isolates)	51.1	99.3–99.6 (2 in 1 O156:H8 isolate, 1 in 4 other O156:H8 isolates)
10	<i>wcaB</i>	96.9 (14 in all O156:H8 isolates)	53.3	98.8 (2 in all isolates)
11	<i>wcaC</i>	99.3–99.5 (9 in 4 of 5 O156:H8 isolates, 6 in 1 other O156:H8 isolate)	54.1	99.5–99.8 (2 in 4 of 5 O156:H8 isolates, 1 in 1 other O156:H8 isolate)
12	<i>wcaD</i>	98.8 (15 in all isolates)	42.5	99.8 (1 in all isolates)
13	<i>wcaE</i>	97.1 (22 in all isolates)	45.9	98.4 (4 in all isolates)
14	<i>wcaF</i>	96.9–97.1 (16 in 1 O156:H8 isolate, 17 in 4 other O156:H8 isolates)	48.0	97.8–98.4 (3 in 1 O156:H8 isolate, 4 in 4 other O156:H8 isolates)
15	<i>gmd</i>	98.5–98.6 (16 in 1 O156:H8 isolate, 17 in 4 other O156:H8 isolates)	42.4	99.7 (1 in all isolates)
16	<i>fcl</i>	96.9 (30 in all isolates)	56.1	98.8 (4 in all isolates)
17	<i>gmm</i>	100	55.7	100
18	<i>wcaI</i>	99.4–99.5 (7 in 1 O156:H8 isolate, 6 in 4 other O156:H8 isolates)	55.7	99.5 (2 in all isolates)
19	<i>cpsB</i>	97.9 (30 in all isolates)	55.3	99.8 (1 in all isolates)
20	<i>cpsG</i>	96.9 (42 in all isolates)	55.2	98.9 (5 in all isolates)
21	<i>wcaJ</i>	97.2 (39 in all isolates)	53.7	99.4 (3 in all isolates)
22	<i>wzxC</i>	97.8 (32 in 4 O156:H8 isolates, 33 in 1 other isolate)	54.5	99.0 (5 in all isolates)
23	<i>wcaK</i>	97.7–97.8 (28 in 4 O156:H8 isolates, 29 in 1 other O156:H8 isolate)	54.6	98.8 (5 in 4 O156:H8 isolates, 6 in 1 other O156:H8 isolate)
24	<i>wcaL</i>	95.1 (60 in all isolates)	55.0	99.3 (3 in all isolates)
25	<i>wcaM</i>	98.5–98.6 (20 in 1 O156:H8 isolate, 21 in 4 other O156:H8 isolates)	47.4	98.7–98.9 (5 in 1 O156:H8 isolate, 6 in 4 other O156:H8 isolates)
26	<i>galF</i>	95.1 (44 in all isolates)	51.1	99.7 (1 in all isolates)
<b>O-antigen gene cluster</b>				
27	<i>gnd</i>	99.9–100 (1 in 4 O156:H8 isolates, 0 in 1 other O156:H8 isolate)	50.4	100
28	UDP-glucose 6-dehydrogenase gene	100	43.4	100
29	<i>wzz</i>	100	47.9	100
30	<i>hisE</i>	100	53.1	100
31	<i>hisF</i>	99.9 (1 in all isolates)	52.6	100
32	<i>hisA</i>	97.4 (19 in all isolates)	57.0	98.0 (5 in all isolates)
33	<i>hisH</i>	97.8 (13 in all isolates)	56.3	99.5 (1 in all isolates)
34	<i>hisB</i>	97.2–97.3 (29 in 1 O156:H8, 30 in 4 other O156:H8)	53.9	99.4 (2 in all isolates)
35	<i>hisC</i>	97.6 (26 in all isolates)	55.1	98.9 (4 in all isolates)
36	<i>hisD</i>	94.7 (69 in all isolates)	57.4	97.7 (10 in all isolates)
37	<i>hisG</i>	98.4–98.6 (13 in 4 O156:H8 isolates, 14 in 1 other O156:H8 isolate)	54.5	99.7 (1 in all isolates)
38	<i>hisL</i>	100	43.1	100
39	<i>yefM</i>	98.0 (5 in all isolates)	47.6	100

(Continued)

TABLE 3 | Continued

No.	Gene	% Identity nucleotide sequence (no. of mismatches)	% GC content	% Identity amino acid sequence (no. of exchanges)
40	<i>yoeB</i>	99.6 (1 in all isolates)	44.1	100
41	<i>yeeZ</i>	99.5 (4 in all isolates)	51.6	99.3 (2 in all isolates)
42	<i>yeeY</i>	100	50.6	100
43	<i>yeeF</i>	100	52.7	100
44	<i>yeeE</i>	99.6 (4 in all isolates)	50.5	99.4 (2 in all isolates)
45	Hypothetical protein gene (1)	100	51.3	100
46	<i>sbcB</i>	99.9 (2 in all isolates)	50.7	99.8 (1 in all isolates)
47	<i>dacD</i>	98.4 (19 in all isolates)	49.3	99.0 (4 in all isolates)
48	<i>sbmC</i>	98.3 (8 in all isolates)	46.3	100
49	<i>yeeA</i>	97.9 (22 in all isolates)	50.3	99.1 (3 in all isolates)
50	Hypothetical protein gene (2)	99.7 (1 in all isolates)	47.7	100
51	<i>yoeF</i>	94.2 (7 in all isolates)	45.0	87.2 (5 in all isolates)
52	<i>cobU</i>	100	50.8	100
53	<i>cobS</i>	99.9 (1 in all isolates)	53.4	100
54	<i>cobT</i>	98.1 (21 in all isolates)	52.5	98.3 (6 in all isolates)
55	<i>erfK</i>	97.2 (26 in all isolates)	50.1	97.4 (9 in all isolates)
56	Transcriptional regulator, DeoR family gene	Missing in all O156:H8	None	None
57	Putative PTS IIA-like nitrogen-regulator protein PtsN gene	Missing in all O156:H8	None	None
58	PTS system, galactitol-specific IIB component gene	Missing in all O156:H8	None	None
59	Sugar-phosphate isomerase, RpiB/LacA/LacB family gene	Missing in all O156:H8	None	None
60	Transcriptional regulator, DeoR family gene 2	Missing in all O156:H8	None	None
61	Putative PTS IIA-like nitrogen-regulatory protein PtsN gene 2	Missing in all O156:H8	None	None
62	PTS system, galactitol-specific IIC component gene	Missing in all O156:H8	None	None
63	Ribulose-phosphate3-epimerase gene	Missing in all O156:H8	None	None
64	Short-chain dehydrogenase/reductase SDR gene	Missing in all O156:H8	None	None
65	Sugar isomerase gene	Missing in all O156:H8	None	None
66	<i>nac</i>	98.7 (12 in all isolates)	48.7	99.7 (1 in all isolates)

was demonstrated in *V. cholerae* (Blokesch and Schoolnik, 2007). Natural transformation following addition of genomic DNA from an O139 donor strain to a competent O1 strain growing as a biofilm on a chitin surface was sufficient for exchange of the O1-AGC against the entire O139-AGC in a single transformation event. Such O-antigen switches can play important roles in at least two steps of the infection process (Lerouge and Vanderleyden, 2002). They can affect colonization via altered adherence and the recombinant strains also have different antigenic properties, which confers a selective advantage as it allows the strains to bypass or overcome host defense responses (Bik et al., 1995). Compared to O1 strains, *V. cholerae* O139 variants, for example, are resistant to an O1 lytic phage (Blokesch and Schoolnik, 2007), colonize a mouse

model with 2-fold higher efficiency (Waldor et al., 1994), are more invasive and damage the mucosal and submucosal layers more aggressively (Amin et al., 2009), and cause disease in persons with preexisting immunity to *V. cholerae* O1 (reviewed in Ramamurthy et al., 2003). In light of the epidemiological data presented herein and previously (Geue et al., 2002; Barth et al., 2016), it is tempting to assume that altered properties have also helped the STEC isolates studied to realize a more persistent lifestyle in the ruminant host.

The question arises how this O-AGC exchange occurred mechanistically. In *V. cholerae*, chitin-induced natural transformation can mediate the switch during a short period of time and with a high frequency (Blokesch and Schoolnik, 2007). *E. coli* has not been shown to be naturally competent

(see Sinha and Redfield, 2012 and references cited therein), although several reports mention uptake of plasmid DNA under specific conditions (Baur et al., 1996; Tsen et al., 2002; Etchuuya et al., 2011; Guo et al., 2015). *E. coli* has homologs to competence genes from *Haemophilus influenzae* and they are expressed sufficiently to allow growth on DNA as sole carbon and energy source (Finkel and Kolter, 2001). If the *E. coli* transcription factor Sxy, whose homolog is indispensable for competence development in *H. influenzae*, and the lambda Red recombinase system are artificially expressed, *E. coli* can take up and incorporate foreign DNA into its genome (Sinha and Redfield, 2012). Natural conditions that induce the expression of *sxy* have not been identified so far. However, extraintestinal pathogenic *E. coli* isolates display higher recombination rates than commensal strains (Rodríguez-Beltrán et al., 2015). Higher recombination frequencies are positively associated with the presence of virulence factors (Rodríguez-Beltrán et al., 2015), suggesting that other *E. coli* pathovars might also display increased recombination activity. It will be interesting to study if this is the case for STEC isolates and if conditions promoting host colonization (La Ragione et al., 2009; Barnett Foster, 2013; Pacheco and Sperandio, 2015) can contribute to natural transformability of *E. coli*.

Another possibility for introducing foreign genetic material is generalized transduction by bacteriophages. In the EHEC strain EDL933, the *stx2AB* genes are located on prophage 933W, which is capable of transducing genetic markers in unmodified EHEC and *E. coli* K-12 strains (Marinus and Poteete, 2013). With ~27.4 kB, the length of the entire O-AGC encoding region from the O156:H25 isolate, which differs in sequence from the O182:H25 isolate, is well within the maximal amount of 61 kB that can be transferred by the phage. Despite all being negative for Stx phages (Barth et al., 2016), the O156:H8 isolates described in this study were originally isolated as *stx1* or *stx2* positive colonies (Geue et al., 2002). The loss of *stx* genes can already occur during the first subcultivation step and appears to be more frequent in non-O157 strains (Joris et al., 2011). It is therefore possible that either a lost *stx*-converting phage or other phages encoded in the O156:H8 genomes might have been involved in the generation of transducing phages containing the O-antigen region.

The O-antigen conversion proposed here for the two different serovars reinforces the importance of the O-antigen

for host colonization and bacterial niche adaptation and adds another facet to the enormous genetic diversity and genomic plasticity of *E. coli* (Lukjancenko et al., 2010; Leimbach et al., 2013), again emphasizing the role of HGT in pathogen evolution. It also points to a probably overlooked aspect of *E. coli*/EHEC/STEC pathogenicity. Many different STEC serovars have been linked with human disease (Werber et al., 2008). The proposed HGT-mediated seroconversion suggests that additional genomic characterization of these isolates could reveal that they belong to only a limited number of sequence types each with its own set of specific virulence factors. Analysis of such “viro-STs” or “viro-clonal complexes” would aid the epidemiological analysis of disease outbreaks and pathogen evolution, help identify virulence factors either common to all or rather specific to only one or a few “viro-ST groups” aiding in their characterization and thereby improving our chances of finding and devising better strategies to combat STEC.

## AUTHOR CONTRIBUTIONS

LG, CM, LW, and SB designed the research; LG, IE, DP, CB, and SB performed the research; LG, SB, CB, and TS analyzed data; LG, CB, and CM wrote the paper.

## FUNDING

This work, including the efforts of SB, CM, and LG was funded by Deutsche Forschungsgemeinschaft (DFG) (GE2509/1-1).

## ACKNOWLEDGMENTS

We thank Birgit Mintel and Susann Schares (both FLI, Greifswald-Insel Riems), Anke Hinsching (FLI, Jena), and Petra Krienke (FU Berlin) for their excellent technical assistance.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2017.00424/full#supplementary-material>

## REFERENCES

- Amin, A., Ali, A., Kurunathan, S., Cheong, T. G., Al-Jashamy, K. A., Jaafar, H., et al. (2009). Comparison of histopathological features of *Vibrio cholerae* O1 El Tor and O139 Bengal infections in rabbit intestinal mucosa. *Histol. Histopathol.* 24, 559–565. doi: 10.14670/HH-24.559
- Barnett Foster, D. (2013). Modulation of the enterohemorrhagic *E. coli* virulence program through the human gastrointestinal tract. *Virulence* 4, 315–323. doi: 10.4161/viru.24318
- Barth, S. A., Menge, C., Eichhorn, I., Semmler, T., Wieler, L. H., Pickard, D., et al. (2016). The accessory genome of Shiga toxin-producing *Escherichia coli* defines a persistent colonization type in cattle. *Appl. Environ. Microbiol.* 82, 5455–5464. doi: 10.1128/AEM.00909-16
- Baur, B., Hanselmann, K., Schlimme, W., and Jenni, B. (1996). Genetic transformation in freshwater: *Escherichia coli* is able to develop natural competence. *Appl. Environ. Microbiol.* 62, 3673–3678.
- Bazaka, K., Crawford, R. J., Nazarenko, E. L., and Ivanova, E. P. (2011). Bacterial extracellular polysaccharides. *Adv. Exp. Med. Biol.* 715, 213–226. doi: 10.1007/978-94-007-0940-9\_13
- Bik, E. M., Bunschoten, A. E., Gouw, R. D., and Mooi, F. R. (1995). Genesis of the novel epidemic *Vibrio cholerae* O139 strain: evidence for horizontal transfer of genes involved in polysaccharide synthesis. *EMBO J.* 14, 209–216.
- Blokesch, M., and Schoolnik, G. K. (2007). Serogroup conversion of *Vibrio cholerae* in aquatic reservoirs. *PLoS Pathog.* 3:e81. doi: 10.1371/journal.ppat.0030081
- Brown, P. K., Dozois, C. M., Nickerson, C. A., Zuppardo, A., Terlonge, J., and Curtiss, R. (2001). MlrA, a novel regulator of curli (AgF) and extracellular matrix synthesis by *Escherichia coli* and *Salmonella enterica*

- serovar Typhimurium. *Mol. Microbiol.* 41, 349–363. doi: 10.1046/j.1365-2958.2001.02529.x
- Brunder, W., Schmidt, H., and Karch, H. (1996). KatP, a novel catalase-peroxidase encoded by the large plasmid of enterohaemorrhagic *Escherichia coli* O157:H7. *Microbiology* 142(Pt 11), 3305–3315. doi: 10.1099/13500872-142-11-3305
- Brunder, W., Schmidt, H., and Karch, H. (1997). EspP, a novel extracellular serine protease of enterohaemorrhagic *Escherichia coli* O157:H7 cleaves human coagulation factor V. *Mol. Microbiol.* 24, 767–778. doi: 10.1046/j.1365-2958.1997.3871751.x
- Brunder, W., Schmidt, H., Frosch, M., and Karch, H. (1999). The large plasmids of Shiga-toxin-producing *Escherichia coli* (STEC) are highly variable genetic elements. *Microbiology* 145(Pt 5), 1005–1014. doi: 10.1099/13500872-145-5-1005
- Darling, A. C., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704
- Diez-Villasenor, C., Almendros, C., Garcia-Martinez, J., and Mojica, F. J. (2010). Diversity of CRISPR loci in *Escherichia coli*. *Microbiology* 156(Pt 5), 1351–1361. doi: 10.1099/mic.0.036046-0
- Etchuuya, R., Ito, M., Kitano, S., Shigi, F., Sobue, R., and Maeda, S. (2011). Cell-to-cell transformation in *Escherichia coli*: a novel type of natural transformation involving cell-derived DNA and a putative promoting pheromone. *PLoS ONE* 6:e16355. doi: 10.1371/journal.pone.0016355
- Finkel, S. E., and Kolter, R. (2001). DNA as a nutrient: novel role for bacterial competence gene homologs. *J. Bacteriol.* 183, 6288–6293. doi: 10.1128/JB.183.21.6288-6293.2001
- Geue, L., Schares, S., Mintel, B., Conraths, F. J., Muller, E., and Ehrlich, R. (2010). Rapid microarray-based genotyping of enterohaemorrhagic *Escherichia coli* serotype O156:H25/H-/Hnt isolates from cattle and clonal relationship analysis. *Appl. Environ. Microbiol.* 76, 5510–5519. doi: 10.1128/AEM.00743-10
- Geue, L., Segura-Alvarez, M., Conraths, F. J., Kuczius, T., Bockemuhl, J., Karch, H., et al. (2002). A long-term study on the prevalence of shiga toxin-producing *Escherichia coli* (STEC) on four German cattle farms. *Epidemiol. Infect.* 129, 173–185. doi: 10.1017/s0950268802007288
- Griffin, P. M., and Tauxe, R. V. (1991). The epidemiology of infections caused by *Escherichia coli* O157:H7, other enterohaemorrhagic *E. coli*, and the associated hemolytic uremic syndrome. *Epidemiol. Rev.* 13, 60–98. doi: 10.1093/oxfordjournals.epirev.a036079
- Guo, M., Wang, H., Xie, N., and Xie, Z. (2015). Positive effect of carbon sources on natural transformation in *Escherichia coli*: role of low-level cyclic AMP (cAMP)-cAMP receptor protein in the derepression of rpoS. *J. Bacteriol.* 197, 3317–3328. doi: 10.1128/JB.00291-15
- Hernandes, R. T., Elias, W. P., Vieira, M. A., and Gomes, T. A. (2009). An overview of atypical enteropathogenic *Escherichia coli*. *FEMS Microbiol. Lett.* 297, 137–149. doi: 10.1111/j.1574-6968.2009.01664.x
- Iguchi, A., Iyoda, S., Kikuchi, T., Ogura, Y., Katsura, K., Ohnishi, M., et al. (2015). A complete view of the genetic diversity of the *Escherichia coli* O-antigen biosynthesis gene cluster. *DNA Res.* 22, 101–107. doi: 10.1093/dnares/dsu043
- Joris, M. A., Verstraete, K., Reu, K. D., and Zutter, L. D. (2011). Loss of vtx genes after the first subcultivation step of verocytotoxigenic *Escherichia coli* O157 and Non-O157 during isolation from naturally contaminated fecal samples. *Toxins* 3, 672–677. doi: 10.3390/toxins3060672
- La Ragione, R. M., Best, A., Woodward, M. J., and Wales, A. D. (2009). *Escherichia coli* O157:H7 colonization in small domestic ruminants. *FEMS Microbiol. Rev.* 33, 394–410. doi: 10.1111/j.1574-6976.2008.00138.x
- 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. doi: 10.3389/fmicb.2011.00118
- Leimbach, A., Hacker, J., and Dobrindt, U. (2013). *E. coli* as an all-rounder: the thin line between commensalism and pathogenicity. *Curr. Top. Microbiol. Immunol.* 358, 3–32. doi: 10.1007/82\_2012\_303
- Lerouge, I., and Vanderleyden, J. (2002). O-antigen structural variation: mechanisms and possible roles in animal/plant-microbe interactions. *FEMS Microbiol. Rev.* 26, 17–47. doi: 10.1111/j.1574-6976.2002.tb00597.x
- Liu, B., Knirel, Y. A., Feng, L., Perepelov, A. V., Senchenkova, S. N., Reeves, P. R., et al. (2014). Structural diversity in Salmonella O antigens and its genetic basis. *FEMS Microbiol. Rev.* 38, 56–89. doi: 10.1111/1574-6976.12034
- Lukjancenko, O., Wassenaar, T. M., and Ussery, D. W. (2010). Comparison of 61 sequenced *Escherichia coli* genomes. *Microb. Ecol.* 60, 708–720. doi: 10.1007/s00248-010-9717-3
- Marinus, M. G., and Poteete, A. R. (2013). High efficiency generalized transduction in *Escherichia coli* O157:H7. *F1000Research* 2:7. doi: 10.12688/f1000research.2-7.v1
- McKee, M. L., Melton-Celsa, A. R., Moxley, R. A., Francis, D. H., and O'Brien, A. D. (1995). Enterohaemorrhagic *Escherichia coli* O157:H7 requires intimin to colonize the gnotobiotic pig intestine and to adhere to HEp-2 cells. *Infect. Immun.* 63, 3739–3744.
- Mekalanos, J. J., Rubin, E. J., and Waldor, M. K. (1997). Cholera: molecular basis for emergence and pathogenesis. *FEMS Immunol. Med. Microbiol.* 18, 241–248. doi: 10.1111/j.1574-695X.1997.tb01052.x
- Nataro, J. P., and Kaper, J. B. (1998). Diarrheagenic *Escherichia coli*. *Clin. Microbiol. Rev.* 11, 142–201.
- Niba, E. T., Naka, Y., Nagase, M., Mori, H., and Kitakawa, M. (2007). A genome-wide approach to identify the genes involved in biofilm formation in *E. coli*. *DNA Res.* 14, 237–246. doi: 10.1093/dnares/dsm024
- Pacheco, A. R., and Sperandio, V. (2015). Enteric pathogens exploit the microbiota-generated nutritional environment of the gut. *Microbiol. Spectr.* doi: 10.1128/microbiolspec.MBP-0001-2014
- Patel, R. K., and Jain, M. (2012). NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE* 7:e30619. doi: 10.1371/journal.pone.0030619
- Paton, A. W., and Paton, J. C. (1998). Detection and characterization of Shiga toxigenic *Escherichia coli* by using multiplex PCR assays for stx1, stx2, eaeA, enterohaemorrhagic *E. coli* hlyA, rfbO111, and rfbO157. *J. Clin. Microbiol.* 36, 598–602.
- Penner, J. L., and Aspinall, G. O. (1997). Diversity of lipopolysaccharide structures in *Campylobacter jejuni*. *J. Infect. Dis.* 176(Suppl. 2), S135–S138. doi: 10.1086/513778
- Ramamurthy, T., Yamasaki, S., Takeda, Y., and Nair, G. B. (2003). Vibrio cholerae O139 Bengal: odyssey of a fortuitous variant. *Microbes Infect.* 5, 329–344. doi: 10.1016/S1286-4579(03)00035-2
- Reeves, P. (1995). Role of O-antigen variation in the immune response. *Trends Microbiol.* 3, 381–386. doi: 10.1016/S0966-842X(00)88983-0
- Remuzzi, G., and Ruggenenti, P. (1998). The hemolytic uremic syndrome. *Kidney Int.* 66(Suppl.), S54–S57.
- Rodríguez-Beltrán, J., Tournet, J., Tenaillon, O., López, E., Bourdelier, E., Costas, C., et al. (2015). High recombinant frequency in extraintestinal pathogenic *Escherichia coli* strains. *Mol. Biol. Evol.* 32, 1708–1716. doi: 10.1093/molbev/msv072
- 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, 2503–2519. doi: 10.1016/j.carres.2003.07.009
- Samuel, G., Hogbin, J. P., Wang, L., and Reeves, P. R. (2004). Relationships of the *Escherichia coli* O157, O111, and O55 O-antigen gene clusters with those of *Salmonella enterica* and *Citrobacter freundii*, which express identical O antigens. *J. Bacteriol.* 186, 6536–6543. doi: 10.1128/JB.186.19.6536-6543.2004
- Schmidt, H., Beutin, L., and Karch, H. (1995). Molecular analysis of the plasmid-encoded hemolysin of *Escherichia coli* O157:H7 strain EDL 933. *Infect. Immun.* 63, 1055–1061.
- Sinha, S., and Redfield, R. J. (2012). Natural DNA uptake by *Escherichia coli*. *PLoS ONE* 7:e35620. doi: 10.1371/journal.pone.0035620
- Sperandio, V. (2001). Genome sequence of *E. coli* O157:H7. *Trends Microbiol.* 9:159. doi: 10.1016/S0966-842X(01)02023-6
- Stentz, R., Weintraub, A., and Widmalm, G. (2006). The structures of *Escherichia coli* O-polysaccharide antigens. *FEMS Microbiol. Rev.* 30, 382–403. doi: 10.1111/j.1574-6976.2006.00016.x
- Su, C., and Brandt, L. J. (1995). *Escherichia coli* O157:H7 infection in humans. *Ann. Intern. Med.* 123, 698–714. doi: 10.7326/0003-4819-123-9-199511010-00009
- Sugiyama, T., Kido, N., Kato, Y., Koide, N., Yoshida, T., and Yokochi, T. (1997). Evolutionary relationship among rfb gene clusters synthesizing mannose homopolymer as O-specific polysaccharides in *Escherichia coli* and *Klebsiella*. *Gene* 198, 111–113. doi: 10.1016/s0378-1119(97)00300-4

- Tarr, P. I., Schoening, L. M., Yea, Y. L., Ward, T. R., Jelacic, S., and Whittam, T. S. (2000). Acquisition of the rfb-gnd cluster in evolution of *Escherichia coli* O55 and O157. *J. Bacteriol.* 182, 6183–6191. doi: 10.1128/JB.182.21.6183-6191.2000
- Touchon, M., Charpentier, S., Clermont, O., Rocha, E. P. C., Denamur, E., and Branger, C. (2011). CRISPR distribution within the *Escherichia coli* species is not suggestive of immunity-associated diversifying selection. *J. Bacteriol.* 193, 2460–2467. doi: 10.1128/JB.01307-10
- Tsen, S. D., Fang, S. S., Chen, M. J., Chien, J. Y., Lee, C. C., and Tsen, D. H. (2002). Natural plasmid transformation in *Escherichia coli*. *J. Biomed. Sci.* 9, 246–252. doi: 10.1159/000059425
- Waldor, M. K., Colwell, R., and Mekalanos, J. J. (1994). The *Vibrio cholerae* O139 serogroup antigen includes an O-antigen capsule and lipopolysaccharide virulence determinants. *Proc. Natl. Acad. Sci. U.S.A.* 91, 11388–11392. doi: 10.1073/pnas.91.24.11388
- Wang, L., Huskic, S., Cisterne, A., Rothmund, D., and Reeves, P. R. (2002). The O-antigen gene cluster of *Escherichia coli* O55:H7 and identification of a new UDP-GlcNAc C4 epimerase gene. *J. Bacteriol.* 184, 2620–2625. doi: 10.1128/JB.184.10.2620-2625.2002
- Wang, L., Qu, W., and Reeves, P. R. (2001). Sequence analysis of four *Shigella boydii* O-antigen loci: implication for *Escherichia coli* and *Shigella* relationships. *Infect. Immun.* 69, 6923–6930. doi: 10.1128/IAI.69.11.6923-6930.2001
- Werber, D., Beutin, L., Pichner, R., Stark, K., and Fruth, A. (2008). Shiga toxin-producing *Escherichia coli* serogroups in food and patients, Germany. *Emerg. Infect. Dis.* 14, 1803–1806. doi: 10.3201/eid1411.080361
- Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L. H., et al. (2006). Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol. Microbiol.* 60, 1136–1151. doi: 10.1111/j.1365-2958.2006.05172.x
- Yin, S., Jensen, M. A., Bai, J., Debröy, C., Barrangou, R., and Dudley, E. G. (2013). The evolutionary divergence of Shiga toxin-producing *Escherichia coli* is reflected in clustered regularly interspaced short palindromic repeat (CRISPR) spacer composition. *Appl. Environ. Microbiol.* 79, 5710–5720. doi: 10.1128/AEM.00950-13

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Geue, Menge, Eichhorn, Semmler, Wieler, Pickard, Berens and Barth. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.